

Задание №2

№1. (К теоретической задаче 1)

Сгенерируйте выборку X_1, \dots, X_N из равномерного распределения на отрезке $[0, \theta]$ для $N = 10^4$. Для всех $n \leq N$ посчитайте оценки параметра θ из теоретической задачи:

$2\bar{X}$, $\bar{X} + \frac{X_{(n)}}{2}$, $(n+1)X_{(1)}$, $X_{(1)} + X_{(n)}$, $\frac{n+1}{n}X_{(n)}$. Постройте на одном графике разными цветами для всех оценок функции модуля разности оценки и истинного значения θ в зависимости от n . Если некоторые оценки (при фиксированном значении θ) сильно отличаются от истинного значения параметра θ , то исключите их и постройте еще один график со всеми кривыми (для измененного значения θ). Для избавления от больших значений разности в начале ограничьте масштаб графика. Для наглядности точки можно соединить линиями. Какая оценка получилась лучше (в смысле упомянутого модуля разности при $n = N$)? Проведите эксперимент для разных значений θ (количество графиков равно количеству значений θ).

In [1]:

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import uniform

%matplotlib inline
```

In [2]:

```
N = 10000

# Сгенерируем выборку для theta = 1.
theta = 1
sample = uniform.rvs(size=N, scale=theta)
```

In [3]:

```
# Посчитаем оценки для каждого n.
sample_mean = sample.cumsum() * [2 / n for n in range(1, N + 1, 1)]
sample_mean_max = sample.cumsum() / [n for n in range(1, N + 1, 1)] +
[np.max(sample[ : n + 1])/2 for n in range(N)]
coeff_min = np.array([(n + 1) * np.min(sample[ : n]) for n in range(1, N + 1, 1)])
min_max = np.array([np.min(sample[ : n + 1]) + np.max(sample[ : n + 1]) for n in range(N)])
coeff_max = np.array([(n + 1) / n * np.max(sample[ : n]) for n in range(1, N + 1, 1)])
```

In [4]:

```
# Построим график функции модуля разности оценки и истинного значения theta.
plt.figure(figsize=(10, 5))
plt.plot(np.linspace(1., N, N), abs(sample_mean - theta), label='sample_mean')
plt.plot(np.linspace(1., N, N), abs(sample_mean_max - theta), label='sample_mean_max')
plt.plot(np.linspace(1., N, N), abs(koeff_min - theta), label='koeff_min')
plt.plot(np.linspace(1., N, N), abs(min_max - theta), label='min_max')
plt.plot(np.linspace(1., N, N), abs(koeff_max - theta), label='koeff_max')
plt.ylim(-100, 100)

