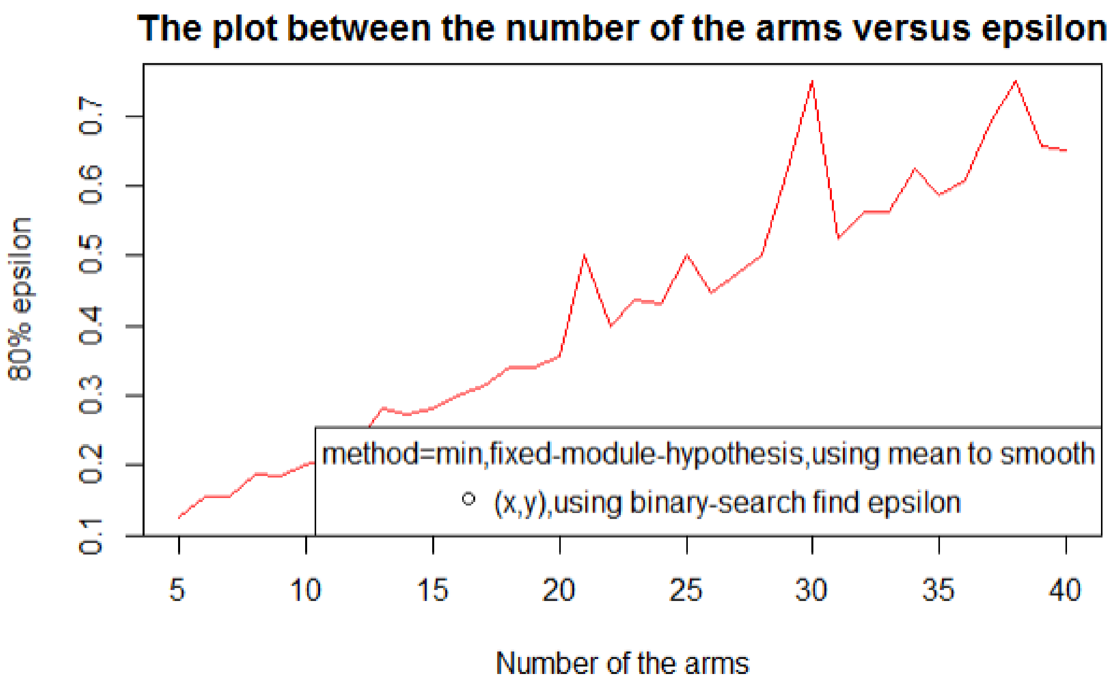
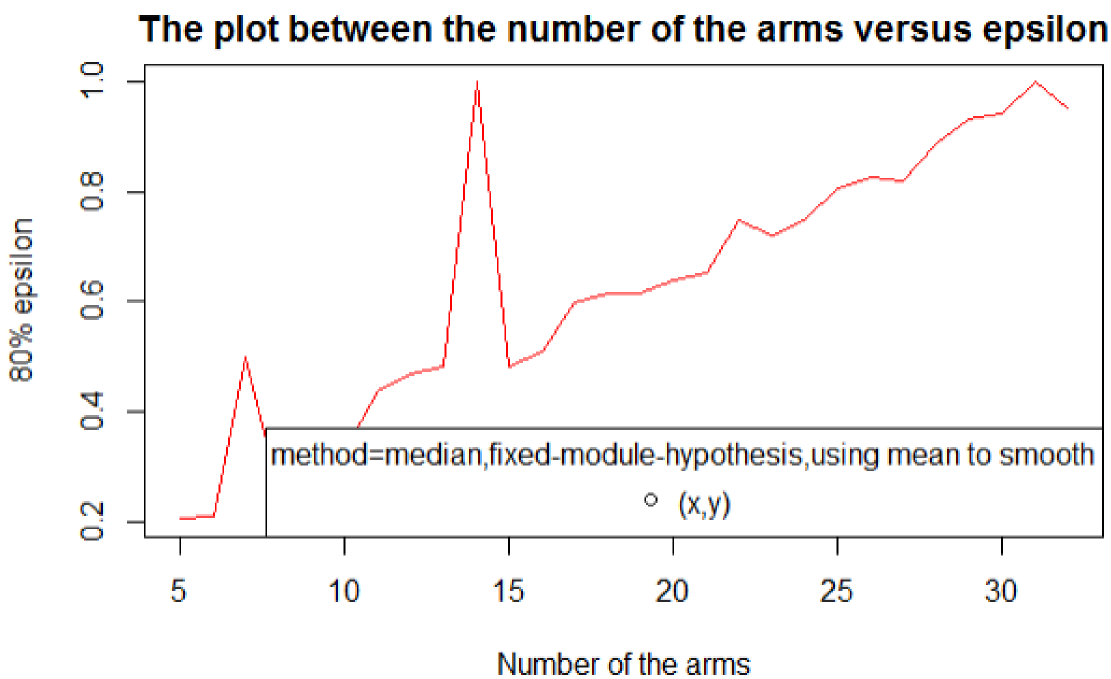
Casual Inference Report Week 6

