Кафедра дискретной математики МФТИ

Курс математической статистики

Игашов Илья, 593 групппа

Задание №3

№2.

На высоте 1 метр от поверхности Земли закреплено устройство, которое периодически излучает лучи на поверхность Земли (считайте, что поверхность Земли представляет из себя прямую). Пусть l — перпендикуляр к поверхности Земли, опущенный из точки, в которой закреплено устройство. Угол к прямой l (под которым происходит излучение) устройство выбирает случайно из равномерного распределения на отрезке $(-\pi/2, \pi/2)$ (все выборы осуществляются независимо). Можно доказать, что в этих предположениях точки пересечения с поверхностью имеют распределение Коши (плотность равна $\frac{\theta}{\pi(\theta^2+(x-x_0)^2)}$) с параметром масштаба $\theta = 1$. Неизвестный параметр сдвига x_0 соответствует проекции (вдоль прямой l) точки расположения устройства на поверхность Земли (направление оси и начало координат на поверхности Земли выбраны заранее некоторым образом независимо от расположения устройства). В файле Cauchy.csv находятся координаты точек пересечения лучей с поверхностью Земли. Оцените параметр сдвига методом максимального правдоподобия а) по половине выборки (первые 500 элементов выборки, т.е. выборка состоит из 1000 наблюдений); б) по всей выборке. Оценку произведите по сетке (т.е. возьмите набор точек с некоторым шагом и верните ту, на которой достигается максимум функции правдоподобия). Известно, что параметр масштаба принадлежит интервалу $[-1000, 1000]$. Выберите шаг равным 0.01. Если получается долго или не хватает памяти, то уменьшите интервал поиска и поясните (в комментариях), почему берете именно такой интервал.

In [1]:

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

%matplotlib inline
```

In [2]:

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

In [3]:

```
# Логарифм функции правдоподобия (сумма логарифмов значений плотности на каждом значении из
 sample).
def likehood(sample, loc_):
    return np.sum(cauchy.logpdf(sample, loc=loc_, scale=1))
```

а) оценка по всей выборке

In [4]:

```
# Ищем максимум логарифмической функции правдоподобия по loc_, перебирая его от -1000 до
 1000 с шагом 0.01.
# Выводится значение параметра сдвига, на котором достигается максимум.
left_bound = -1000
right_bound = 1000
step = 0.01
estimation = left_bound + step * (np.argmax(np.array([likehood(sample, loc_)
                                                 for loc_ in np.arange(left_bound,
 right_bound, step)])))
```

In [5]:

```
print(estimation)
```

```
-208.04
```

б) оценка по половине выборки

In [6]:

```
# Изменяем выборку.
sample = np.array(sample[:500])
```

In [7]:

```
# И снова ищем значение параметра сдвига, на котором достигается максимум логарифмической ф
ункции правдоподобия.
left_bound = -1000
right_bound = 1000
step = 0.01
estimation = left_bound + step * (np.argmax(np.array([likehood(sample, loc_)
                                                 for loc_ in np.arange(left_bound,
 right_bound, step)])))
```

In [8]:

```
print(estimation)
```

```
-208.01
```

Вывод

Мы оценили параметр сдвига в распределении Коши методом правдоподобия по всей выборке и по ее половине. В первом случае оценка параметра $x_0 = -208.04$, во втором — $x_0 = 208.01$. Как мы видим, результаты хоть и различаются, но не намного. Это логично: размеры выборок различны, что дает различие в результате; при этом все значения выборки получены при одинаковых условиях и при конкретном расположении устройства, поэтому все точки пересечения лучей с поверхностью (как из целой выборки, так и из ее половины) подчиняются распределению Коши с одним и тем же параметром сдвига — это дает не очень сильное различие в оценках.