In [4]:

```
import os
import sys
sys.path.append("numpy_path")
import numpy as np
from matplotlib import pyplot as plt
from math import cos, sin, pi
import math
from statistics import mean
import os.path
import random

shape_size = 48
```

In [5]:

```
data_size = 0
dummy1 = [0]*200
dummy2 = [0]*6
SP = np.array(np.reshape(dummy1, (1, 200)))
SH = np.array(np.reshape(dummy2, (1, 6)))
for i in range(2, 65):
    path = 'meep_code/data/DATA'+str(i)
    if not os.path.exists(path):
        #miss.append(i)
        print('Missing batch:' + str(i))
        continue

    files = next(os.walk(path))[2] #dir is your directory path as string]
    num_data = len(files)
    data_size += num_data
    skip = []

    coordinates = np.genfromtxt('meep_code/data/DATA'+str(i)+'_sh.txt')
    xc, yc = coordinates[:, 0], coordinates[:, 1]
    xc = np.reshape(xc, (num_data, shape_size))
    yc = np.reshape(yc, (num_data, shape_size))

    for j in range(num_data):
        tmp = np.genfromtxt(path+'/'+'DATA'+str(i)+'_sp'+str(j)+'.txt')
        valid = True
        for q in range(200):
            if tmp[q] > 3:
                print('Batch '+str(i)+'\tsample '+str(j)+' has extreme value'
                valid = False
                break
        if not valid:
            #skip.append(j)
            continue
        SP = np.concatenate((SP, np.reshape(tmp, (1, 200))))
        tmp = []
        for q in range(6):
            tmp.append(math.sqrt(xc[j][q]**2 + yc[j][q]**2))
        SH = np.concatenate((SH, np.reshape(np.array(tmp), (1, 6))))
    print('Batch '+str(i)+' has \t'+str(num_data))
```

```
Batch 44 has    106
Batch 45 has    35
Batch 46 has    100
Batch 47 has    100
Batch 48 has    287
Batch 49 has    13
Batch 50 has    37
Batch 51 has    37
Batch 52 has    106
Batch 53 has    35
Batch 54 has    100
Batch 55 has    100
Batch 56 has    287
Batch 57 has    35
Batch 58 has    100
Batch 59 has    100
Batch 60 has    287
```

```
Batch 61 has    95
Batch 62 has    272
Batch 63 has    272
```

In [10]:

```
num_peak = 2
distribution = []
print('Total # of data: ' + str(len(SP)))
x = np.genfromtxt('meep_code/data/SP_xaxis.txt')
SP_F, SH_F = np.reshape(SP[1], (1, 200)),np.reshape(SH[1], (1, 6))

for i in range(1, len(SP)):
    p_u, p_d = 0, 0
    p_u_pos, p_d_pos = [], []
    for j in range(1, 200):
        if SP[i][j - 1] >= 0.6 >=SP[i][j]:
            p_d += 1
            p_d_pos.append(j / 2 + 200)
        if SP[i][j - 1] <= 0.6 <=SP[i][j]:
            p_u += 1
            p_u_pos.append(j / 2 + 200)
    if p_d == p_u and p_d == num_peak:
        distribution.append(p_d_pos)
        distribution.append(p_u_pos)
        SP_F = np.concatenate((SP_F, np.reshape(SP[i], (1, 200))))
        SH_F = np.concatenate((SH_F, np.reshape(SH[i], (1, 6))))
#         plt.ylim(0, 1.1)
#         plt.plot(x, SP[i])
#         plt.show()
#         print(p_d_pos)
#         print(p_u_pos)
SP_F = SP_F[1:,:]
SH_F = SH_F[1:,:]
```

```
Total # of data: 4989
```

In [11]:
```
def dis_sec_p(begin = 200, end = 300, show_pic = False, peak = [k for k in ra
    dis = []
    for i in range(num_peak):
        dis.append([])
    for i in range(0, len(distribution) - 1, 2):
        first_peak = (distribution[i][0] + distribution[i + 1][0]) / 2
        if begin <= first_peak < end:
            for j in range(num_peak):
                dis[j].append([random.random(), (distribution[i][j] + distrib
            if show_pic:
                plt.ylim(0, 1.1)
                plt.plot(x, SP_F[i//2])
                plt.show()
    dis = np.array(dis)
    if len(dis[0]) == 0:
        print('No first peak in ['+str(begin)+', '+str(end)+')')
        return
    print(dis.shape)
    plt.xlim(200, 300)
    for p in peak:
        plt.scatter(dis[p - 1,:,1],dis[p - 1,:,0])

    print("{0:.4f}".format(len(dis[0])/len(SP_F)*100)
        +'% ('+str(len(dis[0]))+' out of '+str(len(SP_F))
          +') spectrum with ' + str(num_peak) + ' peaks, its first peak lies
```

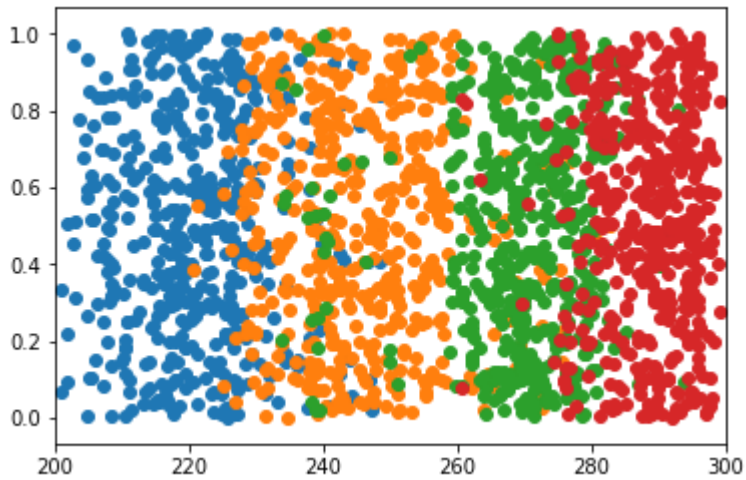
In [12]:
```
dis_sec_p(show_pic = True)
```

. . .

