**Кафедра дискретной математики МФТИ**

**Курс математической статистики**

Игашов Илья, 593 группппа

# Задание №2

## №3. (К теоретической задаче 5) ¶

Придумайте распределение, у которого конечны первые четыре момента, а пятый — нет. Сгенерируйте выборку $X_1, \ldots, X_N$ из этого распределения для $N = 10^4$. Постройте график плотности, а также нанесите точки выборки на график (с нулевой y-координатой). Для всех $n \leq N$ посчитайте оценку $s^2 = s^2(X_1, \ldots, X_N)$ для дисперсии. Постройте график зависимости модуля разности оценки дисперсии и ее истинного значения от $n$. Проведите аналогичное исследование для выборки из распределения Коши, где вместо графика модуля разности оценки дисперсии и ее истинного значения (которого не существует) постройте график оценки дисперсии.

Искомым распредедением является распределение Парето $f_\gamma(x) = \gamma x^{-1-\gamma} I(x \geq 1)$ для $\gamma = 5$.

In [1]:

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

%matplotlib inline
```

In [2]:

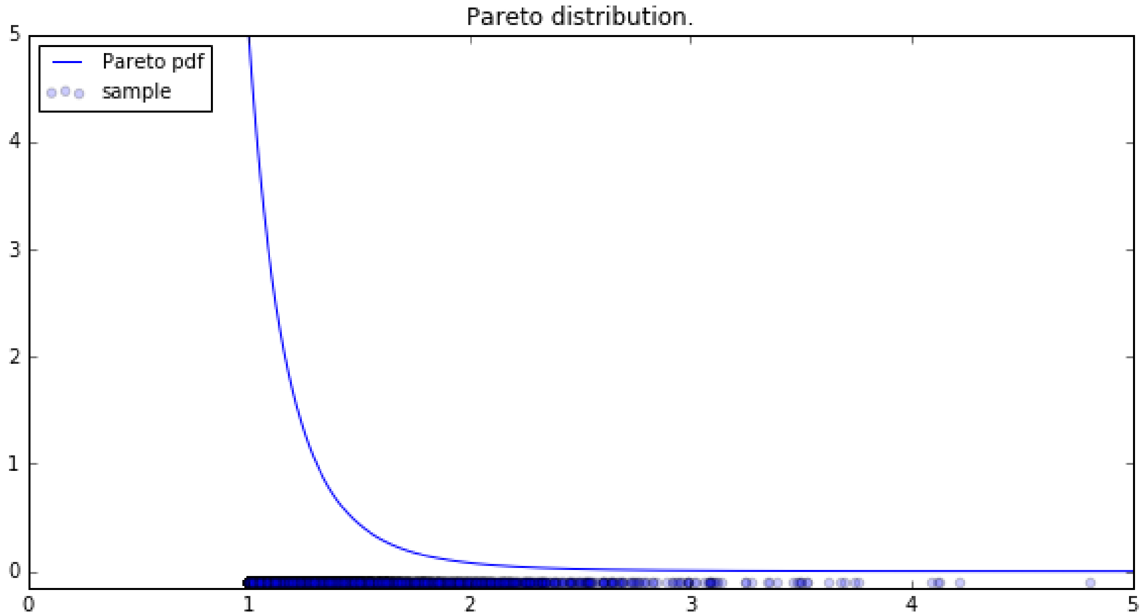
```
# Генерируем выборку размера N = 10000 из распределения Парето.
N = 10000
sample = pareto.rvs(5, size=N)
```

In [3]:

```
# Построим график плотности распределения Парето.
grid = np.linspace(1, 100, 1000000) # Сетка для плотности.

plt.figure(figsize=(10, 5))
plt.plot(grid, pareto.pdf(grid, 5), label='Pareto pdf')
plt.scatter(sample, np.zeros(N) - 0.1, alpha=0.2, label='sample')

