DS 298: Work Assignment - 1

Due Jan 30, 2023

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. In each case, use the cumulative distribution function (CDF) of the original distribution and the sampled distribution to evaluate the above. Note that the CDF can also be evaluated using numerical integration of the probability density function (PDF). 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 sampled 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 relation 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 a sporting league where the players' skills are distributed as a normal distribution $\mathcal{N}(\frac{1}{2}, \frac{1}{36})$. Assume that the probability of a player earning a call for a tournament is proportional to the player's skills, and the career points earned in that tournament are also likely to be proportional. Evaluate the distribution of career points S after the k^{th} tournament in the year, when all players start with an initial point of 1. This probability density in scores S can be evaluated using either a Monte Carlo simulation of S using greater than 10^5 samples of X, or using a relation for transforming the probability density f_X into f_S evaluated at less than 100 values of X in the interval [0,1]. You are expected to do the later and plot S as a function of the tournaments $k = 1, 2 \ldots 8$. Hint: consider a model $S = (1 + cX)^{ckX}$, where X is a random variable representing the skill level of the player, k is the tournament number and c be any constant denoting the maximum possible raise of points in a tournament; c can be set to 1 for the above problem.

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