Distribution of fourth peak

In [8]:
```
dis_sec_p()
```

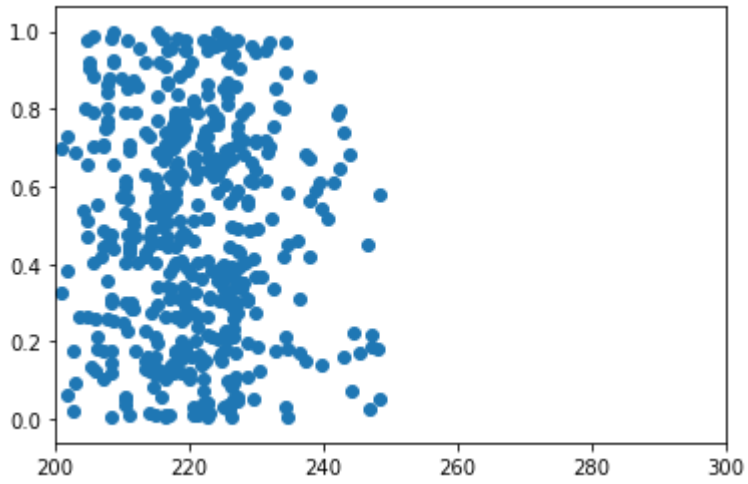
```
(4, 470, 2)
100.0000% (470 out of 470) spectrum with 4 peaks, its first peak lies in [T
rue, 300)
```



First peak distribution

In [196]:
```
dis_sec_p(peak = [1])
```

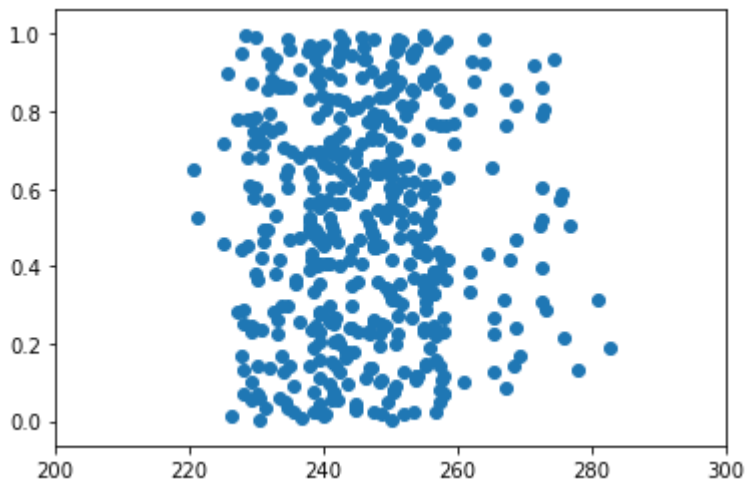
```
(4, 470, 2)
100.0000% (470 out of 470) spectrum with 4 peaks, its first peak lies in [2
00, 300)
```



Second peak distribution

In [197]:
```
dis_sec_p(peak = [2])
```

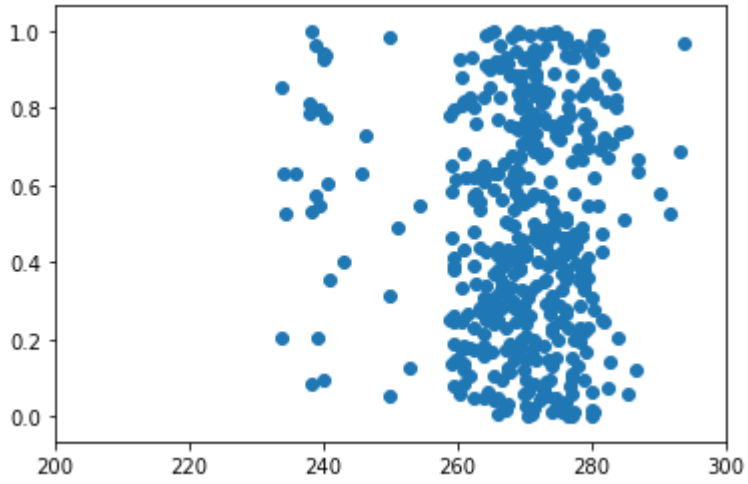
```
(4, 470, 2)
100.0000% (470 out of 470) spectrum with 4 peaks, its first peak lies in [2
00, 300)
```



