COGSCI131–Spring 2019 — Homework 8Solutions

Yuchen Zhou, SID 3034489901

1.

- Namely, to compute likelihood is to compute function $P(D|H)$.
- Likelihood of each data point in hypothesis is $\frac{1}{len(H)}$.
- Likelihood of data points that are not in hypothesis is 0.
- Code as follows:

```
def PDH(x,h):
    p=[0]*len(x)
    pp=1
    for i in range(len(x)):
        for j in range(len(h)):
            if x[i]==h[j]:
                p[i]=1/len(h)
                break
        pp*=p[i]
    if pp==1:
        return 0.0
    else:
        return pp
```

2.

- The general formula are:

$$P(j \in C|D) = \sum_k \{P(j \in C|H_k) \times P(H_k|D)\}$$

$$P(H|D) = \frac{P(D|H) \times P(H)}{\sum_i \{P(D|H_i) \times P(H_i)\}}$$

- 1. No data. In formula, $P(j \in C|H_k)$ are always 0, so all number from 1 through 100 are not in concept.
  2. Data sets: $\{50\}$. Data satisfies hypothesis 1,5,6. By intuition, right answers are tens and multiples of 5.

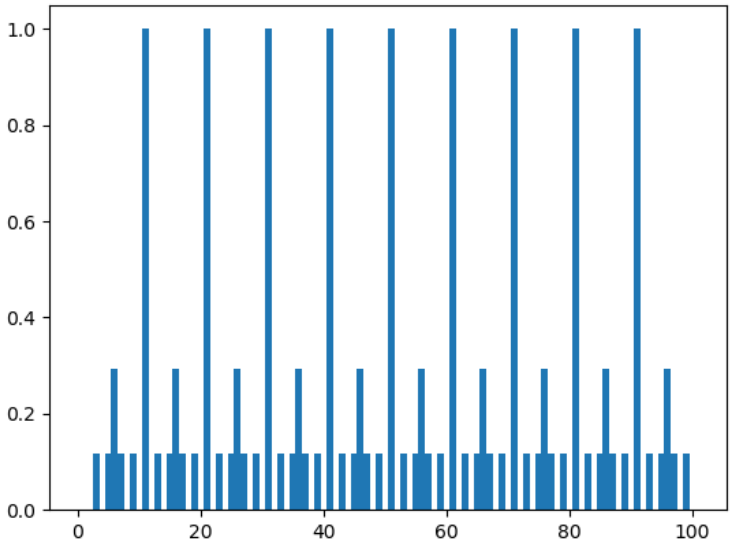

Figure 1: data sets: $\{50\}$

3. Data sets: $\{53\}$. Data satisfies hypothesis 2,4. No distinct intuition about right answer.

4. Data sets: $\{50, 53\}$. Data satisfies no hypothesis, so all number from 1 to 100 are not in the concept. No distinct intuition about right answer.
5. Data sets: $\{16\}$. Data satisfies hypothesis 1,3. By intuition, right answer are square numbers.
6. Data sets: $\{10, 20\}$. Data satisfies hypothesis 1,5,6. By intuition, right answers are tens.

7. Data sets: $\{2, 4, 8\}$. Data satisfies hypothesis 1. By intuition, right answers are even numbers.
8. Data sets: $\{2, 4, 8, 10\}$. Data satisfies hypothesis 1. By intuition, right answers are even numbers.