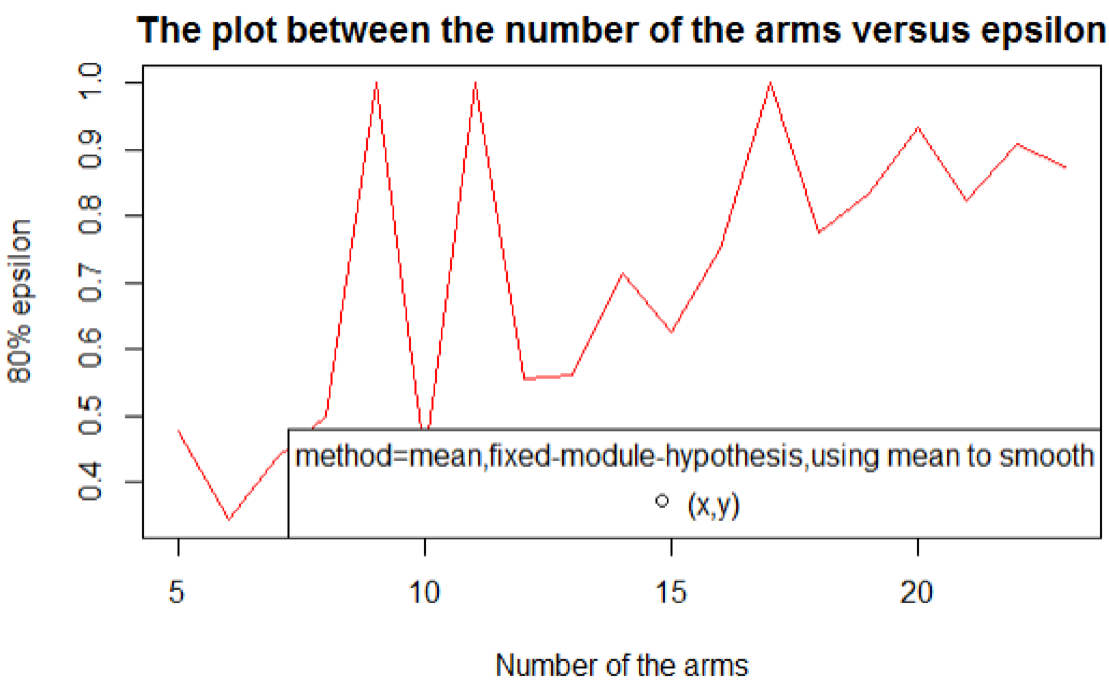
Zihao Wang

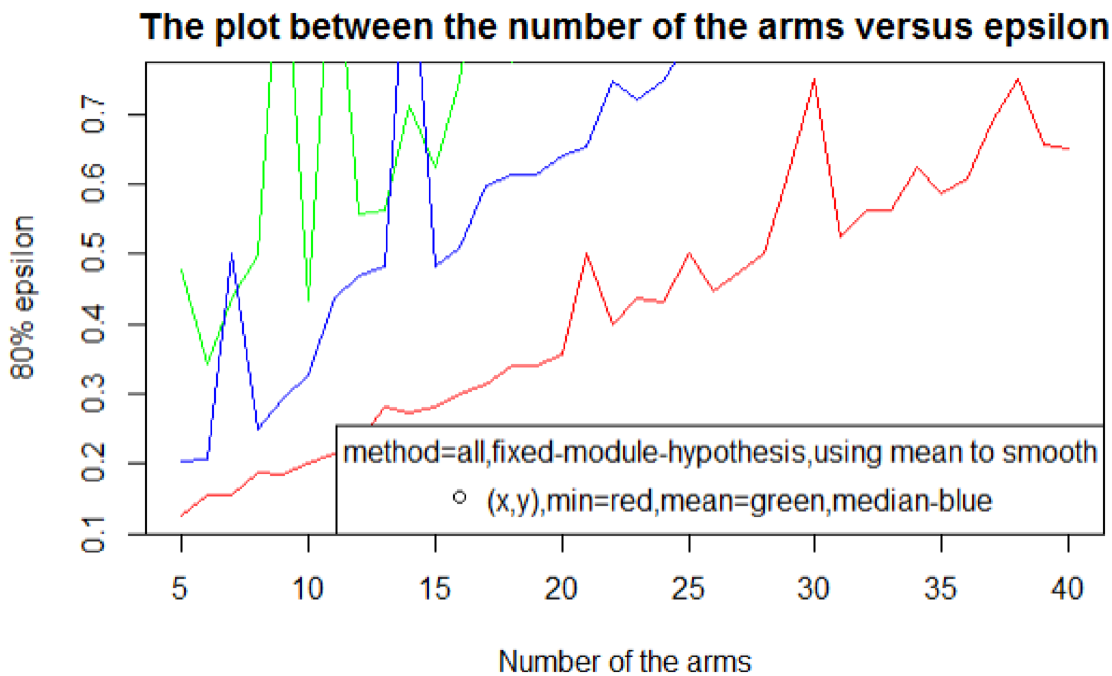
1 Plot the epsilon under which the statistics power is larger than 0.8 versus the number of the arms

1.1 Using binary searching method to find epsilon




The plot between the number of the arms versus epsilon
80% epsilon
0.2
0.4
0.6
0.8
1.0
5
10
15
20
25
30
Number of the arms
method=median,fixed-module-hypothesis,using mean to smooth
(x,y)

