Кафедра дискретной математики МФТИ

Курс математической статистики

Игашов Илья, 593 группппа

Задание №3

№3.

В банке каждую минуту подсчитывается баланс по сравнению с началом дня (6 часов утра). В полночь работники банка измеряют две величины: X_1 — максимальное значение баланса за день, X_2 — значение баланса в полночь. Считается, что величина $X = X_1 - X_2$ имеет распределение Вейбулла с функцией распределения $1 - e^{-x^\gamma} I(x \geq 0)$, где $\gamma > 0$ — параметр формы. В течение 10 лет каждый день банк проводил измерение величины X, получив, в результате выборку $X_1, \ldots, X_{3652}$. В файле Weibull.csv находятся соответствующие измерения. Оцените параметр формы методом максимального правдоподобия а) по первым 4 годам; б) по всей выборке. Оценку произведите по сетке (в логарифмической шкале). Известно, что $\log_{10} \gamma \in [-2, 2]$. Выберите шаг равным 10^{-3}.

In [29]:

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

%matplotlib inline
```

In [30]:

```
# Записываем выборку из файла в список sample
f = open('Weibull.csv.xls', 'r')
text = f.read()
lines = text.split('\n')
sample = np.array([float(lines[n]) if float(lines[n]) else 0.0001 for n in range(len(lines) - 1)])
```

In [31]:

```
# Логарифм функции правдоподобия (сумма логарифмов значений плотности на каждом значении из sample).
# log_shape - десятичный логарифм параметра формы. Аргументтом logpdf является число 10 в степени log_shape.
def likehood(sample, log_shape):
    return np.sum(weibull_min.logpdf(sample, 10 ** log_shape))
```

а) оценка по всей выборке

In [32]:

```
# Ищем максимум логарифмической функции правдоподобия по shape в логарифмическом масштабе,
 перебирая его от -2 до 2 с шагом 0.001.
# Выводится значение параметра формы, на котором достигается максимум.
left_bound = -2
right_bound = 2
step = 0.001
estimation = 10 ** (left_bound + step * (np.argmax(np.array([likehood(sample, lo
g_shape)
                                                    for log_shape in np.arange(left_b
ound, right_bound, step)]))))
```

In [33]:

```
print(estimation)
```

```
0.572796030986
```

б) оценка по первым четырем годам

In [34]:

```
# Изменяем выборку.
sample = sample[:(365 * 4)]
```

In [36]:

```
# И снова ищем значение параметра сдвига, на котором достигается максимум логарифмической ф
ункции правдоподобия.
left_bound = -2
right_bound = 2
step = 0.001
estimation = 10 ** (left_bound + step * (np.argmax(np.array([likehood(sample, lo
g_shape)
                                                    for log_shape in np.arange(left_b
ound, right_bound, step)]))))
```

In [37]:

```
print(estimation)
```

```
0.572796030986
```

Вывод

Мы оценили параметр формы в распределении Вейбулла методом правдоподобия по всей выборке и по ее первым 1460 значениям. В обоих случаях оценка параметра сдвига оказалась равной $\gamma^* = 0,572796030986$.