

経験リスク

名前:松島完忠
学籍番号:T211d070
日付:7/25

[演習 250] 閾値関数の経験リスク

演習 240 で述べた設定 1 の同時分布 \mathbf{D}_Z を用いる。判別関数は閾値関数とし、閾値は $\theta = 0$ とおく。標本サイズ $m \in \{10, 100, 1000, 10000\}$ それぞれに対する標本 $D \sim Z_Z^m$ の経験リスクを m 回繰り返し計算し、その頻度分布をプロットした結果を図 1、図 2、図 3、図 4 に示す。

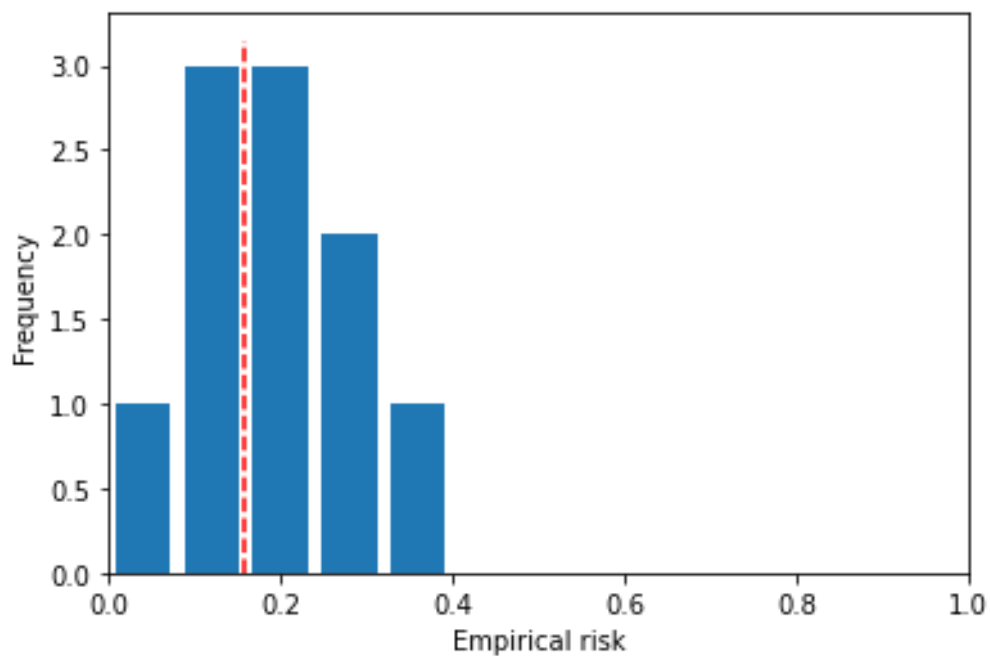


図 1: $m=10$ の経験リスク

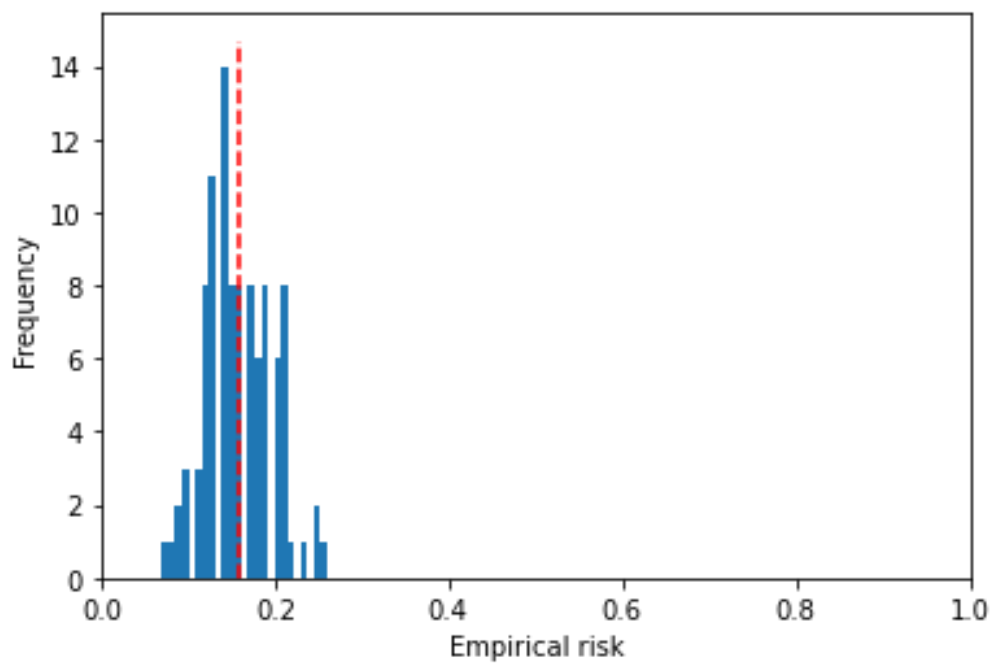


図 2: $m=100$ の経験リスク

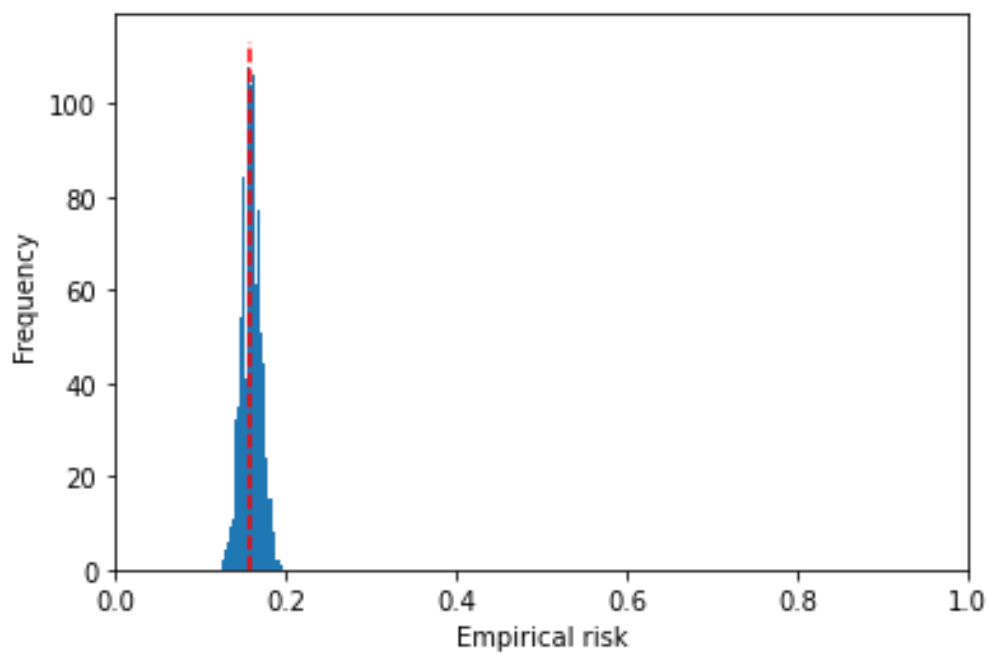


図 3: $m=1000$ の経験リスク

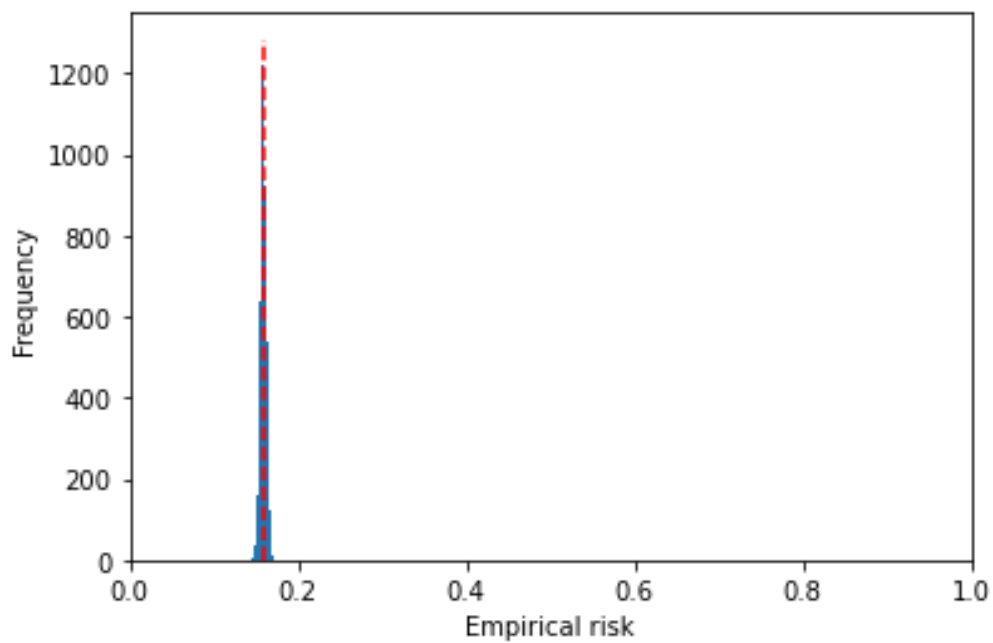


図 4: $m=10000$ の経験リスク

各 m の値のときの頻度分布を観察することから m を増やすほど、経験リスクは汎化リスクに漸近していることがわかる。

作成したプログラム

図 5 に演習 250 にて作成したソースコードを示す。

1	<code>import numpy as np</code>
2	<code>from scipy import stats</code>
3	<code>import matplotlib.pyplot as plt</code>
4	<code>import scipy.stats</code>
5	<code>import matplotlib.ticker as ticker</code>
6	

7	def Genrisk(a, b, theta, N):
8	sum=0
9	