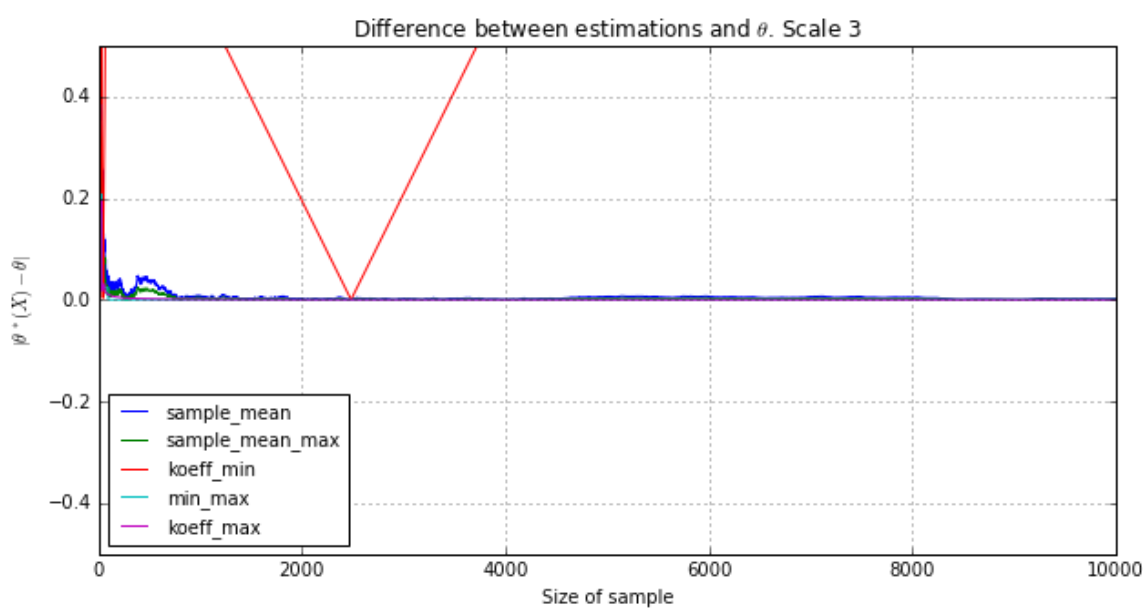
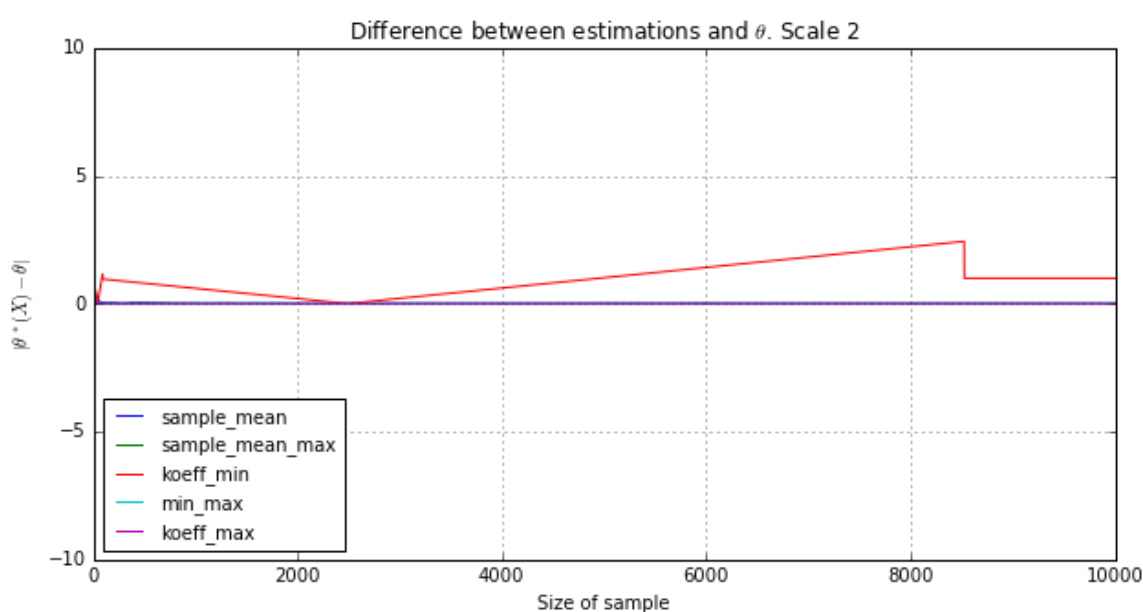
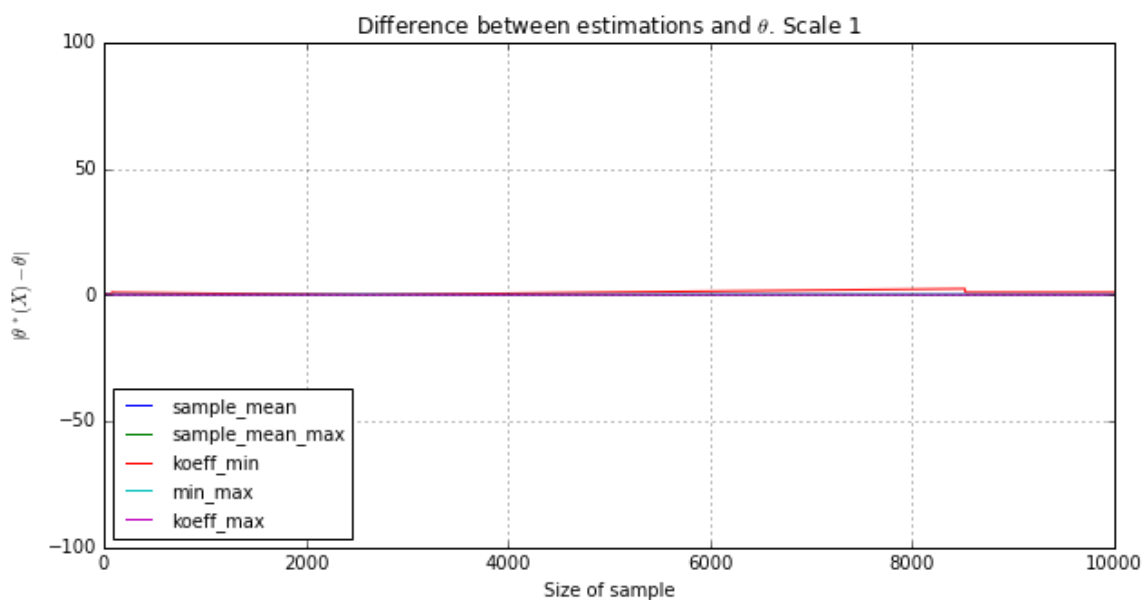
plt.title(r'Difference between estimations and  $\theta$ . Scale 1')
plt.xlabel(r'Size of sample', fontsize='10')
plt.ylabel(r' $|\hat{\theta}(X) - \theta|$ ', fontsize='10')
plt.grid()
plt.legend(fontsize=10, loc=3)
plt.show()

plt.figure(figsize=(10, 5))
plt.plot(np.linspace(1., N, N), abs(sample_mean - theta), label='sample_mean')
plt.plot(np.linspace(1., N, N), abs(sample_mean_max - theta), label='sample_mean_max')
plt.plot(np.linspace(1., N, N), abs(koeff_min - theta), label='koeff_min')
plt.plot(np.linspace(1., N, N), abs(min_max - theta), label='min_max')
plt.plot(np.linspace(1., N, N), abs(koeff_max - theta), label='koeff_max')
plt.ylim(-10, 10)