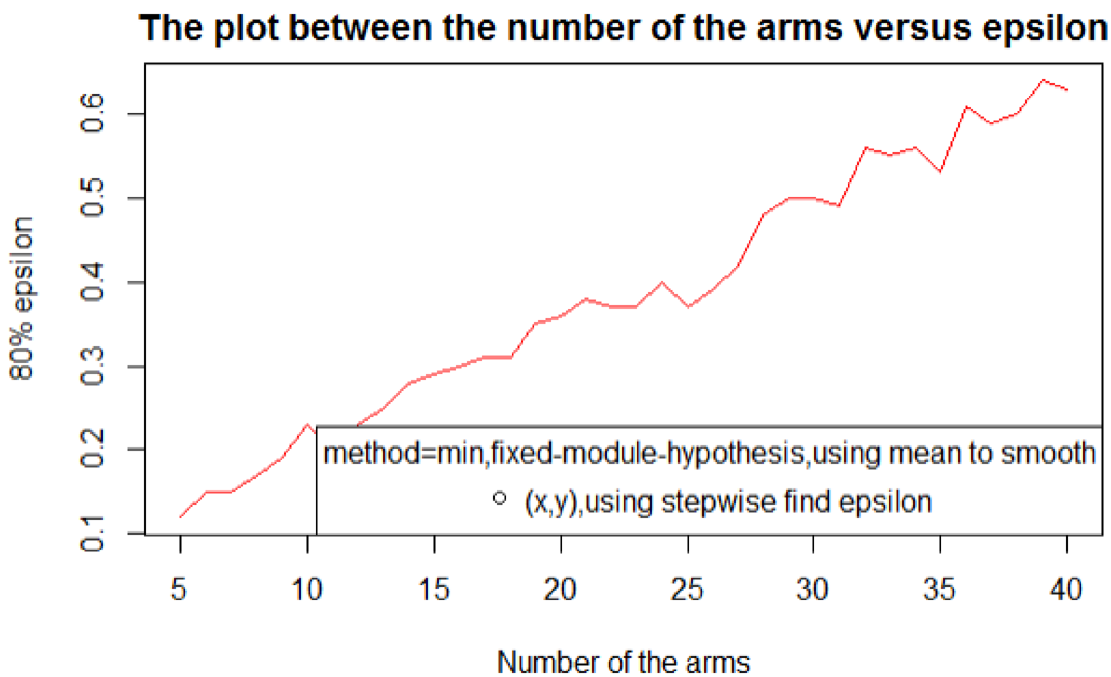

The plot between the number of the arms versus epsilon
80% epsilon
0.4
0.5
0.6
0.7
0.8
0.9
1.0
5
10
15
20
Number of the arms
method=mean,fixed-module-hypothesis,using mean to smooth
(x,y)