Third peak distribution

In [198]:
```
dis_sec_p(peak = [3])
```

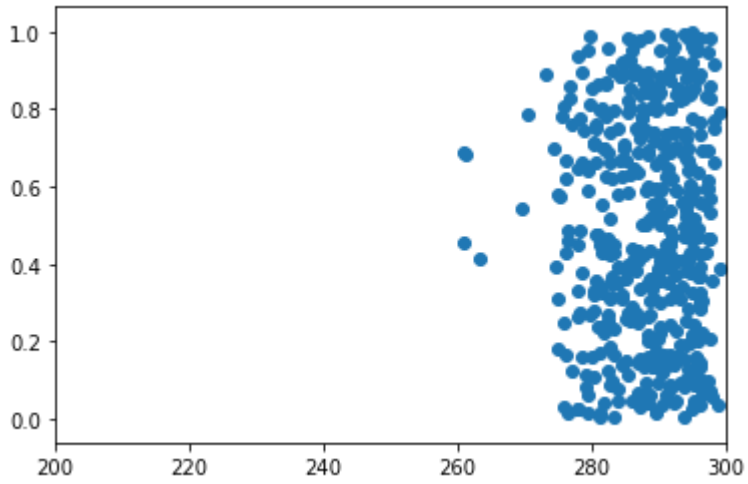
```
(4, 470, 2)
100.0000% (470 out of 470) spectrum with 4 peaks, its first peak lies in [2
00, 300)
```



fourth peak distribution

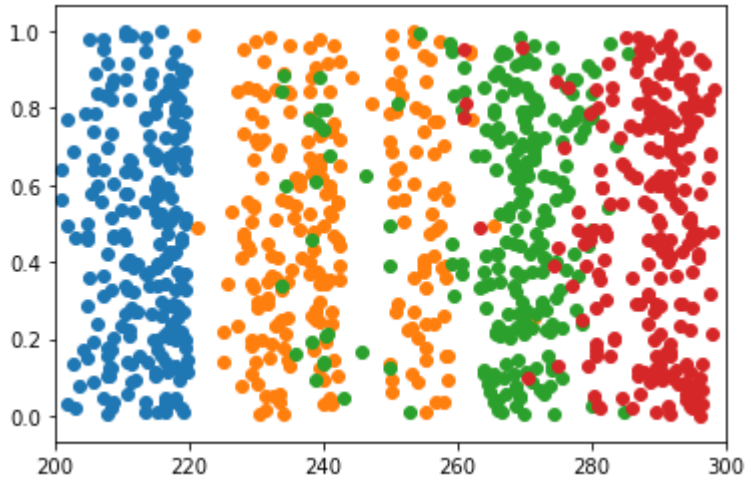
In [199]:
```
dis_sec_p(peak = [4])
```

```
(4, 470, 2)
100.0000% (470 out of 470) spectrum with 4 peaks, its first peak lies in [2
00, 300)
```



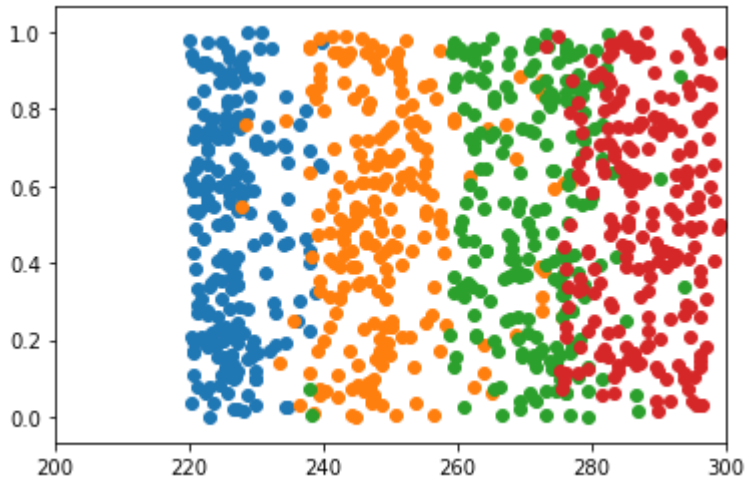
In [166]:
```
dis_sec_p(200,220)
```

```
(4, 233, 2)
49.5745% (233 out of 470) spectrum with 4 peaks, its first peak lies in [20
0, 220)
```



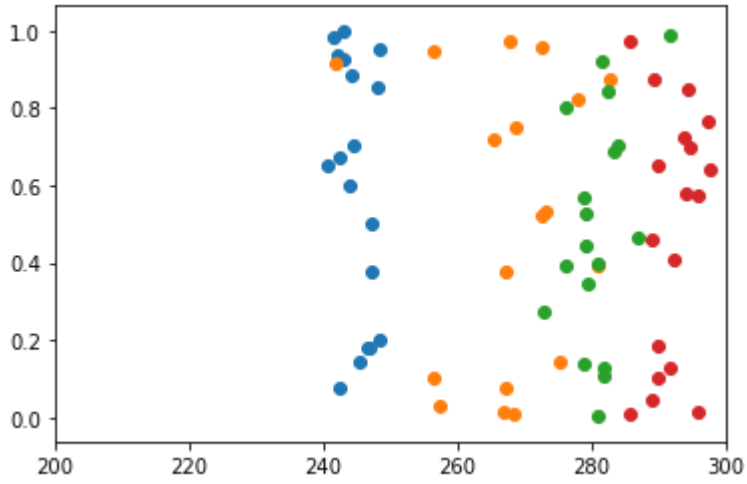
In [167]:
```
dis_sec_p(220,240)
```

```
(4, 219, 2)
46.5957% (219 out of 470) spectrum with 4 peaks, its first peak lies in [22
0, 240)
```



