DS 298: Work Assignment - 1

Due Feb 10, 2025

1. 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 $(0, 1)$.

   a) Plot the Kolmogorov-Smirnoff (K-S) statistic along with the number of samples n for each of the 3 sampled distributions as n varies from 10^2 to 10^5 samples. Note that the empirical CDF can be directly generated as samples are drawn from the distribution. Average over multiple trials and suggest a general function ϵ for the expected K-S statistic as a function of samples n for a given distribution i.e. $E[D_n] \approx \epsilon(n)$.

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

2. Consider the annual rainfall in cm distributed as $\mathcal{N}(73, 20^2)$ for a region. Consider a model of the probability density for the intensity of the daily rainfall in a location : $f_Y(y) = \frac{1}{X} e^{-\frac{y}{X}}$, given the average daily rainfall in mm X at the location during the year. The yield of a particular crop can be estimated as $10P$ tonnes/acre where P is the probability of receiving rainfall between 2mm and 4mm in a day. Evaluate the expected yield/acre of the crop using two approaches for a comparison : (a) using an analytical expression, and (b) using random samples from $\mathcal{N}(0, 1)$ and $\mathcal{U}(0, 1)$.

   Note: Use logarithmic scales in plots wherever appropriate. Submit the responses, plots, and the code as separate files, all zipped into a single folder identified by your name in full, to *suvendukar@iisc.ac.in*