plt.title(r'Difference between estimations and  $\theta$ . Scale 2')
plt.xlabel(r'Size of sample', fontsize='10')
plt.ylabel(r' $|\hat{\theta}(X) - \theta|$ ', fontsize='10')
plt.grid()
plt.legend(fontsize=10, loc=3)
plt.show()

plt.figure(figsize=(10, 5))
plt.plot(np.linspace(1., N, N), abs(sample_mean - theta), label='sample_mean')
plt.plot(np.linspace(1., N, N), abs(sample_mean_max - theta), label='sample_mean_max')
plt.plot(np.linspace(1., N, N), abs(koeff_min - theta), label='koeff_min')
plt.plot(np.linspace(1., N, N), abs(min_max - theta), label='min_max')
plt.plot(np.linspace(1., N, N), abs(koeff_max - theta), label='koeff_max')
plt.ylim(-0.5, 0.5)

plt.title(r'Difference between estimations and  $\theta$ . Scale 3')
plt.xlabel(r'Size of sample', fontsize='10')
plt.ylabel(r' $|\hat{\theta}(X) - \theta|$ ', fontsize='10')
plt.grid()
plt.legend(fontsize=10, loc=3)
plt.show()
```



Исключим из дальнейших экспериментов оценку $(n + 1)X_{(1)}$, поскольку эта оценка сильно отличается от истинного значения θ .

In [5]:

```
# Сгенерируем выборку для значения theta = 10.  
theta = 10  
sample = uniform.rvs(size=N, scale=theta)
```

In [6]:

```
# Посчитаем оценки для каждого n (coeff_min считать не будем).  
sample_mean = sample.cumsum() * [2 / n for n in range(1, N + 1, 1)]  
sample_mean_max = sample.cumsum() / [n for n in range(1, N + 1, 1)] +  
[np.max(sample[ : n + 1])/2 for n in range(N)]  
min_max = np.array([np.min(sample[ : n + 1]) + np.max(sample[ : n + 1]) for n in  
range(N)])  
coeff_max = np.array([(n + 1) / n * np.max(sample[ : n]) for n in range(1, N +  
1, 1)])
```

In [7]:

```
# Построим график функции модуля разности оценки и истинного значения theta.
plt.figure(figsize=(10, 5))
plt.plot(np.linspace(1., N, N), abs(sample_mean - theta), label='sample_mean')
plt.plot(np.linspace(1., N, N), abs(sample_mean_max - theta), label='sample_mean_max')
plt.plot(np.linspace(1., N, N), abs(min_max - theta), label='min_max')
plt.plot(np.linspace(1., N, N), abs(koeff_max - theta), label='koeff_max')
plt.ylim(-2, 2)

