Quantitative Foundations Project 3 Report

Gursimran Singh, Jiangnan Xu

November 2021

1 Results

Function	Valid?	Alpha	Explanation
1	asymptotically	0.15	As N$\uparrow$ $\forall$ θ probability$\uparrow$
2	not valid	NA	As N$\uparrow$ $\exists$ θ probability$\downarrow$
3	not valid	NA	low probability for most values of θ,N
4	valid	a tiny number [0,1)	High probability across all N $\forall$ θ
5	valid	>= 0.11	High probability across all N $\forall$ θ
6	valid when alpha >= 0.4	>= 0.4	High probability across all N $\forall$ θ
7	not valid	NA	low probability for most values of θ,N
8	not valid	NA	As N$\uparrow$ $\exists$ θ probability $\downarrow$
9	not valid	NA	low probability for most values of θ,N
10	not valid	NA	low probability for most values of θ,N

2 Procedures and Input Data

We used three distributions: Normal distribution, Uniform distribution, and Bernoulli distribution. And we captured 10000 pairs of A and B. We tried 100000 pairs of A and B at first, but it does not influence the results much and takes too long to run the codes for 10 functions. For each distribution, we used different values of theta, for each theta we experiment across different values of N, and for each value of N we generated 10000 pairs of A and B from 10000 samples of size N from the given distribution. We used this method to check how confidence interval behaves as the N value increases across different values of theta. Other inputs values are described as follows.

2.1 N

For all the three distributions, we take N for 10, 100, 1000, and 10000 for each function in experiments. Initially, we also tried 100000 and 1000000 for N for confidence Interval 1, because we want to observe how probability increases through increasing N. However, it verified our assumptions based on the observations for smaller N , and does not give us new insights on the pattern. Again, to save time, we did not run the code with N = 100000 or 1000000 for all ten functions.

2.2 theta

Theta or expectation, we take 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9. So we have big amount of data to observe, and also we can see whether the validity of functions varies depending on the value of theta. We also use this value of theta for sampling Bernoulli distribution function.

2.3 mu and sigma

For the normal distribution, we need to set the mu and sigma. In order to set the normal distribution between 0 and 1, we found six set of sigma (s) and mu (m) that work. s:0.1 , m:0.3 ; s:0.1 , m:0.4 ; s:0.1 , m:0.5 ; s:0.2 , m:0.5 ; s:0.1 , m:0.6 ; s:0.1 , m:0.7 ;

3 Findings

3.1 Valid

In our observation, among the 10 functions, confidence intervals of function 4, 5, 6 are valid with our inputs. For all three distributions, we observe a same pattern: The probability of the confidence interval containing the true mean(expectation) is very high across different values of N for any value of theta. Probability for confidence interval 4 is 100%, so we assume the alpha can be a value between 0 and 1, since it will satisfy the condition for all values in the range. For confidence interval 5, for all theta and N, the probability is close or above 89%. So we assume the alpha is around 0.11. It worth to mention, the confidence interval of function 6 is only valid when the alpha is bigger than or equal to 0.4. Since the probability is 100% for all theta for 4th interval, We assume that the interval is larger than the data set domain.

3.2 Asymptotically

Among the 10 functions, only confidence interval of function 1 is asymptotically valid. We observed the pattern that when N is small (10, 100), the percentage of probability is low but increases when N increases; When N is bigger (1000, 10000, 100000, 1000000), the percentage of probability increased dramatically.

3.3 Invalid

We have invalid confidence interval of function 2, 3, 7, 8, 9, 10. Among which, we observed 3 of confidence intervals (function 7, 9, 10) showed similar patterns.

Confidence interval 2: for Bernoulli and Uniform distributions, the value of probability increases when N increases. However, when it comes to the Normal distribution, we observed opposite pattern: the probability decreases when N increases for certain theta values. Also, only when theta = mean +/- sigma, the function is "valid".

For confidence interval 3, 7, 9 and 10, for normal distribution when theta = mean, the probability is very high (close to 100%, 7 and 10 are 100%). For uniform distribution, only when theta = 0.5, the probability is very high for all values of N.

Confidence interval 8 is invalid. Again, we observed in normal distribution, only when theta = mean +/- sigma, the function is "valid". And in Uniform and Bernoulli distributions, the probability decreases as the N increases for certain theta values.