Biswarup Karmakar(21055)

Assignment: DS298 Assignment1

Problem 2:

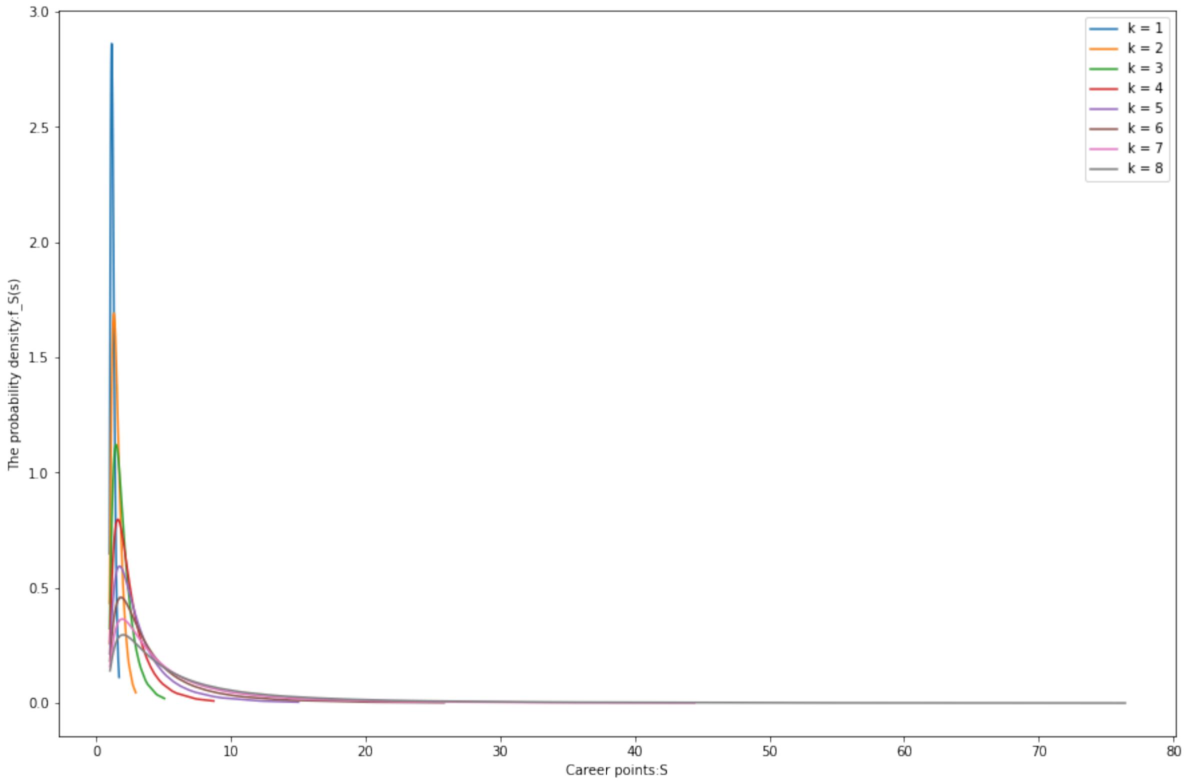
Given that X is a random variable representing the skill level of the player. We start by transforming the probability density of X into the probability density of S using the relation $S = (1 + X)^{kX}$.To do this, we use the change of variable method. The density of S, $f_S(s)$, can be expressed as:

$$f_S(s) = f_X(x) \times |\frac{dx}{ds}|$$

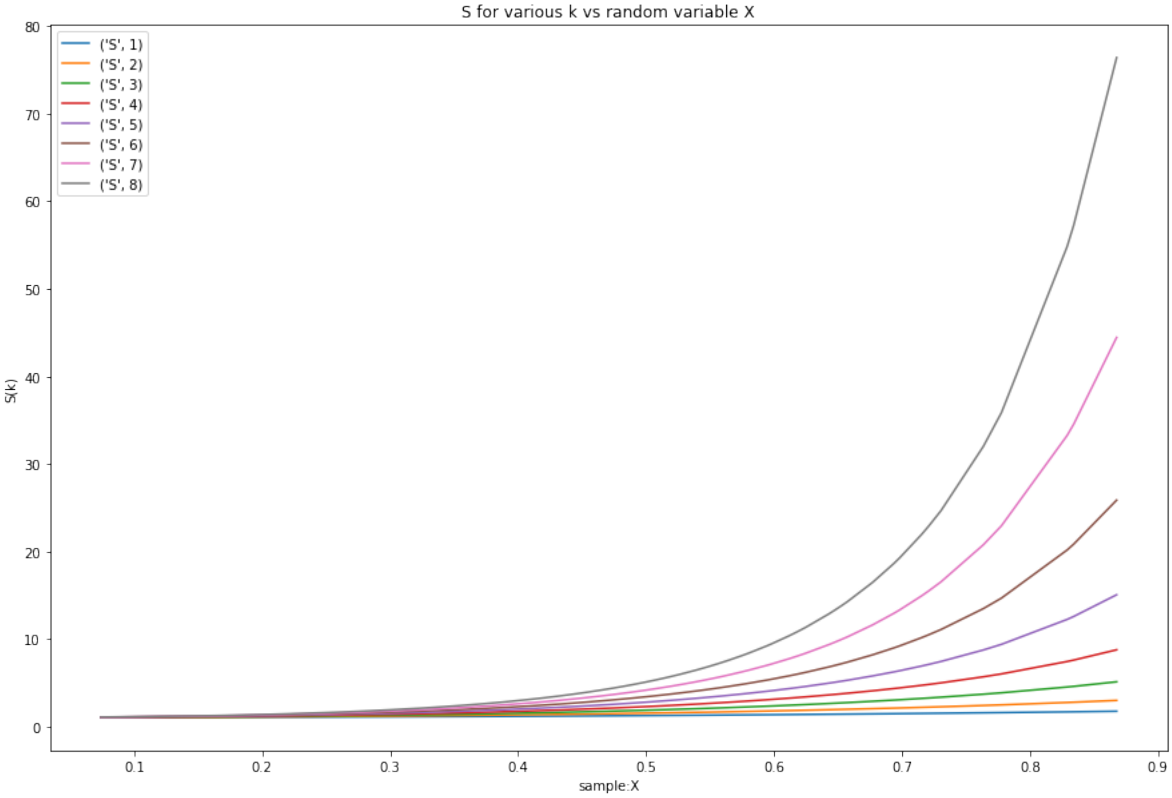
where x is a function of s and can be found by solving the equation $s = (1 + x)^{kx}$ for x. Taking the derivative of s with respect to x, we get:

$$\frac{ds}{dx} = kx(1 + x)^{(kx-1)} + (1 + x)^{kx} \times \ln(1 + x)$$

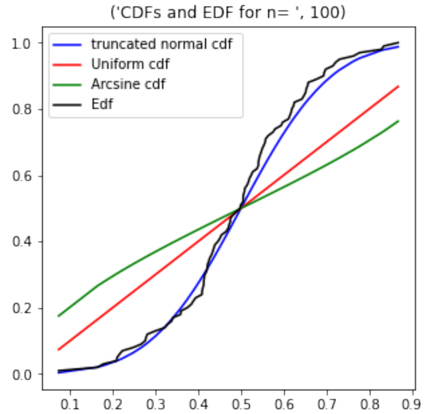
Then we Substitute this into the formula for $f_S(s)$.
After finding x as a function of s, the density of X, $f_X(x)$, can be evaluated at x for each value of s. Finally, we plot $f_S(s)$ as a function of s for each value of $k = 1, 2, ..., 8$.
The resulting plot shows the distribution of career points S after each tournament for $k = 1, 2, ..., 8$. As we can see, the distribution of S becomes more peaked and shifted to the right as the number of tournaments increases, indicating that players who perform well in the early tournaments have a higher probability of earning more career points.

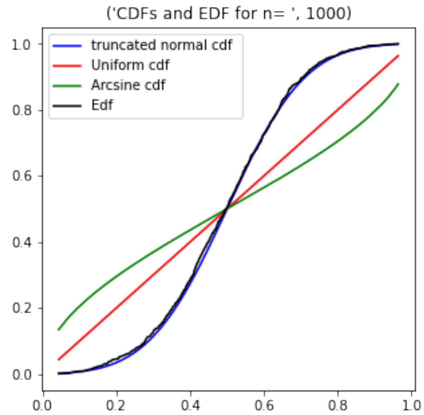


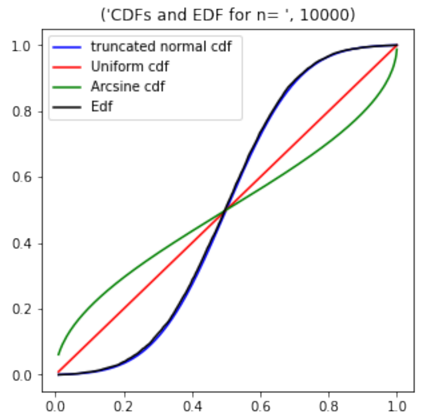
Now we see that career point with the different values of k increases when skill level of player increases.initially for $k = 1$ career point almost constant for different skil level .But for $k = 8$ the career point increases exponentially with increasing value of skill level $(X$.