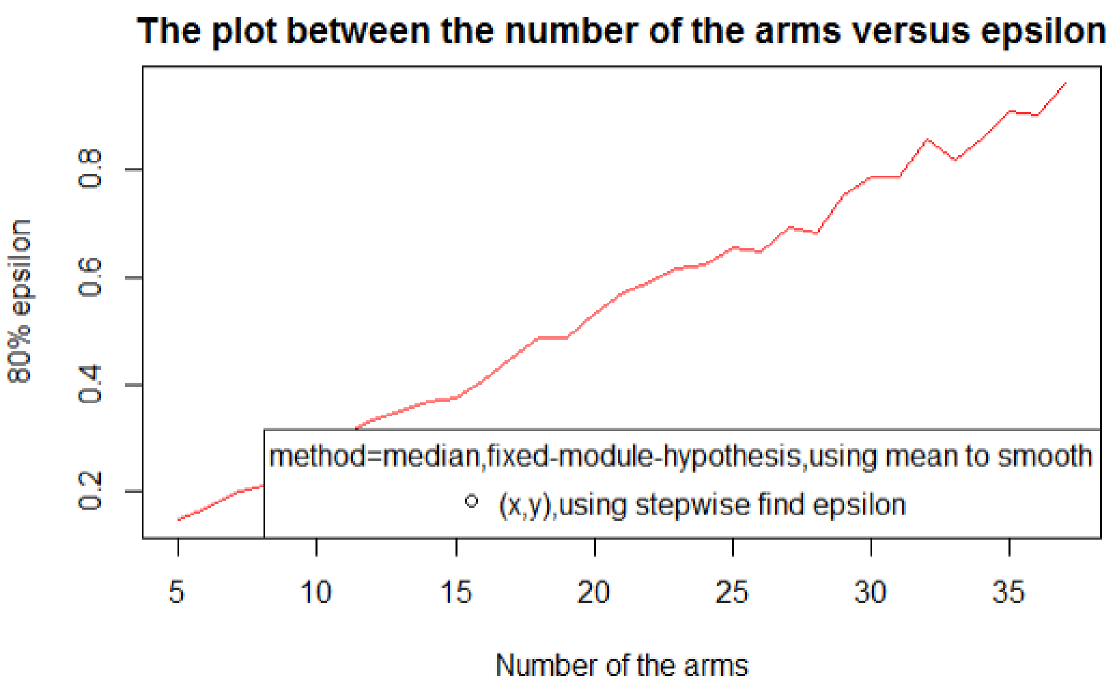


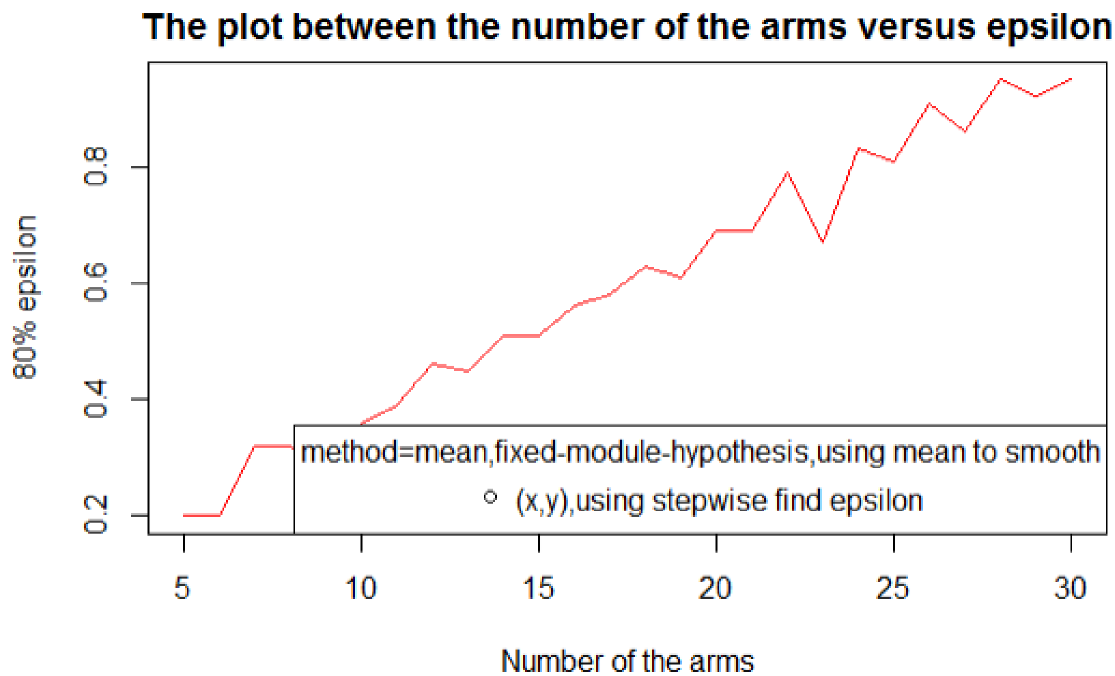
From above I found that using binary-searching method to calculate ϵ would cause some "outlier" in the plot, since there are plenty of sudden jumping points in the plot. The reason for that is that the function we use binary-searching method for might not be strictly increasing, thus, although faster, it cannot always find the ideal root we are looking for.

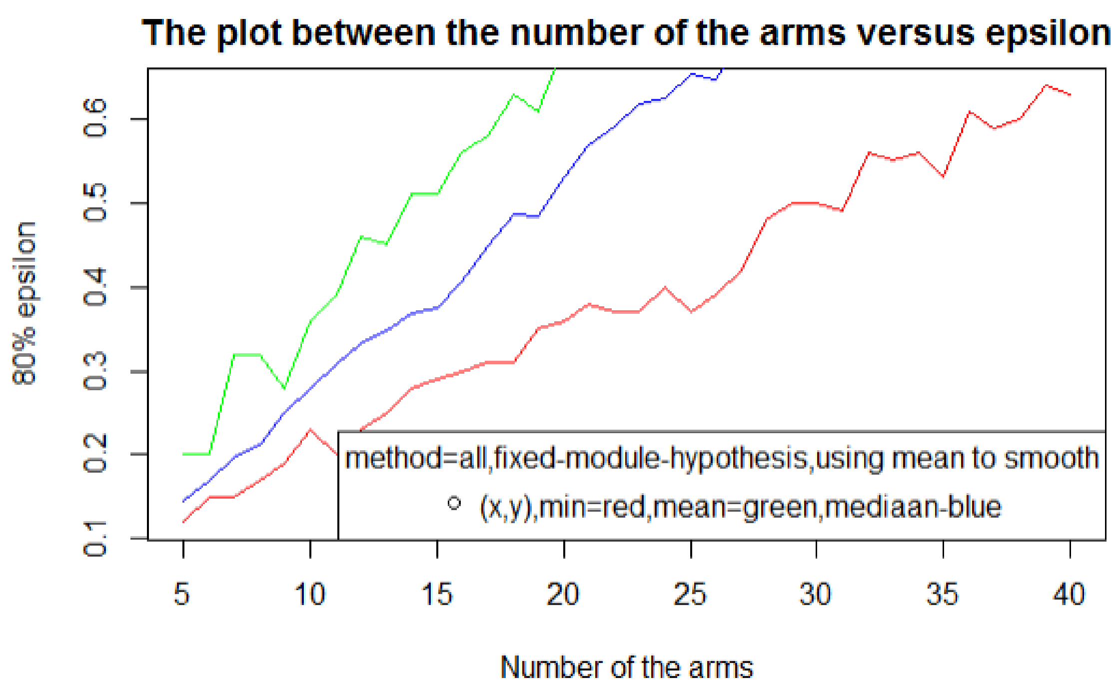
However, this unconvincing plot,especially, the fourth one, still enlightens me some useful information that in the fixing-module-hypothesis, minimum behaves best in reaching the cutoff of statistics power(say 0.8 here) fastest for it needs the smallest ϵ to achieve that, followed by median and mean. This conclusion is consistent with the conclusion I obtained weeks ago that the statistics power under minimum is the largest if all the other factors are the same.