In [168]:
```
dis_sec_p(240,260)
```

```
(4, 18, 2)
3.8298% (18 out of 470) spectrum with 4 peaks, its first peak lies in [240, 260)
```



In [176]:
```
dis_sec_p(260,280)
```

```
No first peak in [260, 280)
```

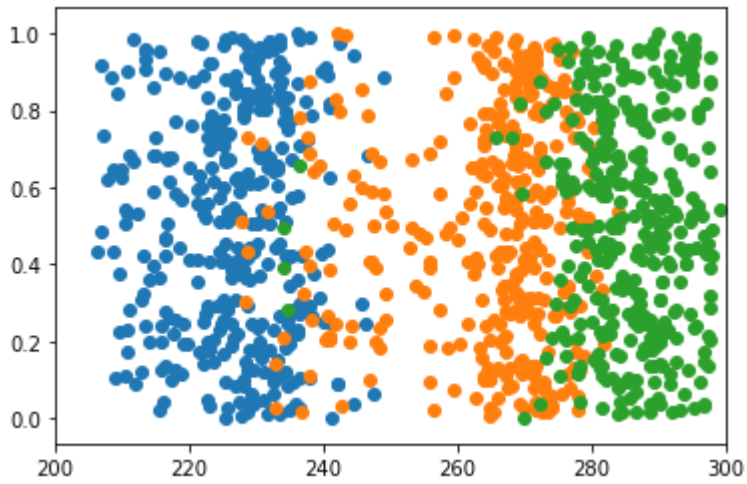
In [177]:
```
dis_sec_p(280,300)
```

```
No first peak in [280, 300)
```

Distribution of Three peak

In [127]:
```
dis_sec_p()
```

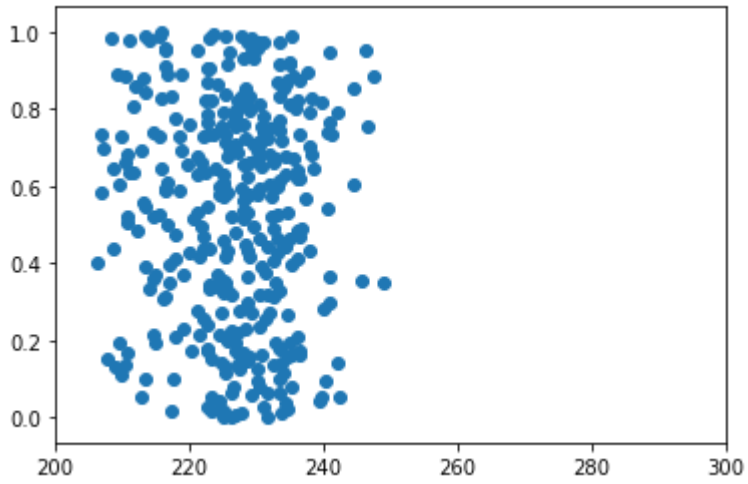
```
(3, 349, 2)
100.0000% (349 out of 349) spectrum with 3peaks, its first peak lies in [200, 300)
```



First peak distribution

In [190]:
```
dis_sec_p(peak = [1])
```

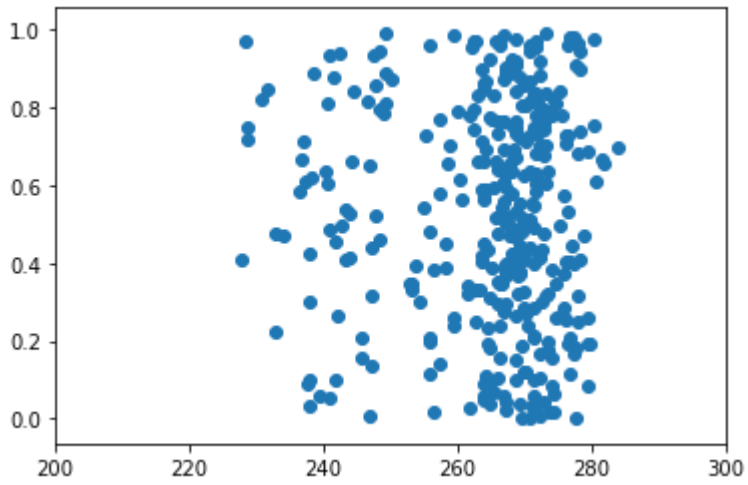
```
(3, 349, 2)
100.0000% (349 out of 349) spectrum with 3 peaks, its first peak lies in [2
00, 300)
```



Second peak distribution

In [192]:
```
dis_sec_p(peak = [2])
```

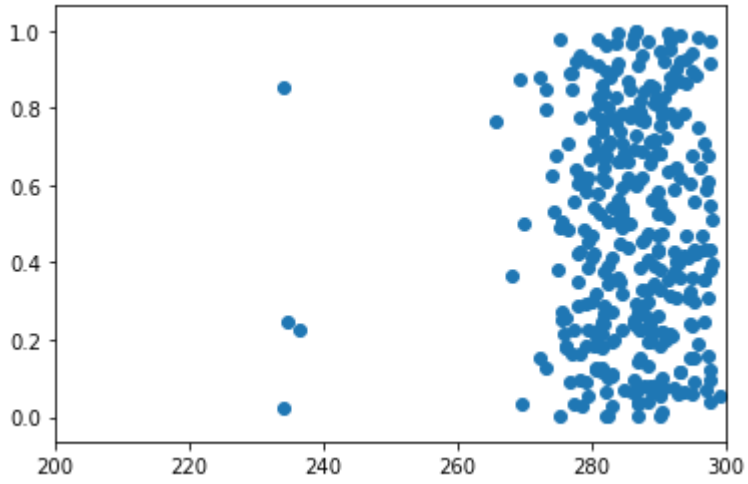
```
(3, 349, 2)
100.0000% (349 out of 349) spectrum with 3 peaks, its first peak lies in [2
00, 300)
```



