

Project 3 Confidence Interval

Reporter: Sumeet Atul Vadhavkar and Ye Zheng

# FUNCTION	VALIDITY	CONFIDENCE
1	Asymptotic	100%
2	Invalid	-
3	Asymptotic	95%
4	Invalid	-
5	Always	90%
6	Invalid	100%
7	Always	90% - 99%
8	Invalid	-
9	Asymptotic	99%
10	Always	96% - 99%

To prove a range estimation is invalid:

- For large enough N , it miss the estimated value with a large probability.

Our testing method:

- Choose $N = 3000$, sample 10,000 data from Uniform, Normal, Bernoulli distribution with different parameters.
- Use CLT confidence interval as reference ($\alpha = 0.05$), observe the missing rate and mean estimated interval of each function.
- For the functions seems asymptotic valid, we use $N = 10$ to check their validity for all datasets.

Uniform Distribution

We use Uniform distribution on $[0, 1]$, $[0, 0]$, $[0.2, 0.3]$, $[0.4, 1]$, $[0.70.72]$, $[0.9, 0.95]$ for testing:

- Function 4 misses 100% for $[0, 0]$, proved to be invalid;
- Function 3, 5, 9 have stable confidence level around 95%, 90%, 99% respectively, but Function 5 has very loose bound;

- Other functions have almost missed rate 0;

Normal Distribution

We use Normal distribution

$\mathcal{N}(0.1, 0.01), \mathcal{N}(0.3, 0.1), \mathcal{N}(0.5, 0.1), \mathcal{N}(0.7, 0.3), \mathcal{N}(0.9, 0.01)$ for testing:

- Function 1 seems to output the minimum and maximum as the estimation.
- Function 6 has fixed bound $[0, 1]$ in all test cases, seems to be valid for all datasets.

Bernoulli Distribution

We use Bernoulli distribution with $\theta = 0, 0.01, 0.1, 0.3, 0.5, 0.7, 0.9, 0.99$ for testing:

- Function 2 has missing rate 100% for $\theta = 0.99$, proved to be invalid;
- Function 8 has missing rate 100% for $\theta = 0.9$, proved to be invalid;
- Function 6 has missing rate 100% for $\theta = 0.01$, proved to be invalid;
- Function 7 has confidence level 90% to 99%, seems to use $[\bar{X} - 0.015, \bar{X} + 0.015]$ as the estimation.
- Function 10 has confidence level 96% to 99%, seems to use $[\bar{X} - 0.018, \bar{X} + 0.018]$ as the estimation.

Checking the Remained

Function 1, 3, 5, 7, 9, 10 need to check validity for small N :

- Function 1, 3, 9 have much larger missing rate than their confidence level, so they are asymptotic valid;
- Function 5, 7, 10 are valid for all datasets.