

データマイニング工学

第3回レポート

濱崎 直紀
(学籍番号 : 28G19096)

令和2年2月3日

問題 1

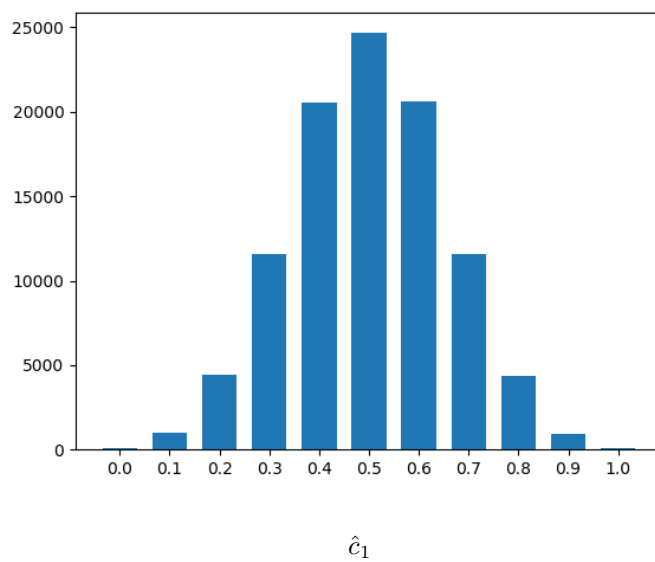
問題 3 で示す.

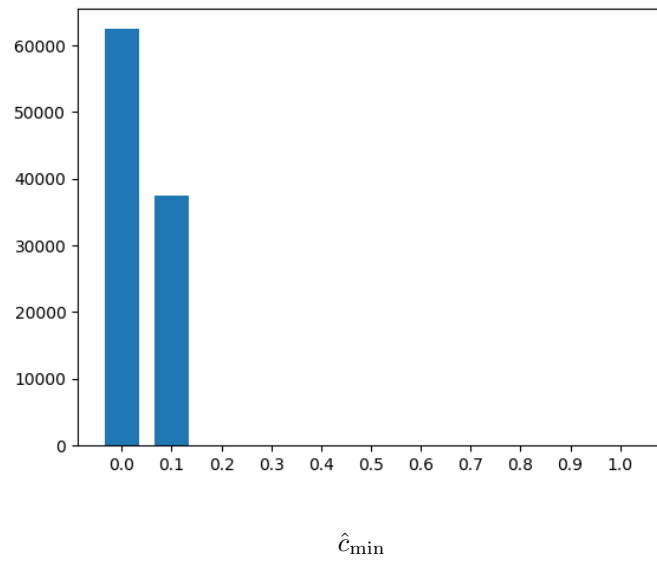
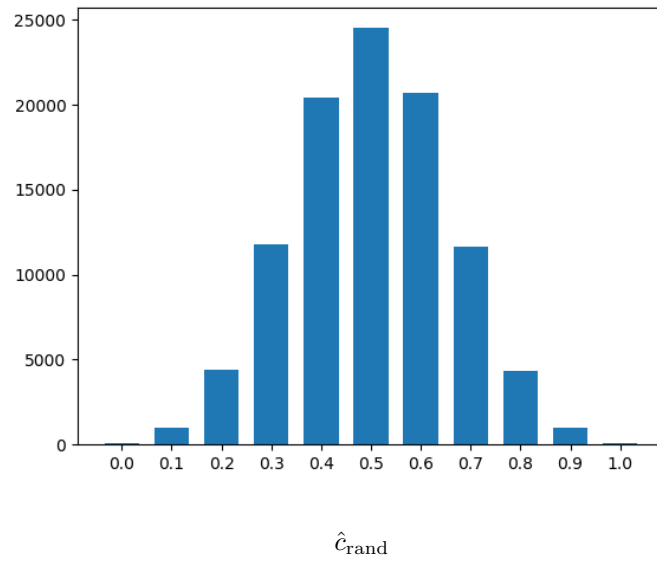
問題 2

問題 3 で示す.

問題 3

\hat{c}_1 , \hat{c}_{rand} , \hat{c}_{min} の分布をヒストグラムで表すと, それぞれ以下の図のようになる.