Third peak distribution

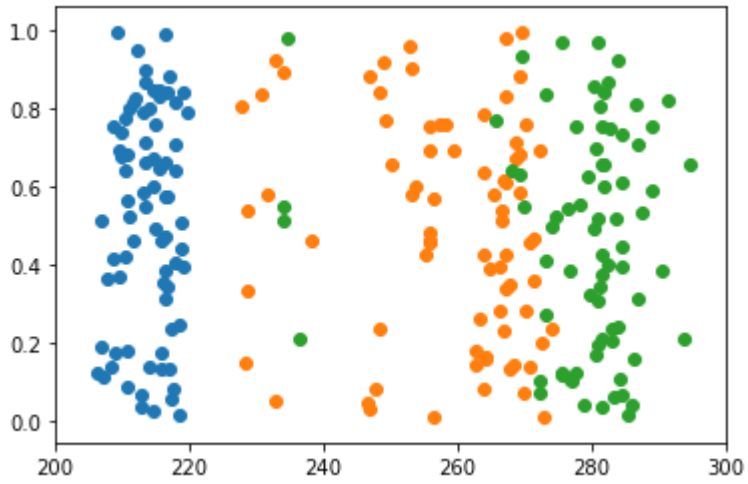
In [193]:
```
dis_sec_p(peak = [3])
```

```
(3, 349, 2)
100.0000% (349 out of 349) spectrum with 3 peaks, its first peak lies in [2
00, 300)
```



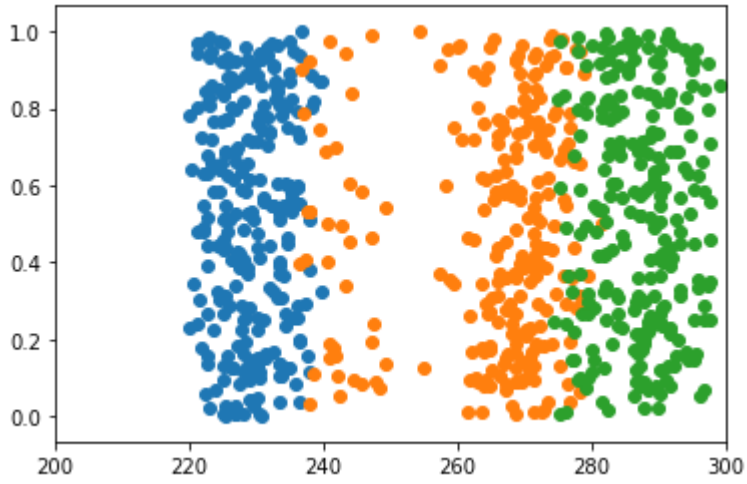
In [183]:
```
dis_sec_p(200,220)
```

```
(3, 76, 2)
21.7765% (76 out of 349) spectrum with 3 peaks, its first peak lies in [20
0, 220)
```



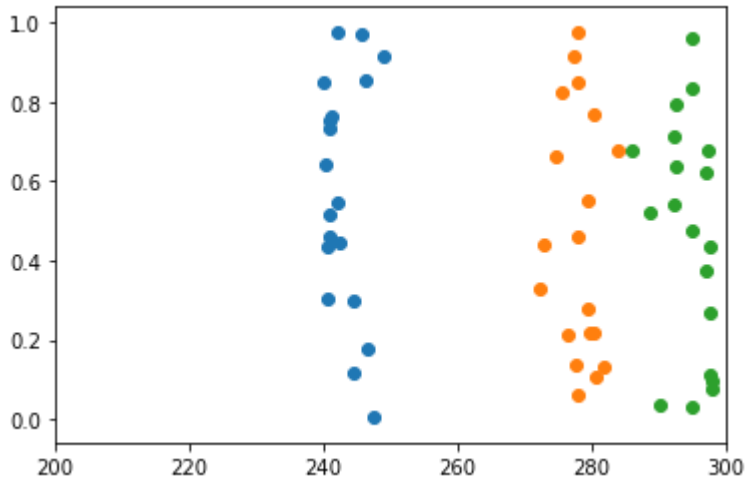
In [184]:
```
dis_sec_p(220,240)
```

```
(3, 254, 2)
72.7794% (254 out of 349) spectrum with 3 peaks, its first peak lies in [22
0, 240)
```



In [185]:
```
dis_sec_p(240,260)
```

```
(3, 19, 2)
5.4441% (19 out of 349) spectrum with 3 peaks, its first peak lies in [240,
260)
```