1.2 Using stepwise method to find ϵ







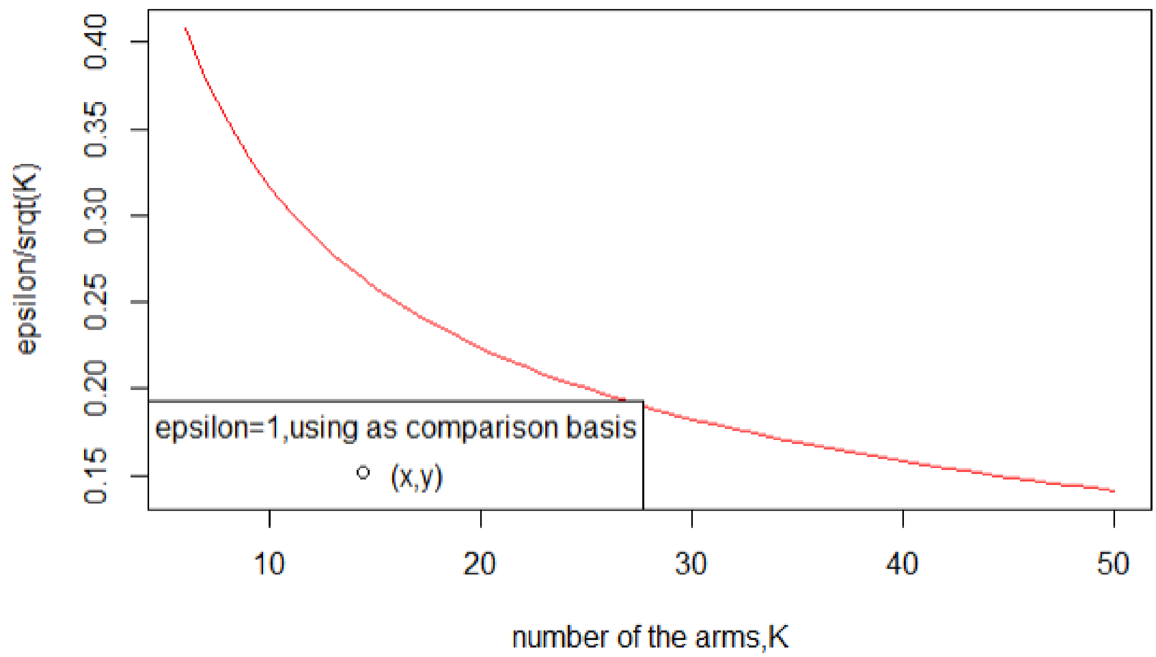


From the above, I found that the plot becomes smoother in this case with less "outliers" occuring in the plot. It indicates that although the stepwise method, iterating over all the number between 0 and 1 according to the precision to find the answer, is of low efficient, it has less demands towards the function we work on.

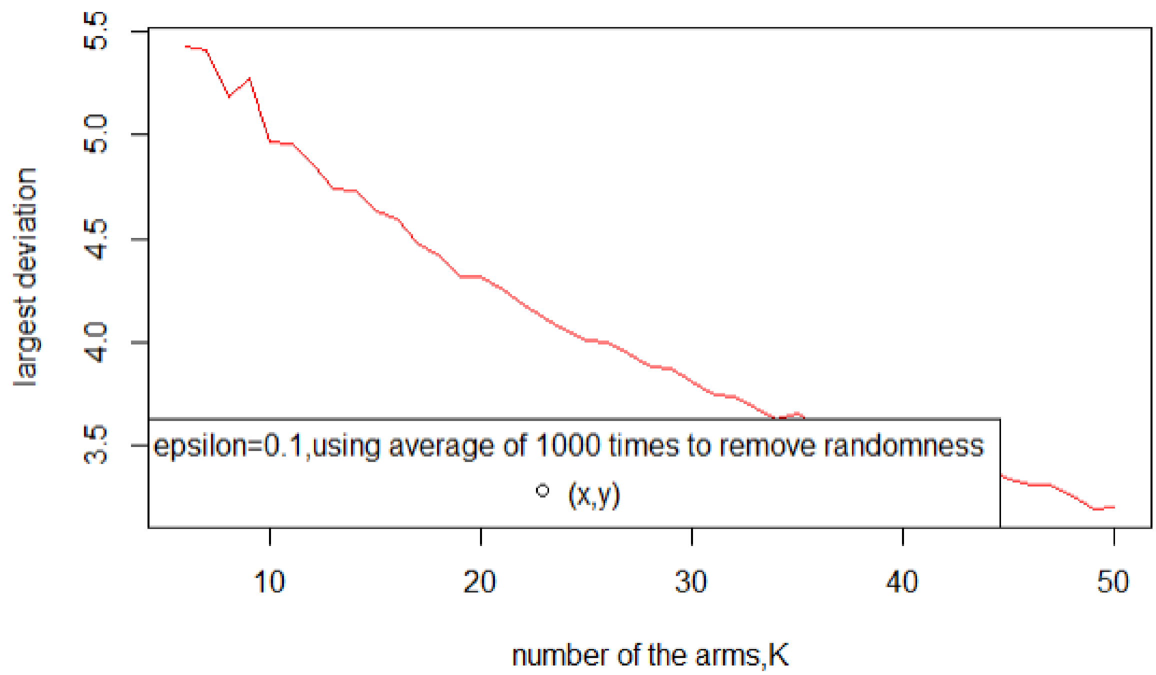
Similarly, we get the same conclusion that in this new alternative hypothesis(fixing the module of Ybar vector to a constant),minimum behaves best to make the statistics power greater than 0.8 by using the smallest ϵ against median and mean.