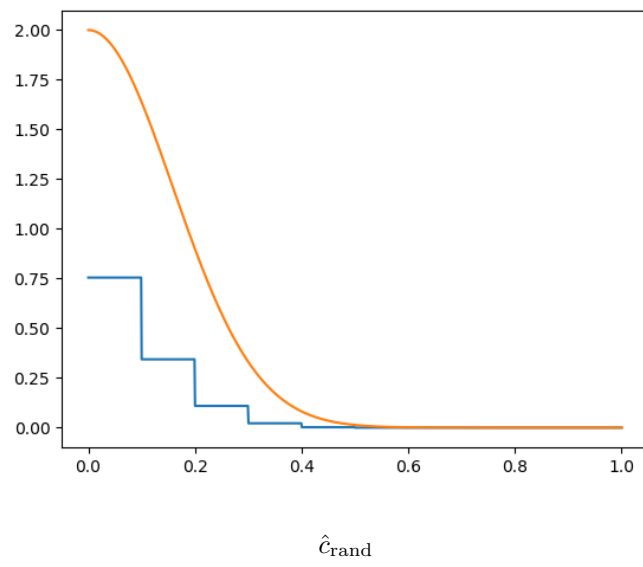
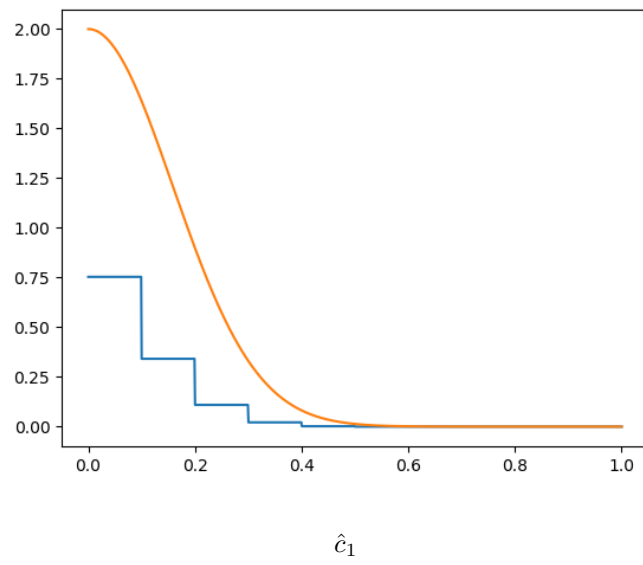


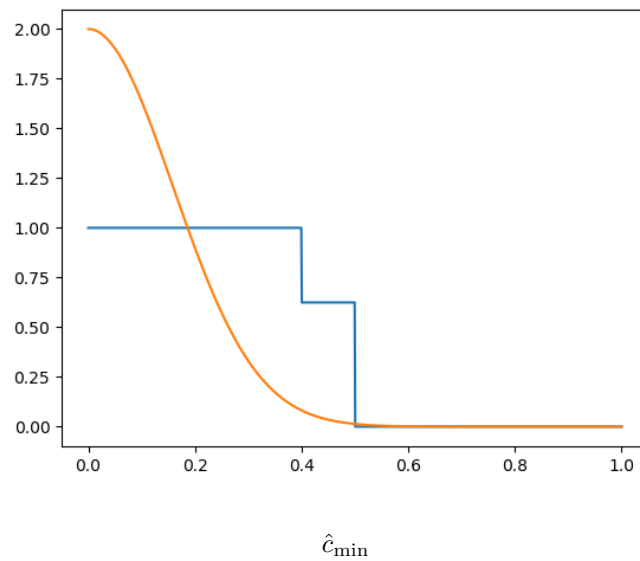


問題 4

A. (B) で示す.

B. $\hat{c}_1, \hat{c}_{\text{rand}}, \hat{c}_{\text{min}}$ の 3 種類のコインにおいて, $\mathbf{P}\{|\hat{c} - \mathbf{E}c| > \varepsilon\}$ を近似し, ε の関数としてそのグラフを描くと, 以下の図のようになる. 尚, 同じプロットにおいて, Hoeffding の上界 ($2e^{-2\varepsilon^2 n}$) をオレンジの線で表している.

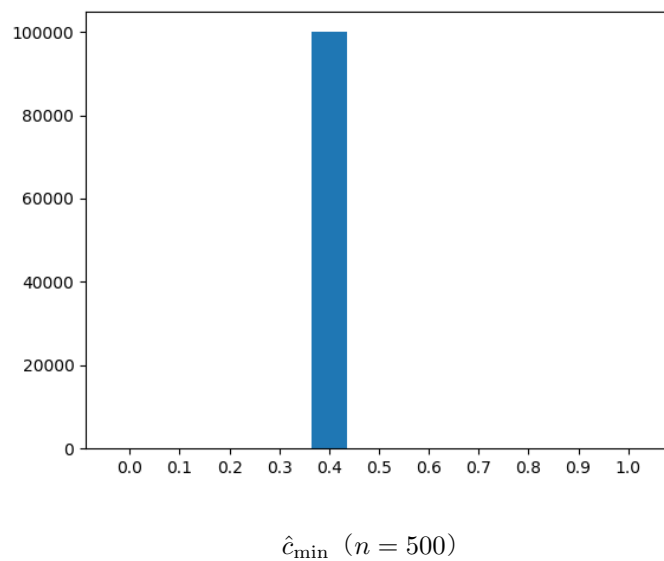




C. c_1 と c_{rand} は Hoeffding の不等式に従うが, c_{min} は従わない. これは, c_{min} が 1000 枚のコインの中の任意のコインではなく, 表の数が最も少なかった特定のコインであるからである.

