

DS 298: Work Assignment - 1

Due Feb 11, 2024

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 $[0, 1]$. Plot the Kolmogorov-Smirnoff (K-S) statistic along with the number of samples n in each of the 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. Also, generate a K-S statistic comparison table in the form of a 3×3 confusion matrix, for sample sizes 10^2 , 10^3 and 10^4 where each empirical distribution is compared with all the three given distributions. Use appropriate averaging over trials to generate an expected confusion matrix, if required. Suggest a general function ϵ to set cut-off values for the K-S statistic as a function of samples n for a given distribution i.e. $D_n \leq \epsilon(n)$ to confirm convergence of samples from an unknown origin, and justify it.
2. Consider the revenue model of a crop across geographical regions where the yield (Y) in tonnes/acre is distributed as $\mathcal{N}(3, 1)$. Consider a speculative local pricing model where the local price/tonne is an exponentially decreasing probability distribution of the local yield ($X = ae^{-\frac{Y}{b}}$), with the potential negative yields representing the loss of crops and its speculative effect on the local prices. Let the stable global market price (representing a zero local yield) be Rs. 5000/tonne and the price representing the mean local yield be Rs $5000e^{-1}$ /tonne. Using empirical distributions with sufficient number of samples, generate the probability distributions of the local price/tonne and the notional income/acre ($Z = XY$), and evaluate the expected income/acre ($E[XY]$ when $Y > 0$) for farmers across the samples representing geographical regions. Using transformation of random variables, you should also realize the probability distribution of prices (X) and its mean without an empirical sampling.

Note: Submit the responses, plots, and the code as separate files, all zipped into a single folder identified by your name in full, to *nagagn@iisc.ac.in* or upload in MS-Teams as specified in its *DS298* channel.