Project 3 Confidence Intervals

Group 13

Gursimran Singh, Jiangnan Xu

Design of input data

- We used **N as 10, 100, 1000, and 10000** for each function.
- We also tried 100000 and 1000000 for CI 1, but didn't get new insights, just verified assumptions.
- We take **theta or expectation values from 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9**. We used these values for Bernoulli distribution as well.
- For Normal distribution, we used six set of sigma and mu values:- **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 ;**

Testing procedure

- 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 theta, we generate **10000** samples of A and B for each value of **N** for a given distribution.

Valid

Confidence interval 4, 5, 6

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.

Confidence Interval 4

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.

Since the probability is 100% for all theta for 4th interval, We assume that the interval is larger than the data set domain.

```
CI 4 - UNIFORM

CI

i =

     4


Outtable =

  36×3 table

      N      theta/mean    percentage (approx > 1-alpha)
    _____    __________    _____________________________

       10         0.1                    100
      100         0.1                    100
     1000         0.1                    100
    10000         0.1                    100
       10         0.2                    100
      100         0.2                    100
     1000         0.2                    100
    10000         0.2                    100
       10         0.3                    100
      100         0.3                    100
     1000         0.3                    100
    10000         0.3                    100
       10         0.4                    100
      100         0.4                    100
     1000         0.4                    100
    10000         0.4                    100
       10         0.5                    100
      100         0.5                    100
     1000         0.5                    100
    10000         0.5                    100
       10         0.6                    100
      100         0.6                    100
     1000         0.6                    100
    10000         0.6                    100
       10         0.7                    100
      100         0.7                    100
     1000         0.7                    100
    10000         0.7                    100
       10         0.8                    100
      100         0.8                    100
     1000         0.8                    100
    10000         0.8                    100
       10         0.9                    100
      100         0.9                    100
     1000         0.9                    100
    10000         0.9                    100
```

Confidence Interval 5

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.

```
CI 5 - UNIFORM

CI

i =

     5


Outtable =

  36×3 table

      N      theta/mean    percentage (approx > 1-alpha)
    _____    __________    _____________________________

       10        0.1                  89.97
      100        0.1                  90.08
     1000        0.1                  89.44
    10000        0.1                  89.84
       10        0.2                  90.28
      100        0.2                  90.32
     1000        0.2                   90.2
    10000        0.2                  89.83
       10        0.3                  89.55
      100        0.3                   90.1
     1000        0.3                  90.29
    10000        0.3                  89.93
       10        0.4                   90.1
      100        0.4                  90.47
     1000        0.4                  90.38
    10000        0.4                  90.41
       10        0.5                  90.27
      100        0.5                  89.75
     1000        0.5                  90.04
    10000        0.5                  90.47
       10        0.6                  90.05
      100        0.6                  89.22
     1000        0.6                  90.02
    10000        0.6                   89.3
       10        0.7                   89.8
      100        0.7                  89.84
     1000        0.7                  88.92
    10000        0.7                  89.33
       10        0.8                  90.15
      100        0.8                  90.43
     1000        0.8                  89.98
    10000        0.8                  89.78
       10        0.9                   90.3
      100        0.9                  90.04
     1000        0.9                   90.1
    10000        0.9                   89.5
```

Confidence Interval 6

The confidence interval of function 6 is only valid when the alpha is bigger than or equal to 0.4.

```
CI 6 - BERNOULLI

CI bernoulli

i =

     6


Outtable =

  36×3 table

      N      theta/mean    percentage (approx > 1-alpha)
    _____    __________    _____________________________

       10        0.1                  65.15
      100        0.1                  97.45
     1000        0.1                    100
    10000        0.1                    100
       10        0.2                  89.37
      100        0.2                    100
     1000        0.2                    100
    10000        0.2                    100
       10        0.3                  97.17
      100        0.3                    100
     1000        0.3                    100
    10000        0.3                    100
       10        0.4                  99.47
      100        0.4                    100
     1000        0.4                    100
    10000        0.4                    100
       10        0.5                   99.9
      100        0.5                    100
     1000        0.5                    100
    10000        0.5                    100
       10        0.6                  99.99
      100        0.6                    100
     1000        0.6                    100
    10000        0.6                    100
       10        0.7                    100
      100        0.7                    100
     1000        0.7                    100
    10000        0.7                    100
       10        0.8                    100
      100        0.8                    100
     1000        0.8                    100
    10000        0.8                    100
       10        0.9                    100
      100        0.9                    100
     1000        0.9                    100
    10000        0.9                    100
```

Asymptotically

Confidence interval 1

Confidence interval 1

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.

```
CI Normal

i =

     1


Outtable =

  45×3 table

      N      theta/mean    percentage (approx > 1-alpha)
    _____    __________    _____________________________

       10          0.1                          20.385
      100          0.1                          89.845
     1000          0.1                             100
    10000          0.1                             100
    1e+05          0.1                             100
       10          0.2                          49.753
      100          0.2                          99.904
     1000          0.2                             100
    10000          0.2                             100
    1e+05          0.2                             100
       10          0.3                          82.055
      100          0.3                             100
     1000          0.3                             100
    10000          0.3                             100
    1e+05          0.3                             100
       10          0.4                          97.539
      100          0.4                             100
     1000          0.4                             100
    10000          0.4                             100
    1e+05          0.4                             100
       10          0.5                          99.807
      100          0.5                             100
     1000          0.5                             100
    10000          0.5                             100
    1e+05          0.5                             100
       10          0.6                          97.448
      100          0.6                             100
     1000          0.6                             100
    10000          0.6                             100
    1e+05          0.6                             100
       10          0.7                          82.474
      100          0.7                             100
     1000          0.7                             100
    10000          0.7                             100
    1e+05          0.7                             100
       10          0.8                          49.823
      100          0.8                          99.899
     1000          0.8                             100
    10000          0.8                             100
    1e+05          0.8                             100
       10          0.9                          20.396
      100          0.9                          89.896
     1000          0.9                             100
```