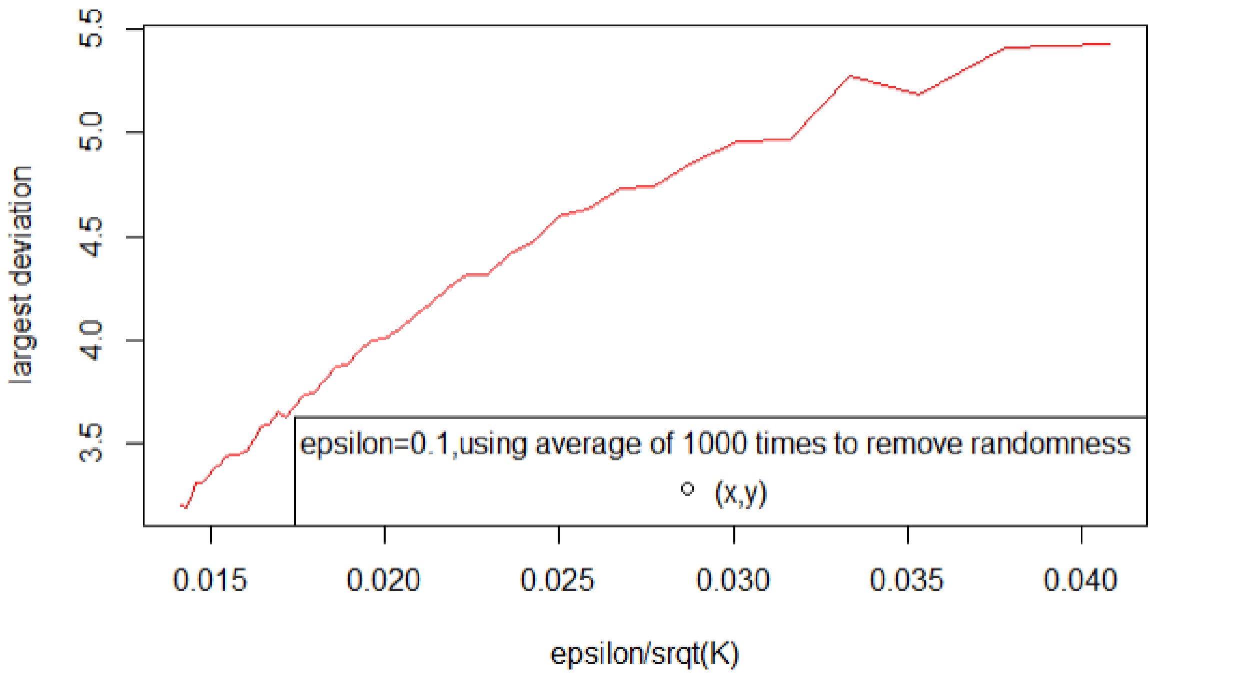
2 The relatiomship between the largest deviation and the number of the arms,K

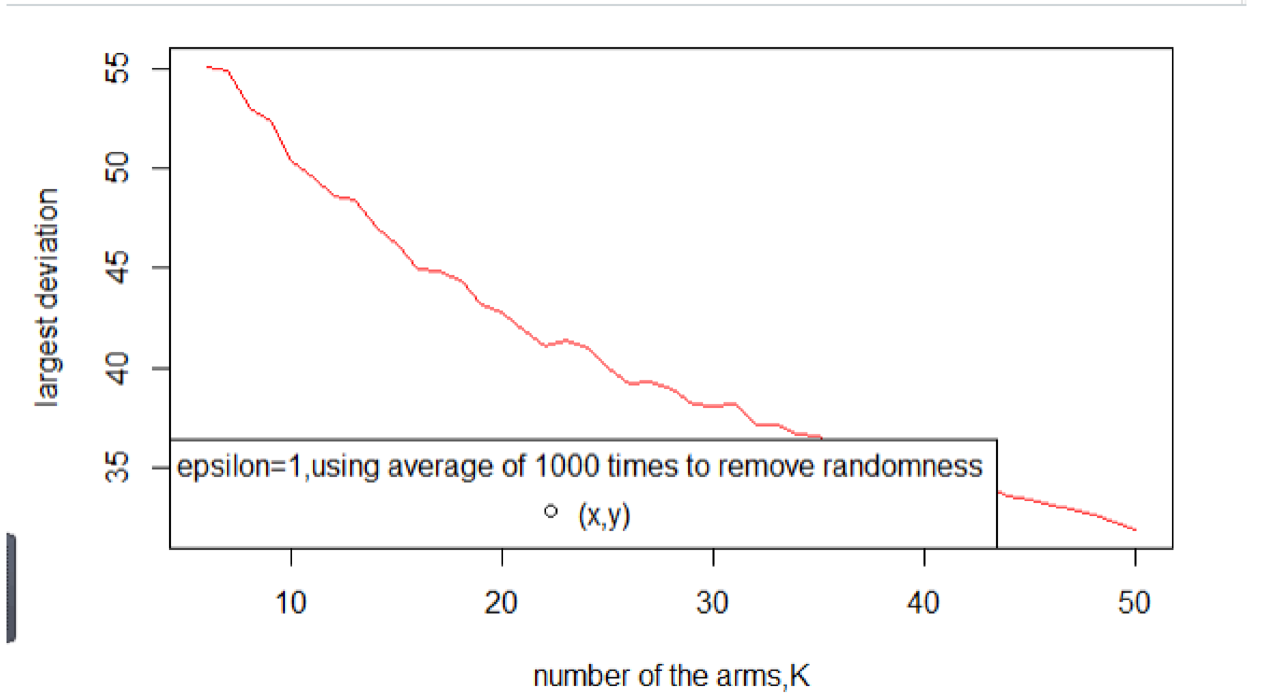
The following is the plot between $\frac{\epsilon}{\sqrt{K}}$ and K