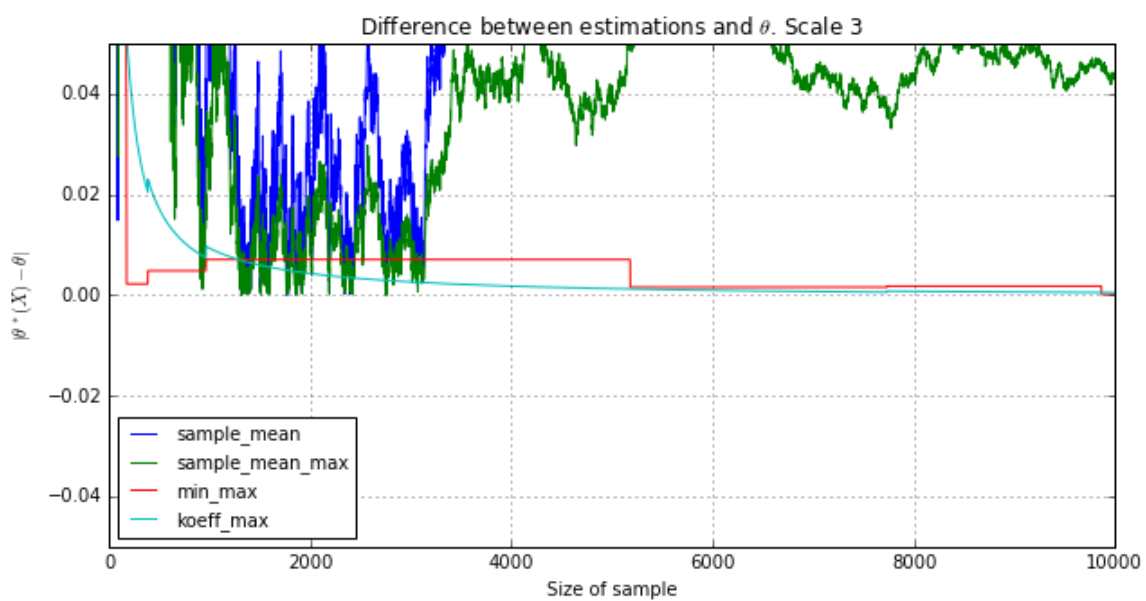
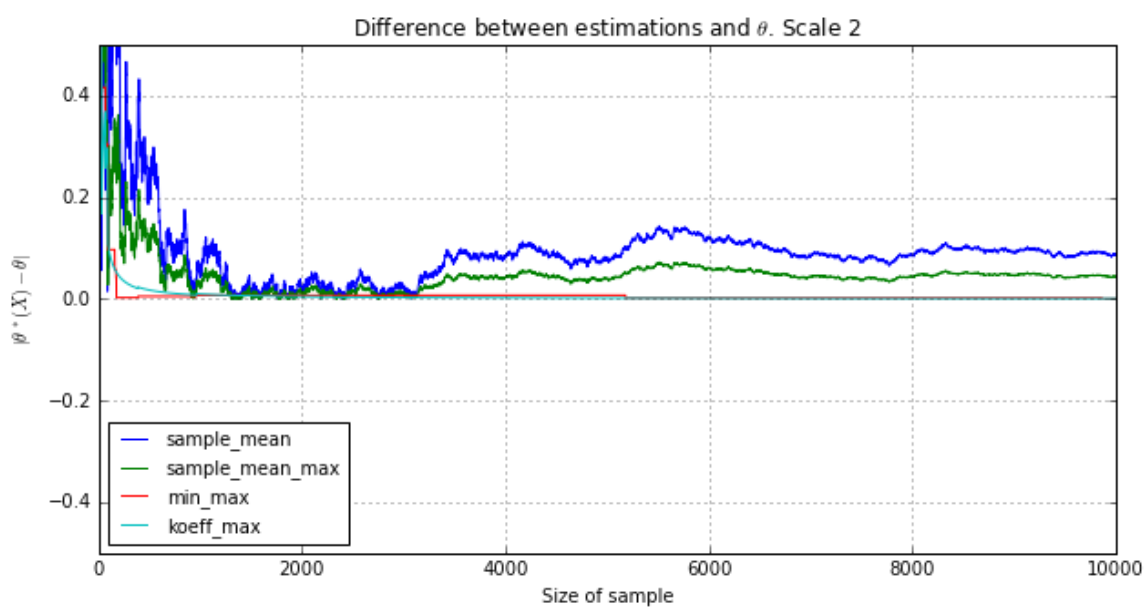
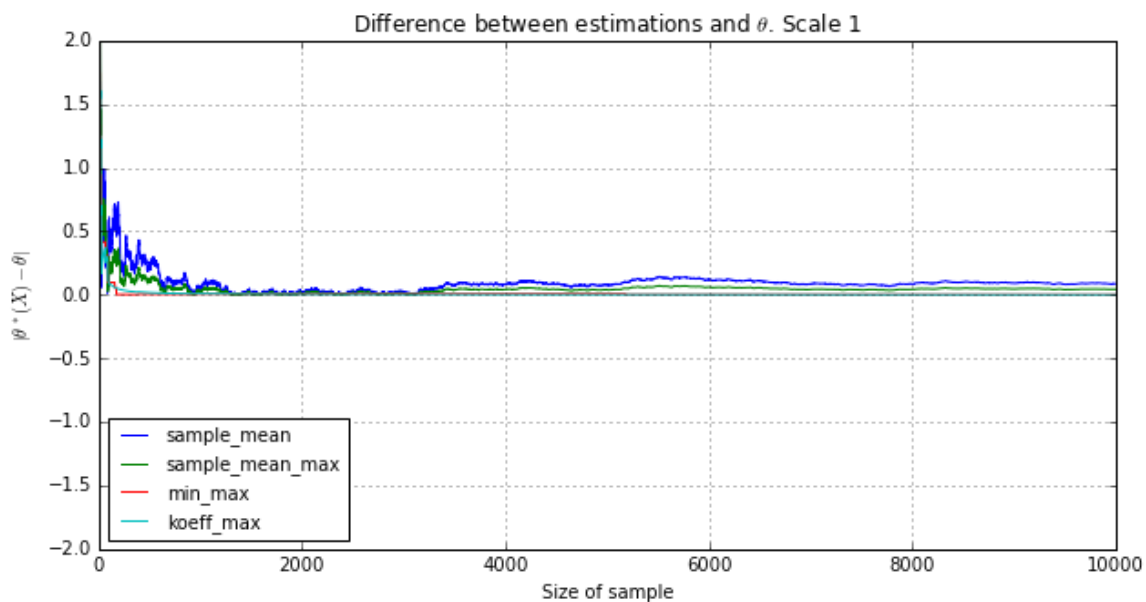
plt.title(r'Difference between estimations and  $\theta$ . Scale 1')
plt.xlabel(r'Size of sample', fontsize='10')
plt.ylabel(r' $|\theta^*(X) - \theta|$ ', fontsize='10')
plt.grid()
plt.legend(fontsize=10, loc=3)
plt.show()

