DS 298: Work Assignment - 1

Due Feb 02, 2026

Sample from a truncated normal distribution $\mathcal{N}(\frac{1}{2}, \frac{1}{36})$, the arcsine distribution (a beta distribution with $\alpha = \beta = 1/2$), and the uniform distribution with the limits of the random variables as $(0, 1)$. These empirical distributions (histograms) can be generated using inbuilt functions of Python or Matlab. Write your description of observations and conclusions including the plots and tables in it.

1. Plot the variation of Kolmogorov-Smirnoff (K-S) statistic with the number of samples n for the above 3 distributions, by comparing the corresponding empirical and reference distributions as n varies from 10^2 to 10^5 samples. Note that when the samples are drawn from a distribution, the empirical CDF can be directly generated without an associated PDF. Average the KS statistic over multiple trials for a smooth plot.

2. Generate a K-S statistic comparison table (again averaging over multiple trials) in the form of a 3×3 symmetric confusion matrix for each of the sample sizes 10^2, 10^3 and 10^4, where now each empirical distribution is compared with all the three reference distributions given.

3. Repeat the construction of the above confusion matrix using the Bhattacharya and the Hellinger distances using 10^4 samples. In these cases, the required distance integral can be replaced by a point-wise summation over the histogram of the empirical density, and its normalization using the point-wise sum of the reference density (which is 1 in the case of an exact integration).

   Note: Use logarithmic scale in an axis/plot wherever appropriate. Submit the descriptive response and the code as separate files, all zipped into a single folder identified by your name in full, on the MS Teams channel for the class.