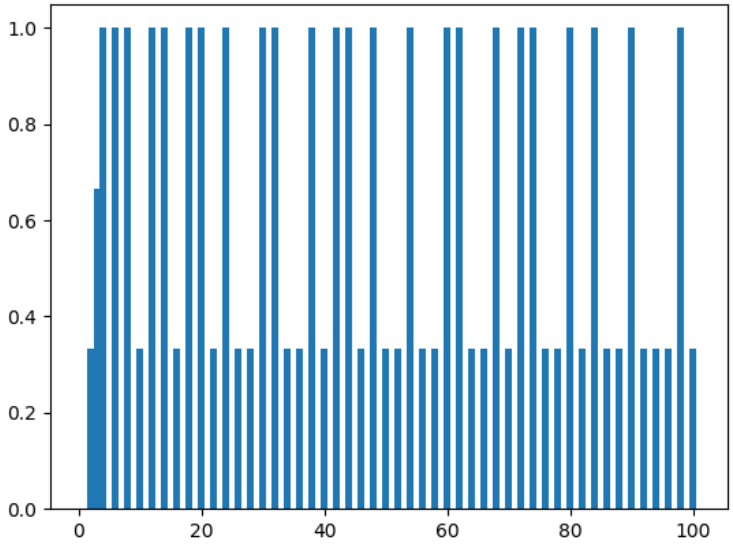

Figure 2: data sets: {53}

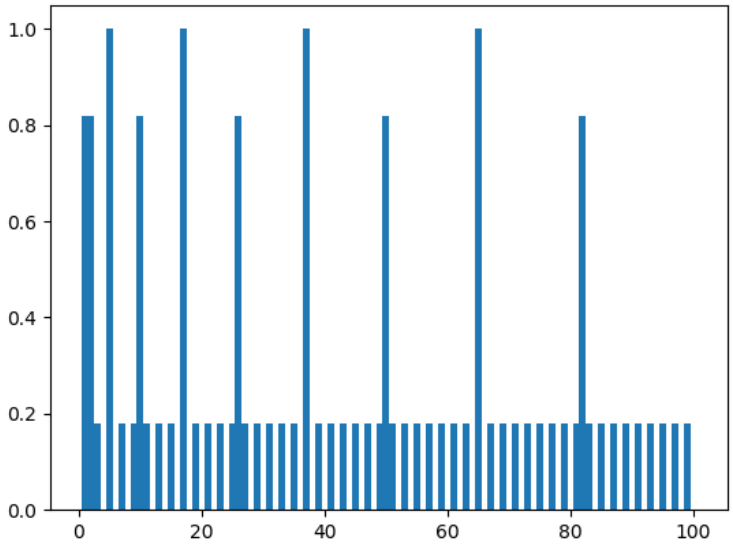

Figure 3: data sets: {16}

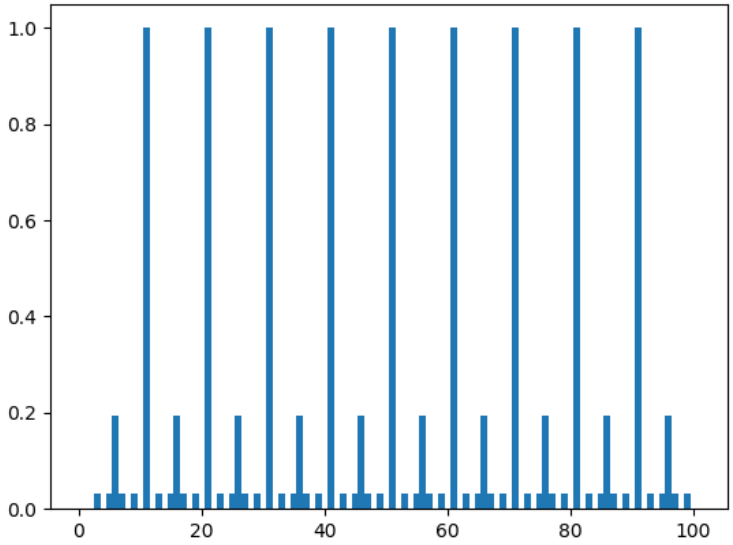

Figure 4: data sets: {10,20}

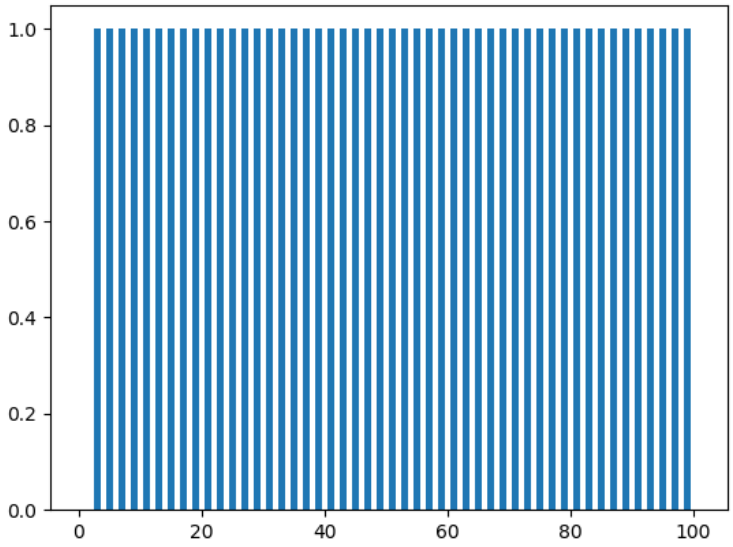

Figure 5: data sets: {2,4,8}

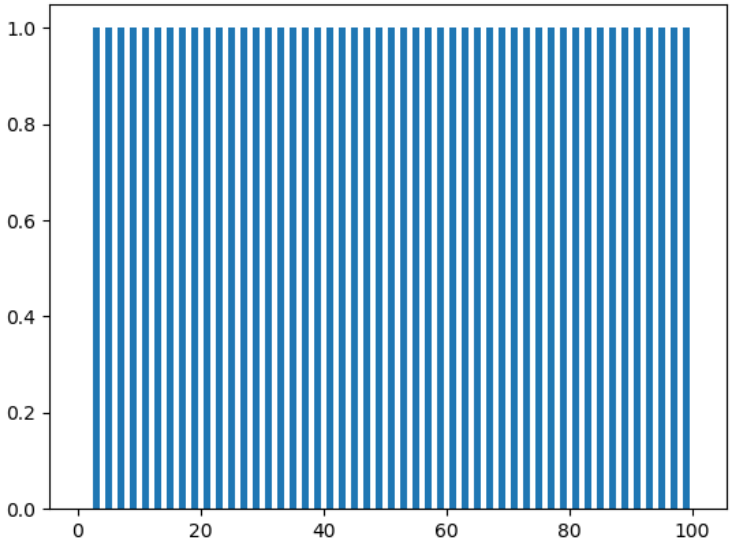

Figure 6: data sets: {2,4,8,10}

- Code as follows:

```
import numpy as np
import matplotlib.pyplot as plt

N=6
H1=[2*i+2 for i in range(50)]
H2=[2*i+1 for i in range(50)]
H3=[]
i3=0
while i3**2<=100:
    H3.append(i3**2)
    i3+=1
H4=[]
i4=0
for i in range(2,100):
    for j in range(2,i):
        if(i%j==0):
            break
    else:
        H4.append(i)
H5=[i*5 for i in range(1,21)]
H6=[i*10 for i in range(1,11)]

def PDH(x,h):
    p=[0]*len(x)
    pp=1
    for i in range(len(x)):
        for j in range(len(h)):
            if x[i]==h[j]:
                p[i]=1/len(h)
                break
        pp*=p[i]
    if pp==1:
        return 0.0
    else:
        return pp

H=[]
H.append(H1)
H.append(H2)
H.append(H3)
H.append(H4)
H.append(H5)
H.append(H6)

prior=[1/6]*N

data=[10,20]
y=np.zeros(100)
PHD=np.zeros(N)
sump=0

for j in range(N):
    sump += PDH(data, H[j]) * prior[j]
for i in range(N):
    if sump==0:
        PHD[i]=0
    else:
```

```
        PHD[i]=PDH(data,H[i])*prior[i]/sump

for i in range(100):
    for j in range(N):
        for k in range(len(H[j])):
            if i==H[j][k]:
                y[i]+=PHD[j]
                break
fig = plt.figure()
ax=fig.add_subplot(111)
#ax.plot(y,color="lightblue",linewidth=2)
plt.bar([i+1 for i in range(100)],y,width=1.2)
plt.show()
```

3.

- There are 5050 intervals in range 1 through 100.
- 1. No data. In formula, $P(j \in C|H_k)$ are always 0, so all number from 1 through 100 are not in concept.
  2. Data sets: $\{50\}$. By intuition, right answers are tens and multiples of 5. The modified hypothesis reveals that number closer to 50 are more likely to be in concept.