plt.figure(figsize=(10, 5))
plt.plot(np.linspace(1., N, N), abs(sample_mean - theta), label='sample_mean')
plt.plot(np.linspace(1., N, N), abs(sample_mean_max - theta), label='sample_mean_max')
plt.plot(np.linspace(1., N, N), abs(min_max - theta), label='min_max')
plt.plot(np.linspace(1., N, N), abs(koeff_max - theta), label='koeff_max')
plt.ylim(-0.5, 0.5)

plt.title(r'Difference between estimations and  $\theta$ . Scale 2')
plt.xlabel(r'Size of sample', fontsize='10')
plt.ylabel(r' $|\theta^*(X) - \theta|$ ', fontsize='10')
plt.grid()
plt.legend(fontsize=10, loc=3)
plt.show()

plt.figure(figsize=(10, 5))
plt.plot(np.linspace(1., N, N), abs(sample_mean - theta), label='sample_mean')
plt.plot(np.linspace(1., N, N), abs(sample_mean_max - theta), label='sample_mean_max')
plt.plot(np.linspace(1., N, N), abs(min_max - theta), label='min_max')
plt.plot(np.linspace(1., N, N), abs(koeff_max - theta), label='koeff_max')
plt.ylim(-0.05, 0.05)

plt.title(r'Difference between estimations and  $\theta$ . Scale 3')
plt.xlabel(r'Size of sample', fontsize='10')
plt.ylabel(r' $|\theta^*(X) - \theta|$ ', fontsize='10')
plt.grid()
plt.legend(fontsize=10, loc=3)
plt.show()
```



Уберем из рассмотрения еще две оценки: $2\bar{X}$, $\bar{X} + \frac{X_{(n)}}{2}$.

In [8]:

```
# Сгенерируем выборку для значения theta = 77.  
theta = 77  
sample = uniform.rvs(size=N, scale=theta)
```

In [9]:

```
# Посчитаем оценки для каждого n (убрали sample_mean и sample_mean_max).  
min_max = np.array([np.min(sample[ : n + 1]) + np.max(sample[ : n + 1]) for n in  
    range(N)])  
coeff_max = np.array([(n + 1) / n * np.max(sample[ : n]) for n in range(1, N +  
    1, 1)])
```

In [10]:

```
# Построим график функции модуля разности оценки и истинного значения theta.
plt.figure(figsize=(10, 5))
plt.plot(np.linspace(1., N, N), abs(min_max - theta), label='min_max')
plt.plot(np.linspace(1., N, N), abs(koeff_max - theta), label='koeff_max')
plt.ylim(-2, 2)