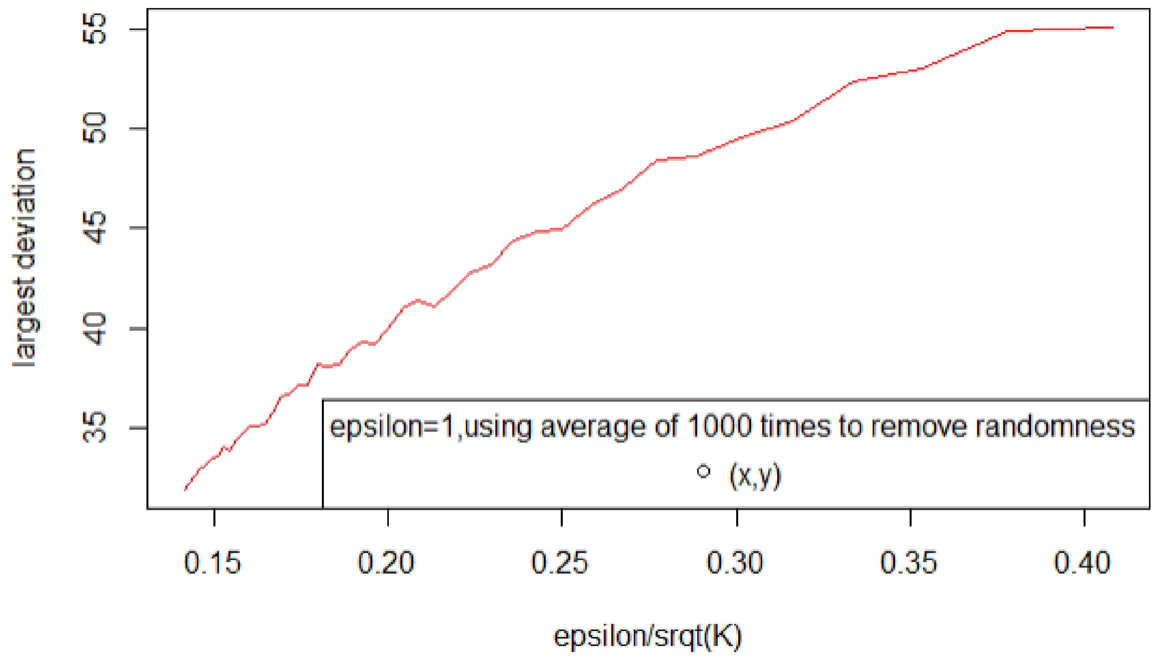


Then I set *epsilon* to 0.1, 1 and 10.