10	for (x, y) in zip(a, b):
11	f=x-theta
12	if (y*f<=0):
13	sum+=1
14	return float(sum)/float(N)
15	
16	def empricalrisk(sigma, M):
17	Y=[]
18	x=[]
19	empricalrisk=[]
20	y=stats.uniform.rvs(0, scale=1, size=M)
21	for s in y:
22	if s<0.5:
23	t=stats.norm.rvs(loc=-1, scale=sigma)
24	x.append(t)
25	Y.append(-1.0)
26	else:
27	t=stats.norm.rvs(loc=1, scale=sigma)
28	x.append(t)
29	Y.append(1.0)
30	return Genrisk(x, Y, 0, M)

31	
32	size = 10
33	sigma = 1
34	
35	Empricalrisk=[]
36	
37	genrisk = scipy.stats.norm.cdf(x=-1, loc=0, scale=1)
38	
39	for th in range(size):
40	Empricalrisk.append(empricalrisk(sigma, size))
41	
42	fig=plt.figure()
43	ax2=fig.add_subplot(111,xlabel='Empirical risk',ylabel='Frequency',xli m=(0,1))
44	ax2.hist(Empricalrisk,bins=5,rwidth=0.8)
45	y_min,y_max=ax2.get_ylim()
46	plt.vlines(genrisk,0,y_max,colors="red",linestyle="dashed")

図 5:作成したプログラム