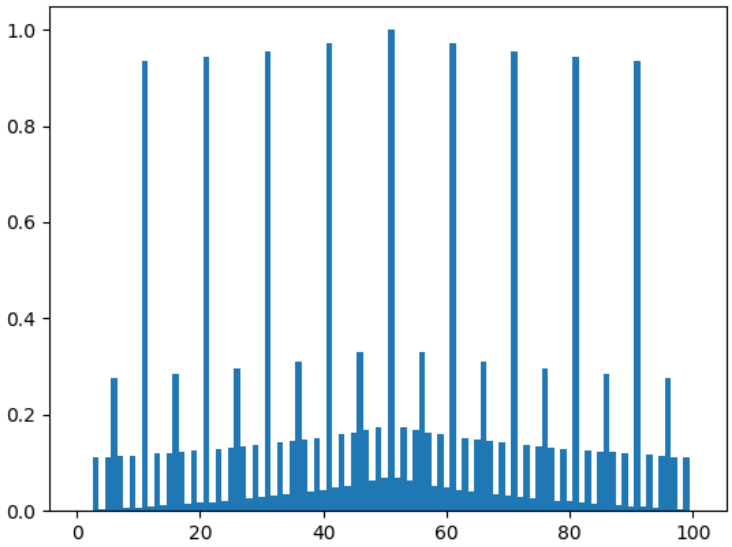

Figure 7: data sets: $\{50\}$

3. Data sets: $\{53\}$.No distinct intuition about right answer, also.
4. Data sets: $\{50, 53\}$. Data satisfies no hypothesis in 1-6. Intuition is that more closer, more likely. This model better match my thought.
5. Data sets: $\{16\}$. By intuition, right answer are square numbers. Modified model makes little improvement.
6. Data sets: $\{10, 20\}$.By intuition, right answers are tens. Modified model is very similar to unmodified one.
7. Data sets: $\{2, 4, 8\}$.By intuition, right answers are small even numbers. New modified model better match my intuition.
8. Data sets: $\{2, 4, 8, 10\}$. By intuition, right answers are small even numbers. New modified model better match my intuition.