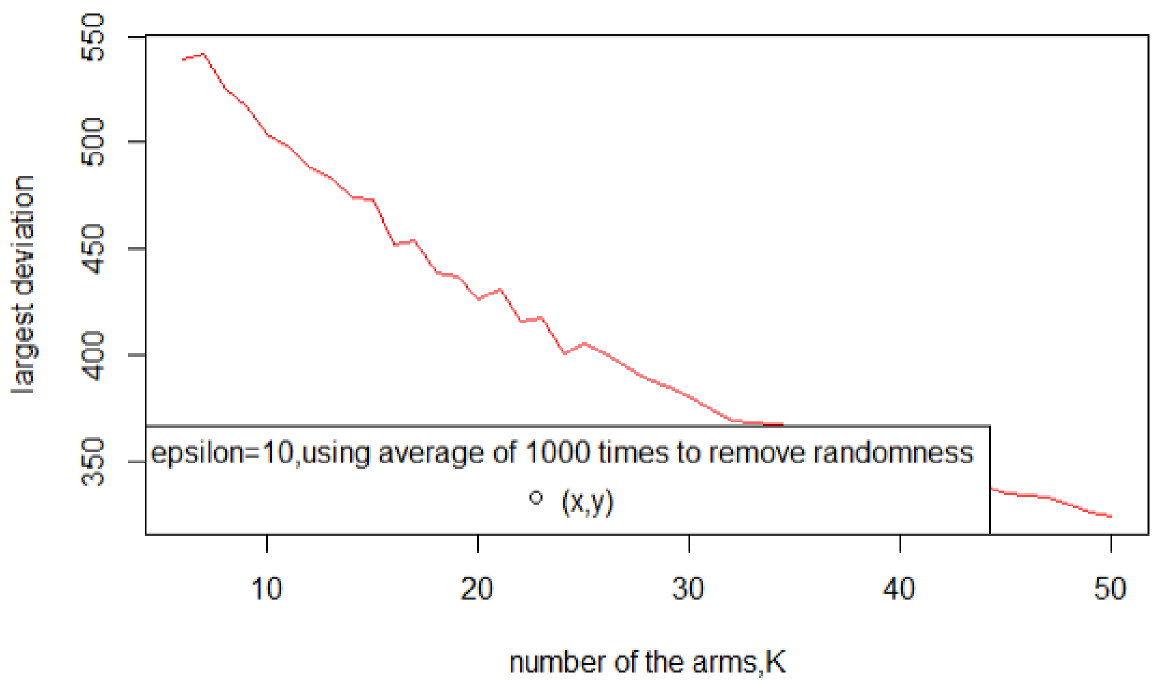



largest deviation
5.5
5.0
4.5
4.0
3.5
0.015
0.020
0.025
0.030
0.035
0.040
epsilon/srqt(K)
epsilon=0.1,using average of 1000 times to remove randomness
(x,y)

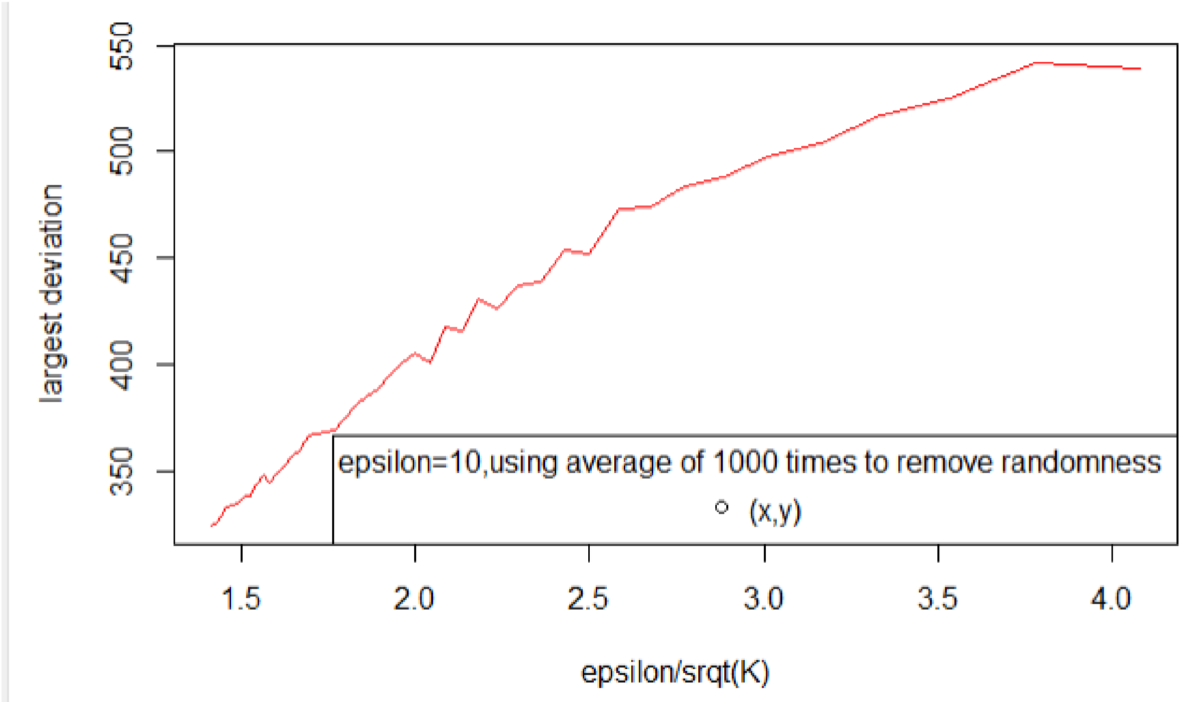

largest deviation
55
50
45
40
35
10
20
30
40
50
number of the arms,K
epsilon=1,using average of 1000 times to remove randomness
(x,y)



largest deviation
55
50
45
40
35
epsilon=1,using average of 1000 times to remove randomness
(x,y)
0.15
0.20
0.25
0.30
0.35
0.40
epsilon/srqt(K)



largest deviation
550
500
450
400
350
epsilon=10,using average of 1000 times to remove randomness
(x,y)
10
20
30
40
50
number of the arms,K



From the above, I found that no matter ϵ equals to what, the relationship between the largest-deviation and K is similar with that between $\frac{\epsilon}{\sqrt{K}}$ and K. If I plot the relationship between the largest deviation between $\frac{\epsilon}{\sqrt{K}}$, what I obtained is a nearly-linear plot. It indicates that the largest deviation of vector Ybar has a strong linear relationship with $\frac{\epsilon}{\sqrt{K}}$,which explains why in the sparse alternative hypothesis, the 0.8-ϵ has the strongest linear relationship between $\sqrt{K}$