問題 5

標本数 n (コインを投げる回数) を 500 としたときの \hat{c}_{min} の分布のヒストグラムは以下ようになる.



標本数が 10 である問題 3 の \hat{c}_{min} では, 分布が $\hat{c} = 0.0, 0.1$ (コインが表である割合) に集まっているのに

対して，標本数が 500 のときは分布が $\hat{c} = 0.4$ に集まっている．標本数が 10 のように少ない場合は，1000 枚のコインの中に期待値 $\hat{c} = 0.5$ から大きく外れた \hat{c} の非常に小さいコインが紛れている可能性は非常に高い．しかし，標本数が 500 のように多い場合は，確率が収束し， \hat{c} が最も小さいコインでも期待値から大きく外れず，分布が $\hat{c} = 0.4$ 辺りに集まっていると考えられる．

プログラム

ソースコード 1: 数値シミュレーションプログラム

```
1  import matplotlib.pyplot as plt
2  import numpy as np
3  import os
4  from tqdm import tqdm
5  os.chdir(os.path.dirname(os.path.abspath(__file__)))
6
7
8  # コインを投げてc_1, c_rand, c_minを求める
9  # trial : 試行回数
10 # t_num : コイン 1枚に対する試行回数
11 # c_num : コインの枚数
12 def coin(trial, t_num, c_num):
13     c_1 = []
14     c_rand = []
15     c_min = []
16
17     print('トス開始')
18     for _ in tqdm(range(trial)):
19         front = np.random.binomial(t_num, 0.5, size=c_num)
20         f_rate = front / t_num
21         c_1.append(f_rate[0])
22         rnd = np.random.randint(0, c_num)
23         c_rand.append(f_rate[rnd])
24         c_min.append(min(f_rate))
25     print('トス終了')
26
27     return c_1, c_rand, c_min
28
29
30 # グラフの計算
31 def prob(data):
32     elem = 1000
33     data_num = len(data)
34
35     x = np.linspace(0, 1, elem)
36     y = []
37
38     print('グラフ計算開始')
39     for i in tqdm(x):
40         num = 0
```

```

41         for d in data:
42             if abs(d - 0.5) > i:
43                 num += 1
44             rate = num / data_num
45             y.append(rate)
46         print('グラフ計算終了')
47
48     return x, y
49
50
51 # Hoeffding
52 def hoeffding():
53     n = 10
54     elem = 1000
55
56     x = np.linspace(0, 1, elem)
57     y = []
58
59     print('Hoeffding の計算開始')
60     for i in tqdm(x):
61         ans = 2*np.e**(-2*i**2*n)
62         y.append(ans)
63     print('Hoeffding の計算終了')
64
65     return x, y
66
67
68 if __name__ == '__main__':
69     os.makedirs('./hist', exist_ok=True)
70
71     # 課題 3
72     c_1, c_rand, c_min = coin(100000, 10, 1000)
73
74     fig = plt.figure(1)
75     plt.hist(c_1, bins=11, range=(-0.05, 1.05), rwidth=0.7)
76     plt.xticks([0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0])
77     fig.savefig('./hist/c_1.png')
78     fig = plt.figure(2)
79     plt.hist(c_rand, bins=11, range=(-0.05, 1.05), rwidth=0.7)
80     plt.xticks([0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0])
81     fig.savefig('./hist/c_rand.png')
82     fig = plt.figure(3)
83     plt.hist(c_min, bins=11, range=(-0.05, 1.05), rwidth=0.7)
84     plt.xticks([0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0])
85     fig.savefig('./hist/c_min.png')
86

```



```
87     # 課題 4
88     a, b = hoeffding()
89     x, y = prob(c_1)
90     fig = plt.figure(4)
91     plt.plot(x, y)
92     plt.plot(a, b)
93     fig.savefig('./hist/Hoeffding_c1.png')
94     x, y = prob(c_rand)
95     fig = plt.figure(5)
96     plt.plot(x, y)
97     plt.plot(a, b)
98     fig.savefig('./hist/Hoeffding_crand.png')
99     x, y = prob(c_min)
100    fig = plt.figure(6)
101    plt.plot(x, y)
102    plt.plot(a, b)
103    fig.savefig('./hist/Hoeffding_cmin.png')
104
105    # 課題 5
106    c_1_500, c_rand_500, c_min_500 = coin(100000, 500, 1000)
107    fig = plt.figure(7)
108    plt.hist(c_min_500, bins=11, range=(-0.05, 1.05), rwidth=0.7)
109    plt.xticks([0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0])
110    fig.savefig('./hist/c_min_500.png')
```