plt.ylim(-0.15, 5)
plt.xlim(0, 5)
plt.title(r'Pareto distribution.')
plt.legend(fontsize=10, loc=2)
plt.show()
```



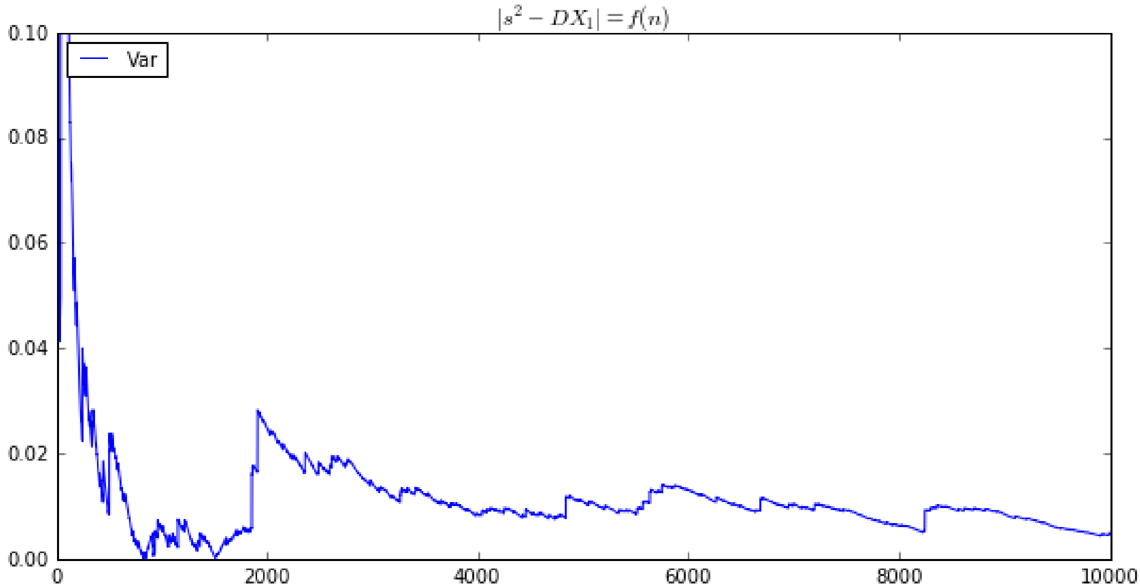

In [4]:

```
# Строим выборочную дисперсию
s = np.cumsum(sample ** 2) * [1/(n + 1) for n in range(N)] - (np.cumsum(sample)
* [1/(n + 1) for n in range(N)]) ** 2
```

In [5]:

```
# Построим график зависимости модуля разности оценки дисперсии s и истинной дисперсии.
plt.figure(figsize=(10, 5))
plt.plot(np.linspace(1, N, N), np.abs(s -  pareto.var(5)), label='Var')

plt.ylim(0, 0.1)
plt.title(r'$|s^2 - DX_1| = f(n)$')
plt.legend(fontsize=10, loc=2)
plt.show()
```




In [6]:

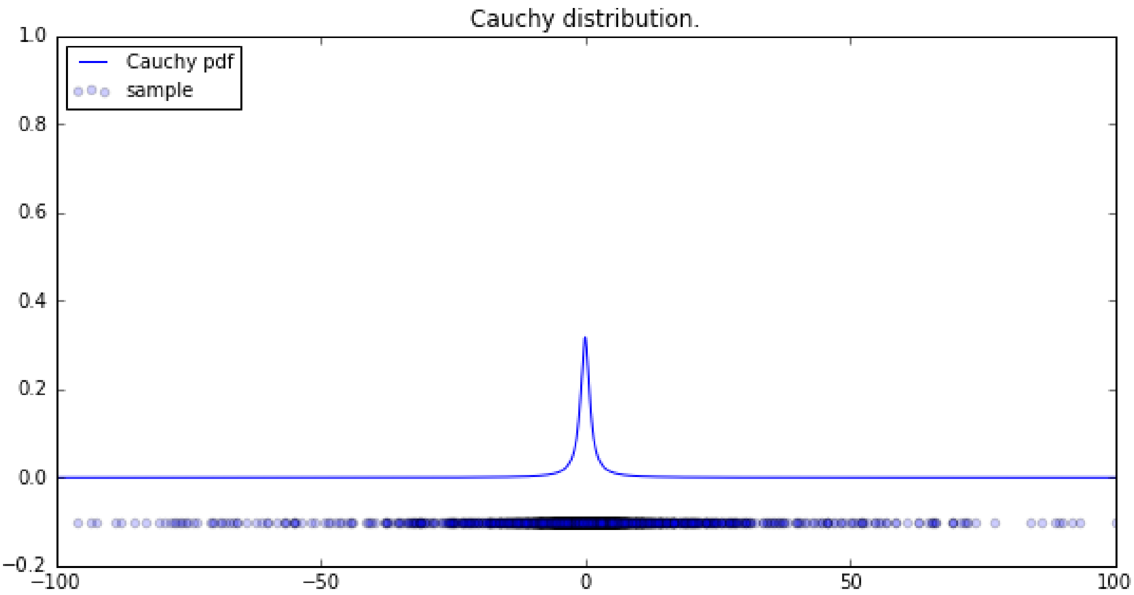
```
# Генерируем выборку размера N = 10000 из распределения Коши.
sample = cauchy.rvs(size=N)
```

In [7]:

```
# Постороим график плотности распределения Коши.
grid = np.linspace(-100, 100, 1000000)

plt.figure(figsize=(10, 5))
plt.plot(grid, cauchy.pdf(grid), label='Cauchy pdf')
plt.scatter(sample, np.zeros(N) - 0.1, alpha=0.2, label='sample')

plt.ylim(-0.2, 1)
plt.xlim(-100, 100)
plt.title(r'Cauchy distribution.')
plt.legend(fontsize=10, loc=2)
plt.show()
```



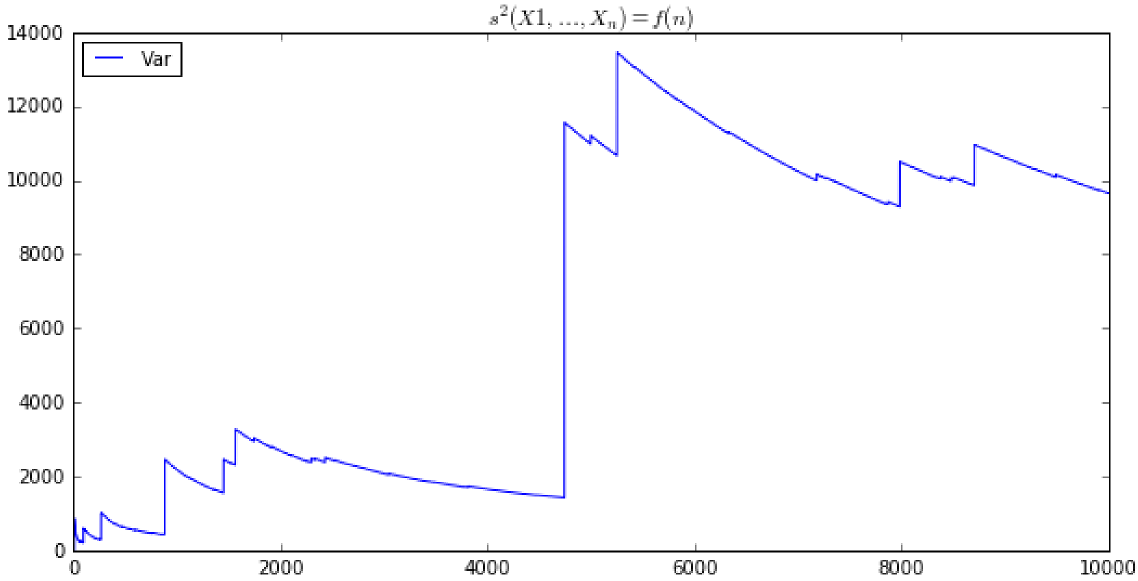

In [8]:

```
# Строим выборочную дисперсию
s = np.cumsum(sample ** 2) * [1/(n + 1) for n in range(N)] - (np.cumsum(sample)
* [1/(n + 1) for n in range(N)]) ** 2
```

In [9]:

```
# Постороим график зависимости выборочной дисперсии от n.
plt.figure(figsize=(10, 5))
plt.plot(np.linspace(1, N, N), s, label='Var')

plt.title(r'$s^2(X1,\dots,X_n) = f(n)$')
plt.legend(fontsize=10, loc=2)
plt.show()
```




# Вывод

У распределения Парето $f_\gamma(x) = \gamma x^{-1-\gamma} I(x \geq 1)$ с параметром $\gamma = 5$ первые четыре момента конечны, а пятый — нет. Мы построили график его плотности, посчитали выборочную дисперсию и построили график $|s^2 - DX_1| = f(n)$, который показал, что при больших n выборочная дисперсия не сильно отличается от истинного значения дисперсии данного распределения. Также мы рассмотрели распределение Коши, у которого, как известно, не существует математического ожидания и дисперсии. Мы построили график его плотности, посчитали выборочную дисперсию и построили ее график. Ничего определенного по последнему графику сказать нельзя.