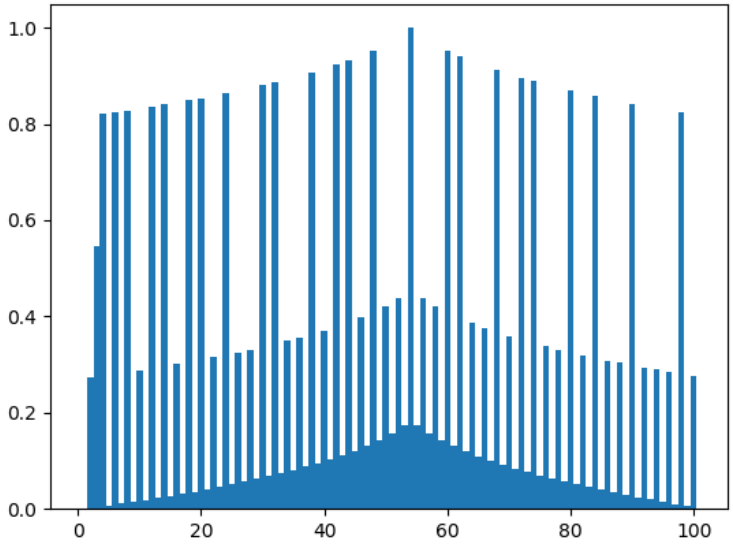

Figure 8: data sets: {53}

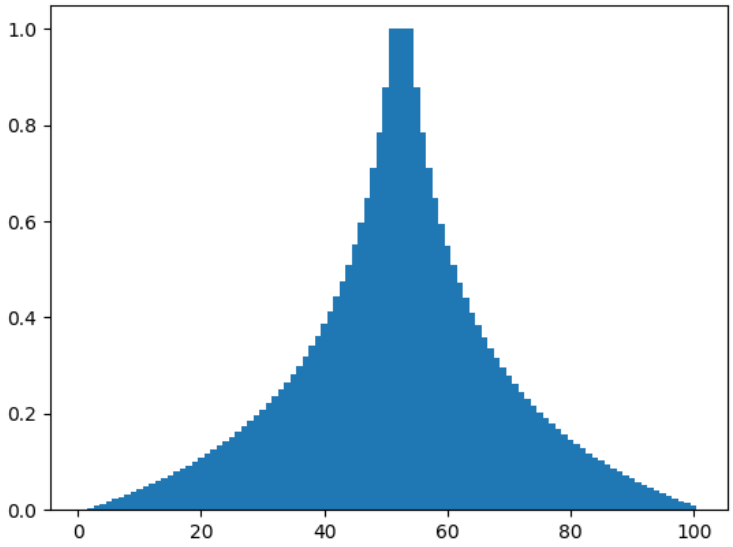

Figure 9: data sets: {50,53}

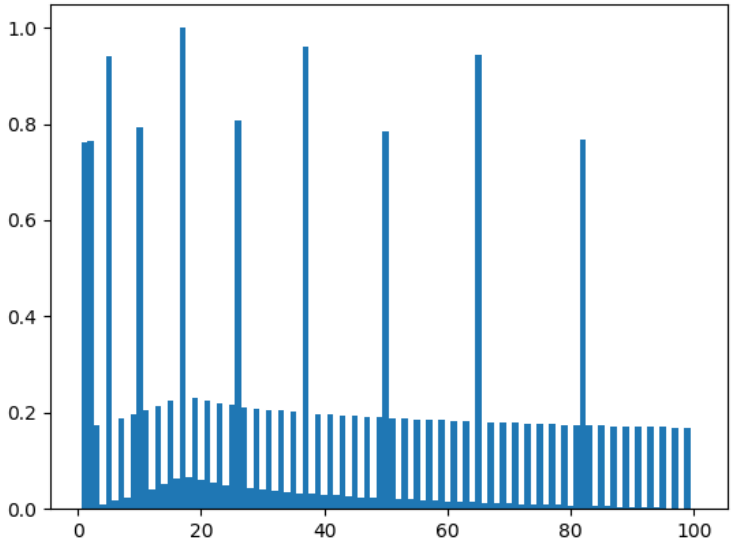

Figure 10: data sets: {16}

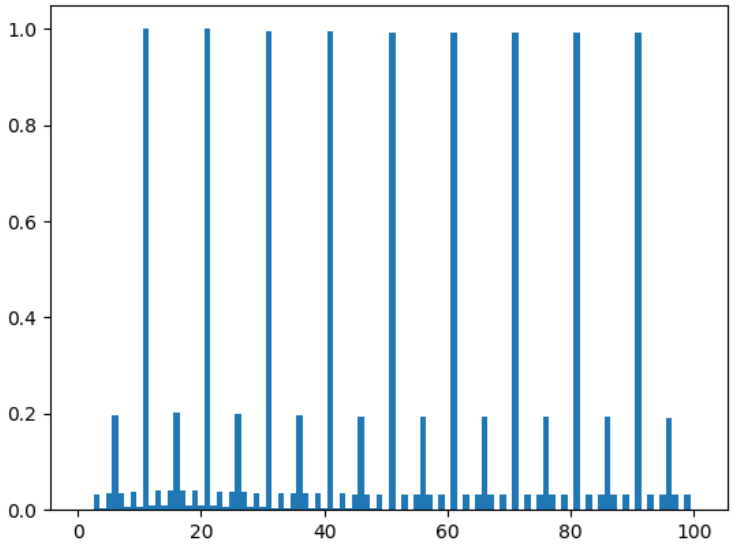

Figure 11: data sets: {10,20}

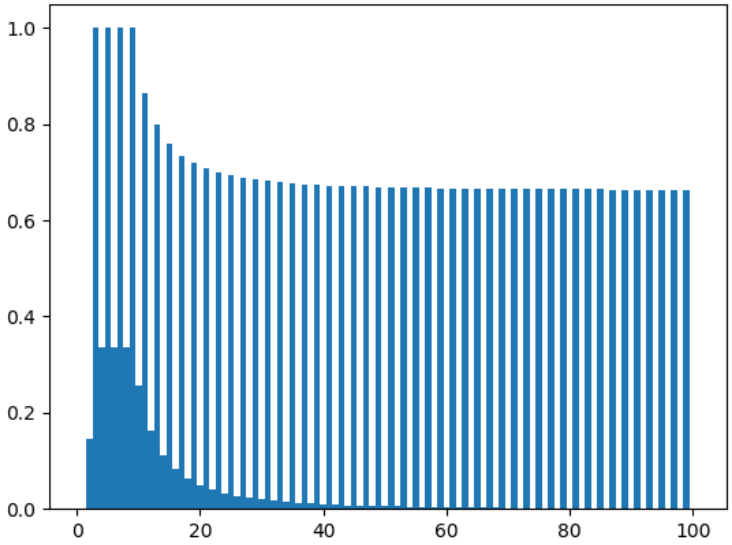

Figure 12: data sets: {2,4,8}

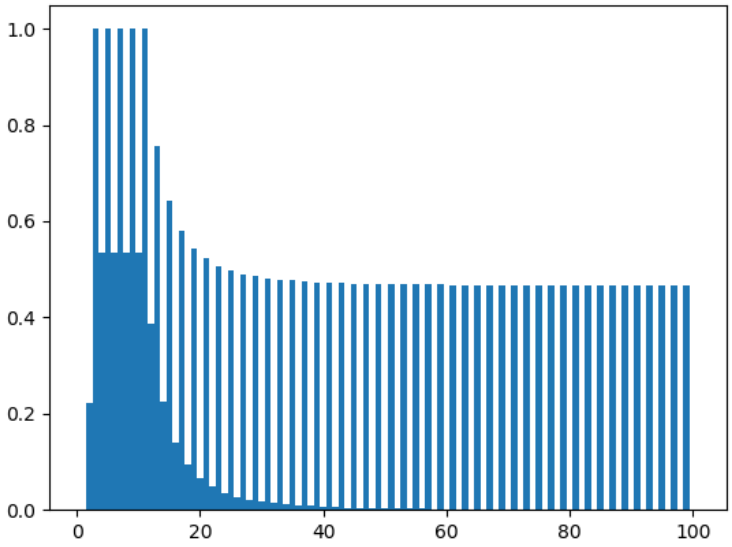

Figure 13: data sets: {2,4,8,10}

- Codes as follows:

```
import numpy as np
import matplotlib.pyplot as plt

N=6
H1=[2*i+2 for i in range(50)]
H2=[2*i+1 for i in range(50)]
H3=[]
i3=0
while i3**2<=100:
    H3.append(i3**2)
    i3+=1
H4=[]
i4=0
for i in range(2,100):
    for j in range(2,i):
        if(i%j==0):
            break
    else:
        H4.append(i)
H5=[i*5 for i in range(1,21)]
H6=[i*10 for i in range(1,11)]

H=[]
H.append(H1)
H.append(H2)
H.append(H3)
H.append(H4)
H.append(H5)
H.append(H6)

NN=0
for gap in range(1,101):
    st = 1
    while st+gap<=101:
        NN+=1
        H.append([i+st for i in range(gap)])
        st+=1

def PDH(x,h):
    p=[0]*len(x)
    pp=1
    for i in range(len(x)):
        for j in range(len(h)):
            if x[i]==h[j]:
                p[i]=1/len(h)
                break
        pp*=p[i]
    if pp==1:
        return 0.0
    else:
        return pp

prior=np.zeros(N+NN)
for i in range(6):
    prior[i]=1/7
for i in range(NN):
    prior[6+i]=1/7/NN
```

```
data=[2,4,8,10]
y=np.zeros(100)
PHD=np.zeros(N+NN)
sump=0

for j in range(N+NN):
    sump += PDH(data, H[j]) * prior[j]
for i in range(N+NN):
    if sump==0:
        PHD[i]=0
    else:
        PHD[i]=PDH(data,H[i])*prior[i]/sump

for i in range(100):
    for j in range(N+NN):
        for k in range(len(H[j])):
            if i==H[j][k]:
                y[i]+=PHD[j]
                break
fig = plt.figure()
ax=fig.add_subplot(111)
#ax.plot(y,color="lightblue",linewidth=2)
plt.bar([i+1 for i in range(100)],y,width=1)
plt.show()
```