1155093445 Leung Wang Fat

CSCI 3320: Fundamental of Machine Learning

• Describe how δ and n affect the PMF.

The smaller the δ, the higher the number of bins, the higher the number of available outcomes, the smaller the probability mass for an outcome (a value of the sampling mean).
The higher the n, the smaller the difference between sampling mean, the higher chance that the probability of a sampling mean converge to the population mean and less variance in the pmf.

• Describe how close the different PMF is to the true sampling distribution of the mean and your E[X] to the population mean with respect to different δ and n.

We can see that when the δ is smaller, the result PMF will be closer to the true sampling distribution. Also, the result E[X] will be closer to the population mean.
We can see that when the n is higher, the result PMF will be closer to the true sampling distribution. Also, the result E[X] will be closer to the population mean.

• Give the clear explanation of why the result would be like that
The sampling distribution follows the normal distribution. By the weak law of large number, as the sample size increase. The mean of the samples will converge to the true population mean. Thus, increasing n will reduce the variance of the sampling distribution of means (denoted as $\sigma_{\bar{X}}^2$ and true variance as σ^2. Thus, the pmf will be closer to the true sampling distribution and the true population mean.

$$E[(\bar{X} - \mu)^2] = \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$$

Increasing the bin number by reducing the δ will reduce the error for approximating a sampling mean to a particular bin value, which gives a better approximation of the true population mean and sampling distribution.