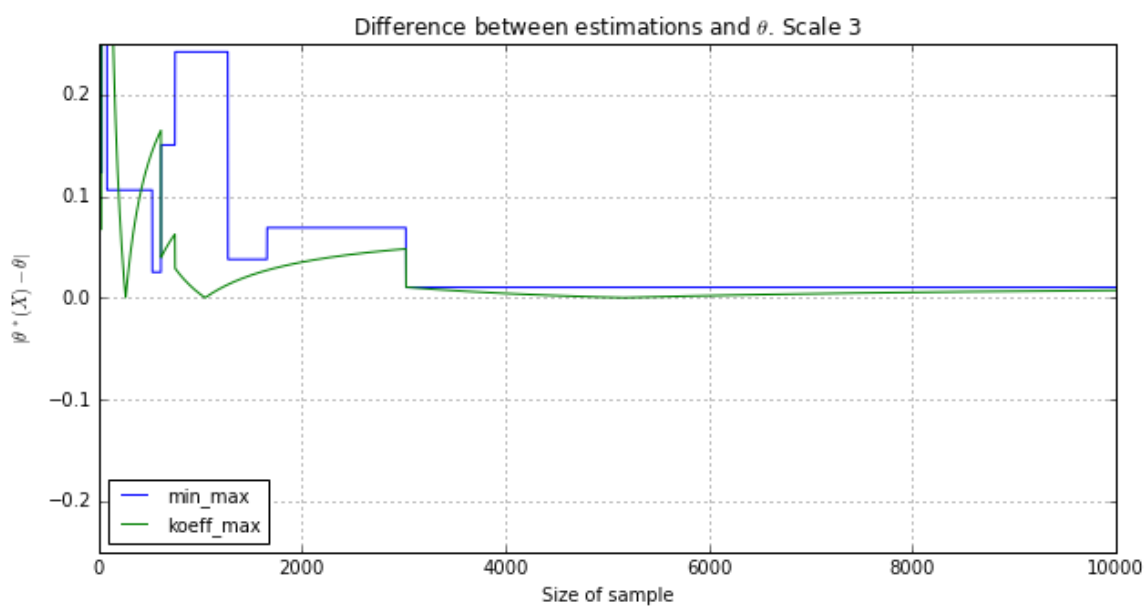
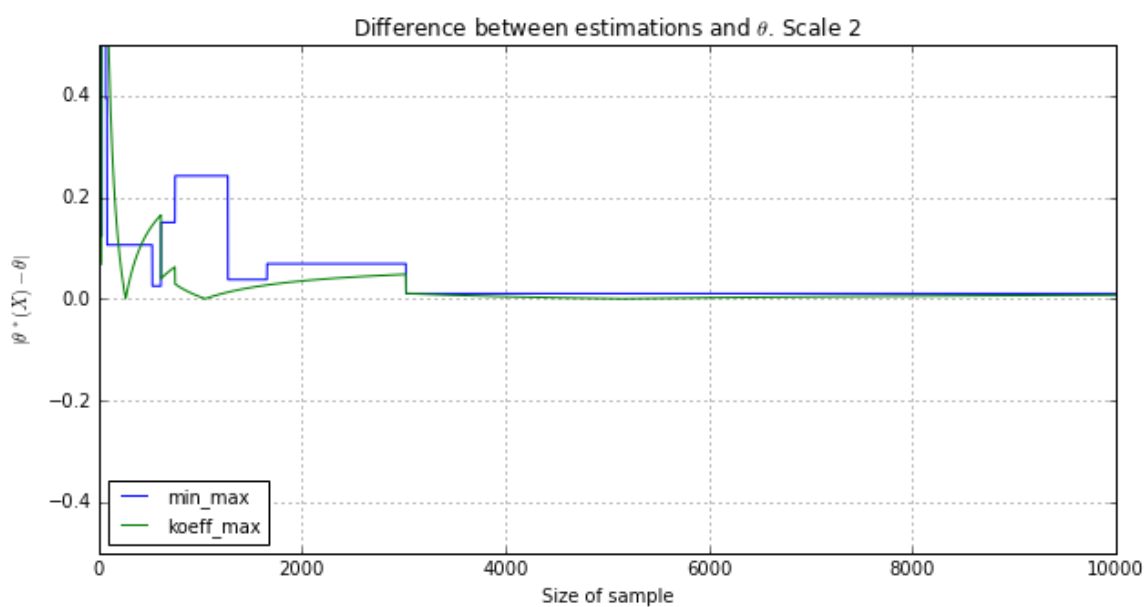
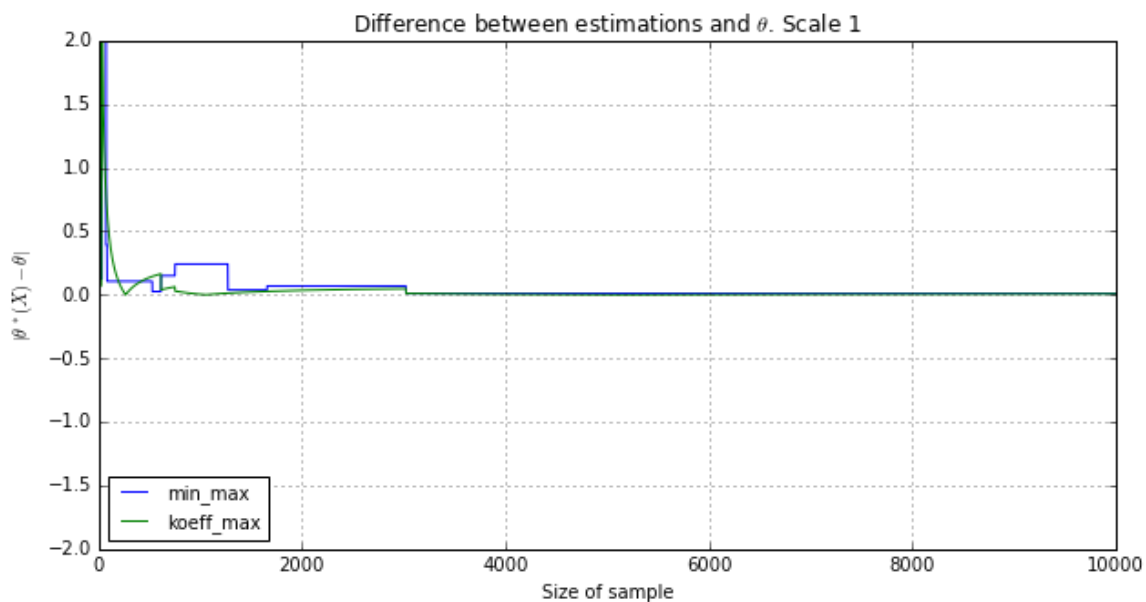
plt.title(r'Difference between estimations and  $\theta$ . Scale 1')
plt.xlabel(r'Size of sample', fontsize='10')
plt.ylabel(r' $|\theta^*(X) - \theta|$ ', fontsize='10')
plt.grid()
plt.legend(fontsize=10, loc=3)
plt.show()

plt.figure(figsize=(10, 5))
plt.plot(np.linspace(1., N, N), abs(min_max - theta), label='min_max')
plt.plot(np.linspace(1., N, N), abs(koeff_max - theta), label='koeff_max')
plt.ylim(-0.5, 0.5)

plt.title(r'Difference between estimations and  $\theta$ . Scale 2')
plt.xlabel(r'Size of sample', fontsize='10')
plt.ylabel(r' $|\theta^*(X) - \theta|$ ', fontsize='10')
plt.grid()
plt.legend(fontsize=10, loc=3)
plt.show()

plt.figure(figsize=(10, 5))
plt.plot(np.linspace(1., N, N), abs(min_max - theta), label='min_max')
plt.plot(np.linspace(1., N, N), abs(koeff_max - theta), label='koeff_max')
plt.ylim(-0.25, 0.25)

plt.title(r'Difference between estimations and  $\theta$ . Scale 3')
plt.xlabel(r'Size of sample', fontsize='10')
plt.ylabel(r' $|\theta^*(X) - \theta|$ ', fontsize='10')
plt.grid()
plt.legend(fontsize=10, loc=3)
plt.show()
```

In [11]:

```
# Сгенерируем еще для значения theta = 0.5.  
theta = 0.5  
sample = uniform.rvs(size=N, scale=theta)
```

In [12]:

```
# Посчитаем оценки для каждого n.  
min_max = np.array([np.min(sample[ : n + 1]) + np.max(sample[ : n + 1]) for n in  
    range(N)])  
coeff_max = np.array([(n + 1) / n * np.max(sample[ : n]) for n in range(1, N +  
    1, 1)])
```

In [13]:

```
# Построим график функции модуля разности оценки и истинного значения theta.
plt.figure(figsize=(10, 5))
plt.plot(np.linspace(1., N, N), abs(min_max - theta), label='min_max')
plt.plot(np.linspace(1., N, N), abs(koeff_max - theta), label='koeff_max')
plt.ylim(-0.02, 0.02)

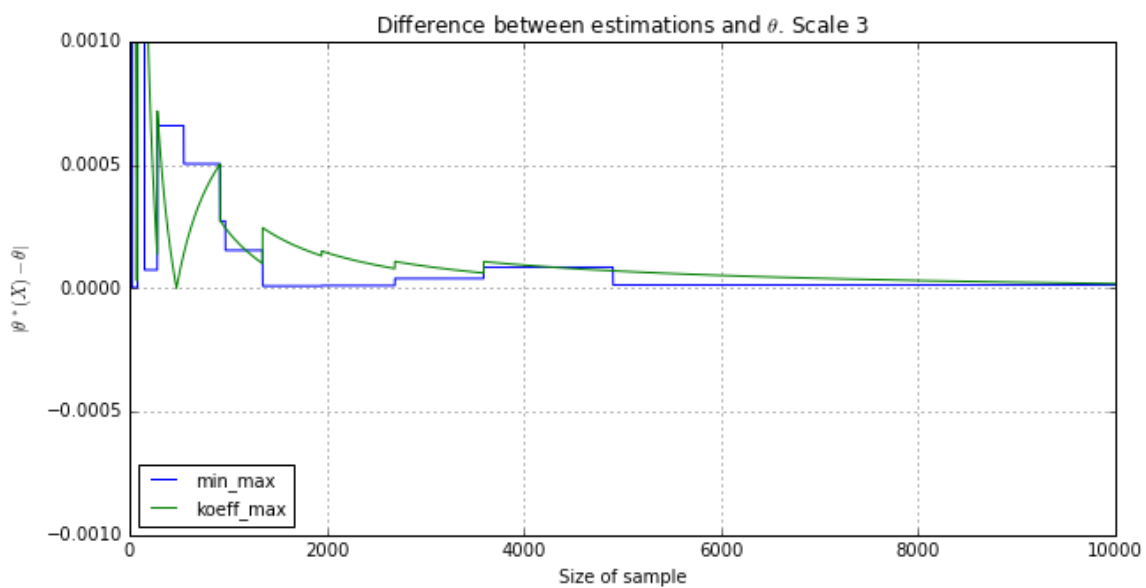
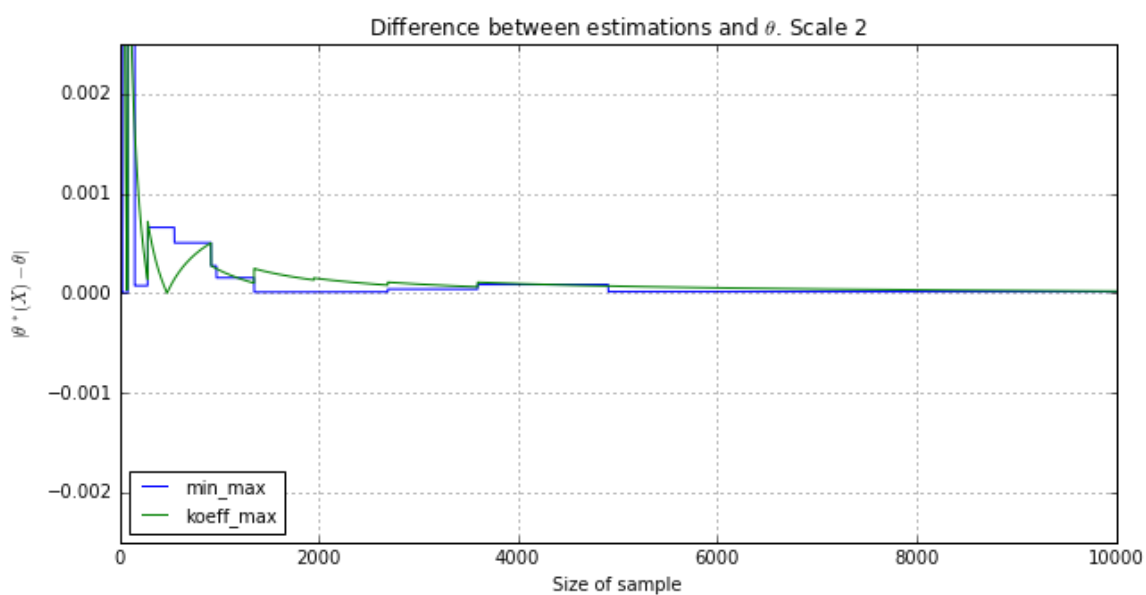
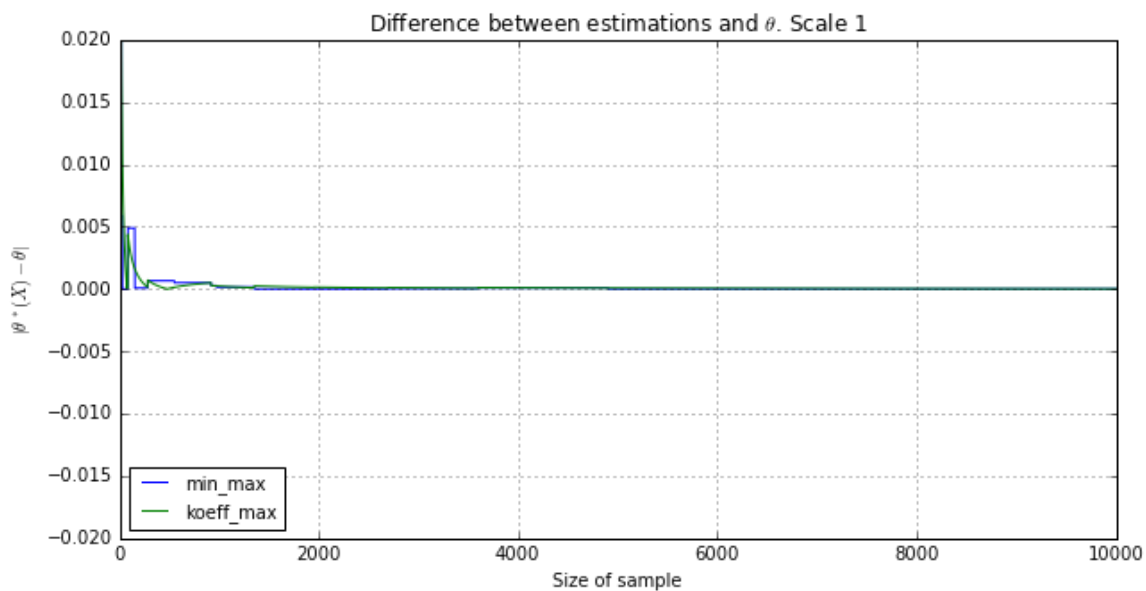
plt.title(r'Difference between estimations and  $\theta$ . Scale 1')
plt.xlabel(r'Size of sample', fontsize='10')
plt.ylabel(r' $|\hat{\theta}^*(X) - \theta|$ ', fontsize='10')
plt.grid()
plt.legend(fontsize=10, loc=3)
plt.show()