In [187]:
```
dis_sec_p(260,280)
```

```
No first peak in [260, 280)
```

In [189]:
```
dis_sec_p(280,300)
```

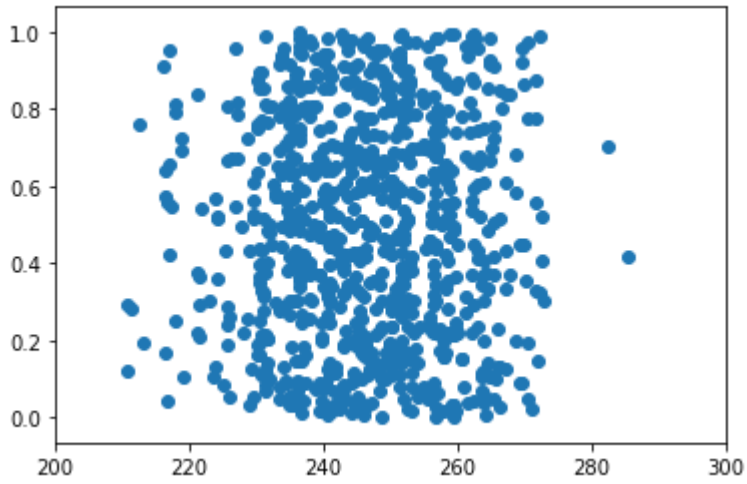
```
No first peak in [280, 300)
```

Distribution of Two peak

Distribution of the first of Two peak

In [152]:
```
dis_sec_p(peak = [1])
```

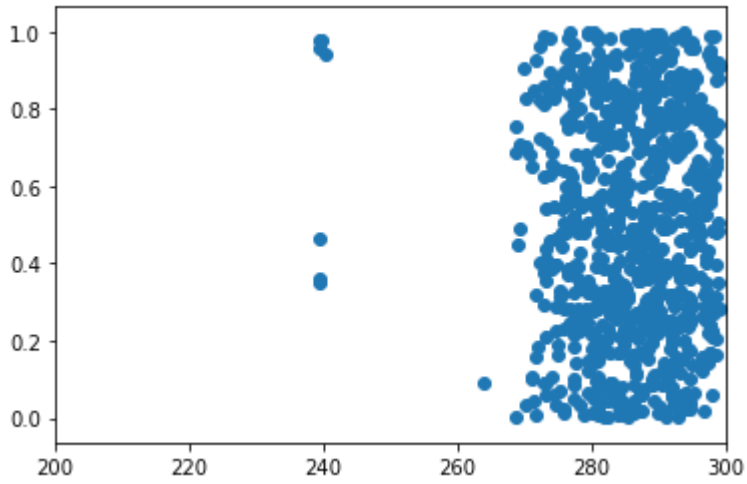
```
(2, 838, 2)
100.0000% (838 out of 838) spectrum with 2 peaks, its first peak lies in [2
00, 300)
```



Distribution of the second of Two peak

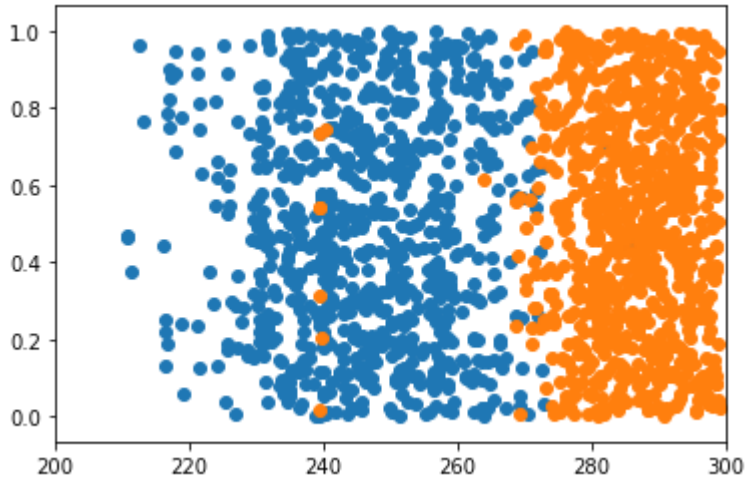
In [153]:
```
dis_sec_p(peak=[2])
```

```
(2, 838, 2)
100.0000% (838 out of 838) spectrum with 2 peaks, its first peak lies in [2
00, 300)
```



In [154]:
```
dis_sec_p()
```

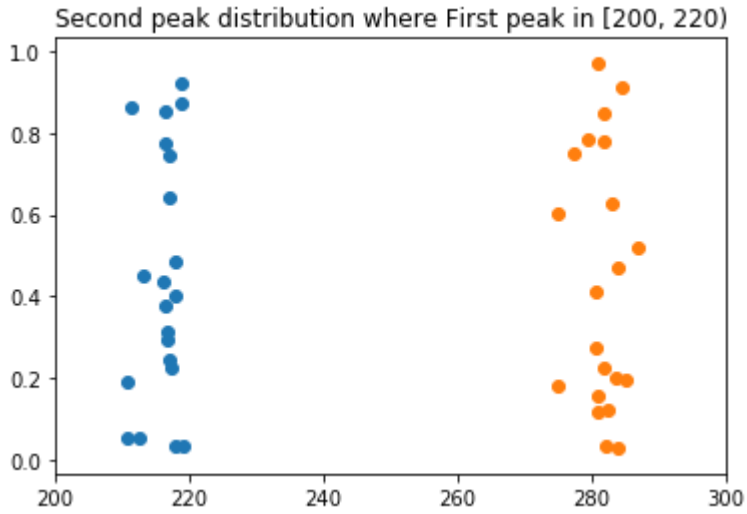
```
(2, 838, 2)
100.0000% (838 out of 838) spectrum with 2 peaks, its first peak lies in [2
00, 300)
```



Distribution of the second peak where the first peak is in [200, 220)