Invalid

Confidence interval 2, 3, 7, 8, 9, 10

Among which, we observed 4 of confidence intervals (function 3, 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" with minor outliers.

```
sample_normal(50,0.2,0.5)

CI Normal

i =

     2


Outtable =

  36×3 table

      N      theta/mean    percentage (approx > 1-alpha)
    _____    __________    _____________________________

       10         0.1                  20.65
      100         0.1                  66.93
     1000         0.1                  75.63
    10000         0.1                   97.3
       10         0.2                  49.99
      100         0.2                  99.13
     1000         0.2                    100
    10000         0.2                    100
       10         0.3                  81.95
      100         0.3                    100
     1000         0.3                    100
    10000         0.3                    100
       10         0.4                  97.28
      100         0.4                    100
     1000         0.4                    100
    10000         0.4                    100
       10         0.5                  99.76
      100         0.5                    100
     1000         0.5                    100
    10000         0.5                    100
       10         0.6                  97.77
      100         0.6                    100
     1000         0.6                    100
    10000         0.6                    100
       10         0.7                  82.28
      100         0.7                    100
     1000         0.7                    100
    10000         0.7                    100
       10         0.8                   49.4
      100         0.8                  96.48
     1000         0.8                    100
    10000         0.8                    100
       10         0.9                  21.45
      100         0.9                  38.95
     1000         0.9                   8.25
    10000         0.9                      0
```

Confidence Interval 3, 7, 9, 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.

CI 3 - UNIFORM

CI

i =

3

Outtable =

36×3 table

N	theta/mean	percentage (approx > 1-alpha)
10	0.1	0.31
100	0.1	0
1000	0.1	0
10000	0.1	0
10	0.2	7.2
100	0.2	0
1000	0.2	0
10000	0.2	0
10	0.3	38.13
100	0.3	0
1000	0.3	0
10000	0.3	0
10	0.4	75.84
100	0.4	6.45
1000	0.4	0
10000	0.4	0
10	0.5	90.4
100	0.5	94.95
1000	0.5	95.11
10000	0.5	95.17
10	0.6	76.01
100	0.6	6.61
1000	0.6	0
10000	0.6	0
10	0.7	38.51
100	0.7	0
1000	0.7	0
10000	0.7	0
10	0.8	7.9
100	0.8	0
1000	0.8	0
10000	0.8	0
10	0.9	0.19
100	0.9	0
1000	0.9	0
10000	0.9	0

sample_normal(50,0.1,0.5)

CI Normal

i =

3

g =

0.5000

Outtable =

36×3 table

N	theta/mean	percentage (approx > 1-alpha)
10	0.1	0
100	0.1	0
1000	0.1	0
10000	0.1	0
10	0.2	0
100	0.2	0
1000	0.2	0
10000	0.2	0
10	0.3	0
100	0.3	0
1000	0.3	0
10000	0.3	0
10	0.4	10.28
100	0.4	0
1000	0.4	0
10000	0.4	0
10	0.5	90.57
100	0.5	94.59
1000	0.5	95.18
10000	0.5	94.89
10	0.6	10.58
100	0.6	0
1000	0.6	0
10000	0.6	0
10	0.7	0
100	0.7	0
1000	0.7	0
10000	0.7	0
10	0.8	0
100	0.8	0
1000	0.8	0
10000	0.8	0
10	0.9	0
100	0.9	0
1000	0.9	0
10000	0.9	0

Confidence Interval 8

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.

```
sample_normal(50,0.1,0.5)

CI Normal

i =

     8


g =

    0.5000


Outtable =

  36×3 table

      N      theta/mean    percentage (approx > 1-alpha)
    _____    __________    _____________________________

       10        0.1                      0
      100        0.1                      0
     1000        0.1                      0
    10000        0.1                      0
       10        0.2                   0.01
      100        0.2                      0
     1000        0.2                      0
    10000        0.2                      0
       10        0.3                   1.97
      100        0.3                      0
     1000        0.3                      0
    10000        0.3                      0
       10        0.4                  50.16
      100        0.4                  82.73
     1000        0.4                  99.89
    10000        0.4                    100
       10        0.5                  98.99
      100        0.5                    100
     1000        0.5                    100
    10000        0.5                    100
       10        0.6                  81.75
      100        0.6                  99.36
     1000        0.6                    100
    10000        0.6                    100
       10        0.7                  20.92
      100        0.7                   0.21
     1000        0.7                      0
    10000        0.7                      0
       10        0.8                   1.47
      100        0.8                      0
     1000        0.8                      0
    10000        0.8                      0
       10        0.9                   0.06
      100        0.9                      0
     1000        0.9                      0
    10000        0.9                      0
```

Observations

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

Thank you!

Q & A