plt.figure(figsize=(10, 5))
plt.plot(np.linspace(1., N, N), abs(min_max - theta), label='min_max')
plt.plot(np.linspace(1., N, N), abs(koeff_max - theta), label='koeff_max')
plt.ylim(-0.0025, 0.0025)

plt.title(r'Difference between estimations and  $\theta$ . Scale 2')
plt.xlabel(r'Size of sample', fontsize='10')
plt.ylabel(r' $|\hat{\theta}^*(X) - \theta|$ ', fontsize='10')
plt.grid()
plt.legend(fontsize=10, loc=3)
plt.show()

plt.figure(figsize=(10, 5))
plt.plot(np.linspace(1., N, N), abs(min_max - theta), label='min_max')
plt.plot(np.linspace(1., N, N), abs(koeff_max - theta), label='koeff_max')
plt.ylim(-0.001, 0.001)

plt.title(r'Difference between estimations and  $\theta$ . Scale 3')
plt.xlabel(r'Size of sample', fontsize='10')
plt.ylabel(r' $|\hat{\theta}^*(X) - \theta|$ ', fontsize='10')
plt.grid()
plt.legend(fontsize=10, loc=3)
plt.show()
```



Вывод:

Самой плохой оказалась оценка $(n + 1)X_{(1)}$, поэтому из всех экспериментов, кроме первого, она была исключена. Также оценки $2\bar{X}$, $\bar{X} + \frac{X_{(n)}}{2}$ оказались не очень точными (видно из графиков экспериментов 1 и 2), поэтому они также были исключены из последующих измерений. Самыми точными (в смысле упомянутого модуля разности при $n = N$) оказались оценки $X_{(1)} + X_{(n)}$, $\frac{n+1}{n}X_{(n)}$.