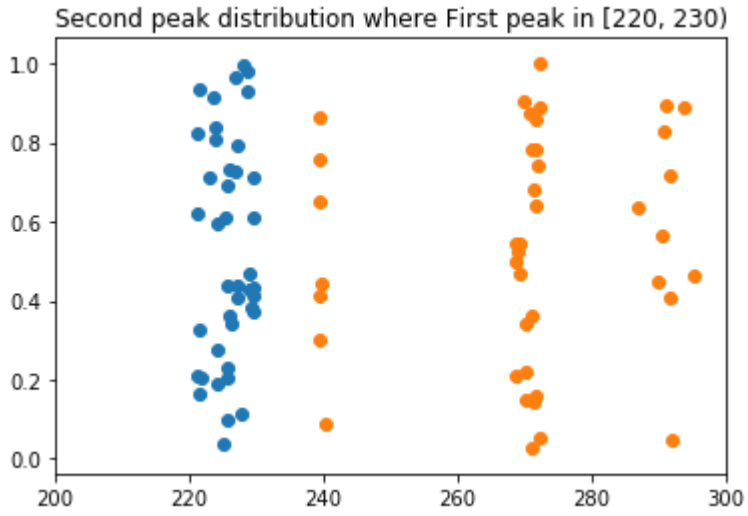
In [50]:
```
dis_sec_p(200, 220)
```

```
2.5060% (21 out of 838) spectrum with two peaks, its first peak lies in [20
0, 220)
```



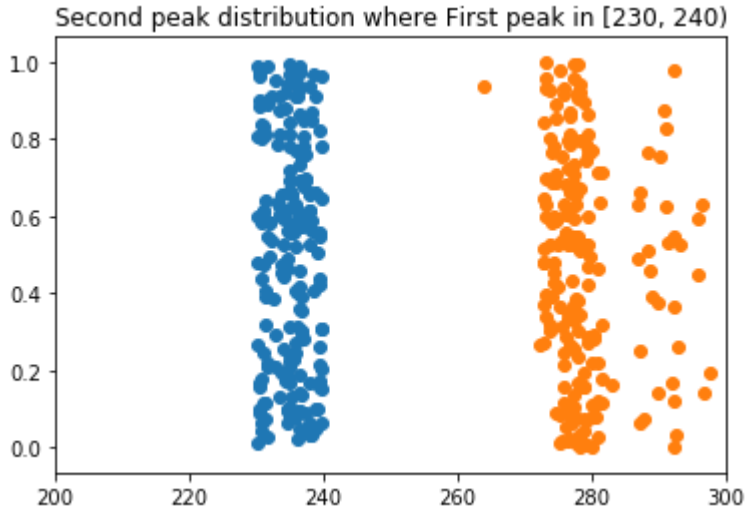
In [52]:
```
dis_sec_p(220, 230)
```

```
4.8926% (41 out of 838) spectrum with two peaks, its first peak lies in [22
0, 230)
```

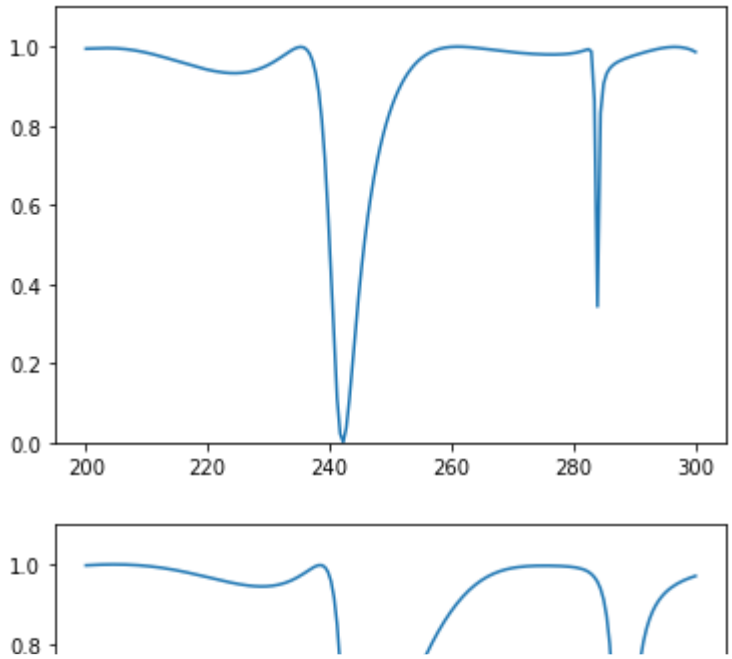


In [54]:
```
dis_sec_p(230, 240)
```

```
23.5084% (197 out of 838) spectrum with two peaks, its first peak lies in
[230, 240)
```

