ML703/803: (Advanced) Probabilistic and Statistical Inference (Spring 2024)

Week 1 Lab: Understanding Uncertainty in Samples
12 January 2024

The objective of this lab is to have a clear understanding of uncertainty in a sample and uncertainty of a statistic (which is a quantity computed from values in a sample), as well as basic concepts in statistics.

For illustrative purposes, we will use a very simple statistical problem (concerning the estimation of the mean from a sample), but the observations apply to other statistical procedures as well.

You are free to use any programming tool (e.g., Python, MATLAB, R, etc.) that you are comfortable with. Furthermore, you can download and use built-in functions in any package to visualize graphs. However, you are expected to implement the procedures to perform tasks 1-6 below yourself, i.e., you are not allowed to use built-in functions to directly perform the following tasks. We also provide the sample code in Python (in the format of Jupyter notebook) for this lab.

- **Task 1**: Draw a sample of size $n = 4$ from the normal distribution with mean 0 and standard deviation 1 (denoted by $\mathcal{P}_1$), and calculate its sample mean (or average), which is the statistic we are concerned with.

- **Task 2**: Repeat the above Task 1000 times. Draw the histogram of the calculated sample means, which shows the sampling distribution of the statistic; compare this empirical distribution with the theoretical sampling distribution, which is a normal distribution with mean 0 and standard deviation $1/\sqrt{n} = 1/2$.

- **Task 3**: Do Task 2 again, but this time use sample size $n = 100$ instead.

- **Task 4**: Do Task 2 again, but this time let us draw samples of size $n = 4$ from the Laplace distribution with mean 0 and variance 1 (denoted by $\mathcal{P}_2$).

- **Task 5**: Do Task 4 again, but this time use sample size $n = 100$ instead.

- **Task 6**: Do Task 2 again, but this time let us draw samples of size $n = 4$ from the Uniform distribution on [0, 1] (denoted by $\mathcal{P}_3$). Is it more similar to normal compared to the Laplace one? How many samples we need to make the difference with the normal one not-distinguishable by eye?

- **Task 7**: Do Task 2 again, but this time let us draw samples of size $n = 4$ from the Poisson distribution with parameter equal to 1 (denoted by $\mathcal{P}_4$). What is the key feature of this distribution compared to previous ones?

- **Task 8**: Do Task 7 again, but this time use sample size $n = 100$ instead.