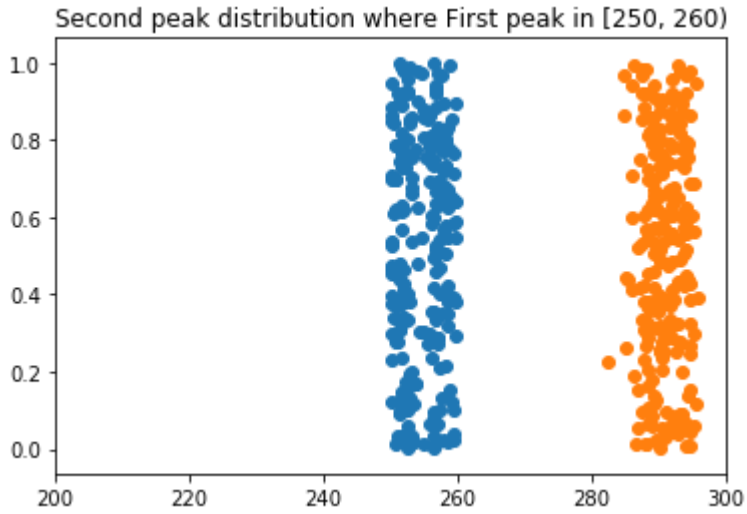


In [57]:
```
dis_sec_p(240, 250, True)
```



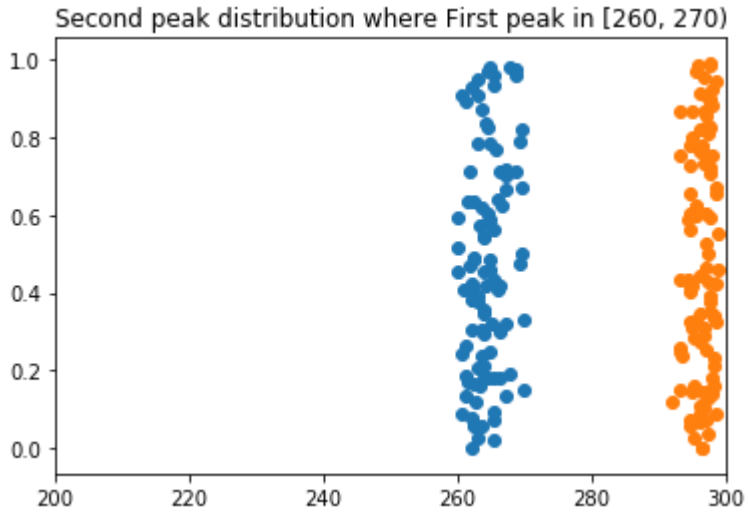
In [37]:
```
dis_sec_p(250, 260, False)
```

```
27.3270% (229 out of 838) spectrum with two peaks, its first peak lies in
[250, 260)
```



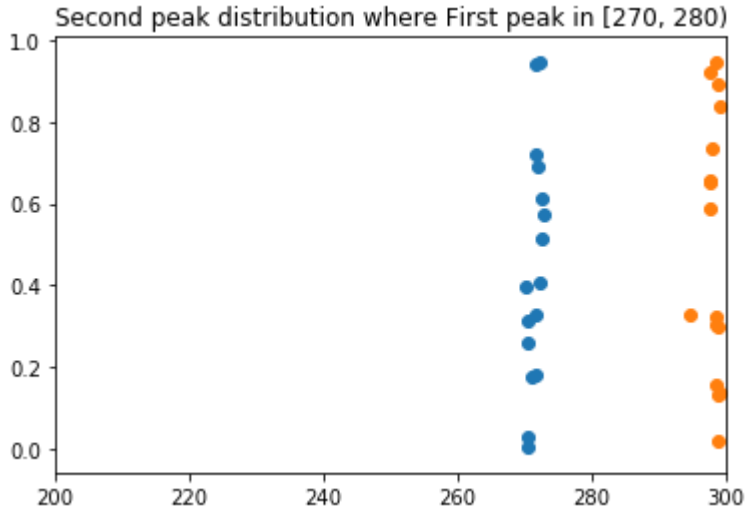
In [42]:
```
dis_sec_p(260, 270)
```

```
11.9332% (100 out of 838) spectrum with two peaks, its first peak lies in
[260, 270)
```



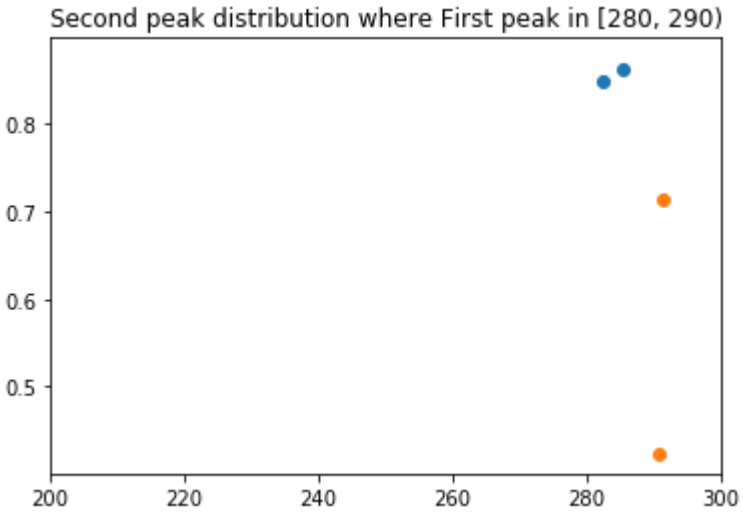
In [45]:
```
dis_sec_p(270, 280)
```

```
1.9093% (16 out of 838) spectrum with two peaks, its first peak lies in [27
0, 280)
```



In [46]:
```
dis_sec_p(280, 290)
```

```
0.2387% (2 out of 838) spectrum with two peaks, its first peak lies in [28
0, 290)
```



Distribution of One peak

In [146]:
```
dis_sec_p()
```

```
(1, 413, 2)
100.0000% (413 out of 413) spectrum with 1peaks, its first peak lies in [20
0, 300)
```