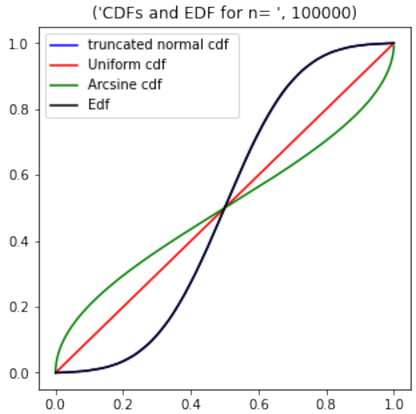


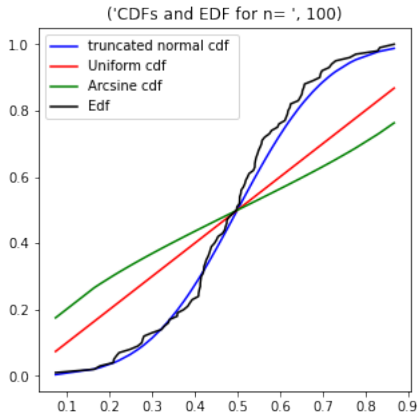
Problem1: Samples X are drawn from Truncated Normal Distribution $N(1/2, 1/36)$ and EDF is plotted along with CDFs corresponding to X assuming it to be drawn from Truncated Normal, Uniform and Arcsine distribution.Here we tested for 3 trials, and see that the EDF(empirical distribution function) nearly close to Truncated Normal CDF.Hence most likely the drawn samples are from Truncated Distribution.

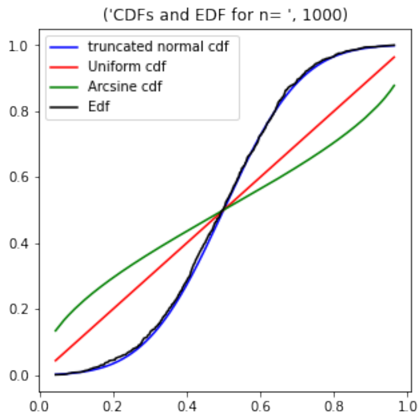


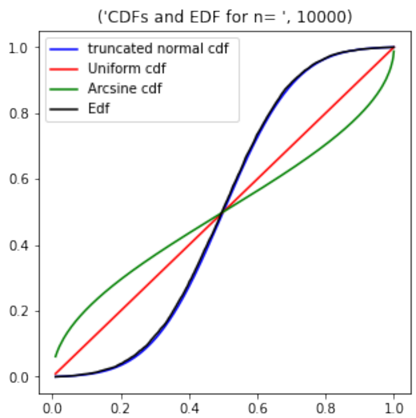


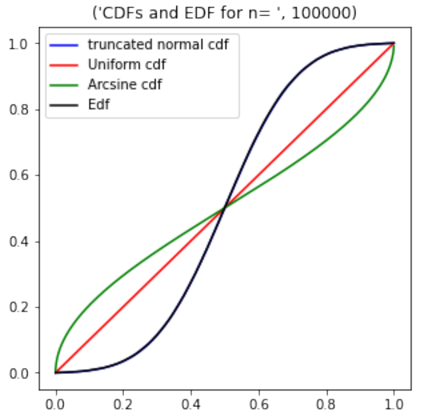


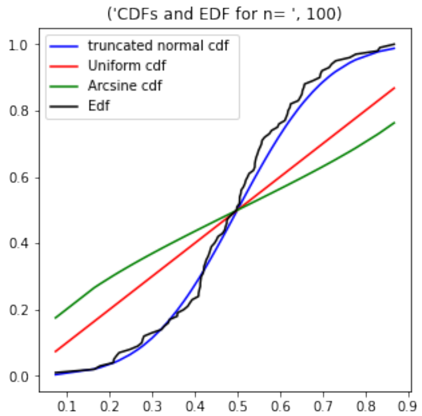

('CDFs and EDF for n= ', 100000)
truncated normal cdf
Uniform cdf
Arcsine cdf
Edf

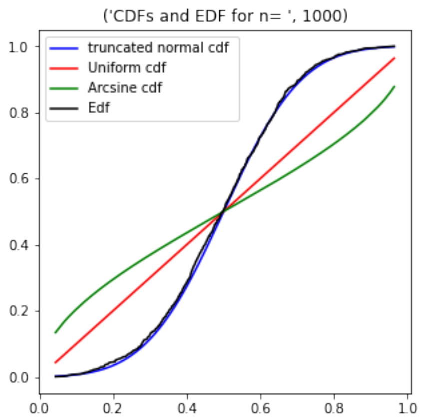

('CDFs and EDF for n= ', 100)
truncated normal cdf
Uniform cdf
Arcsine cdf
Edf

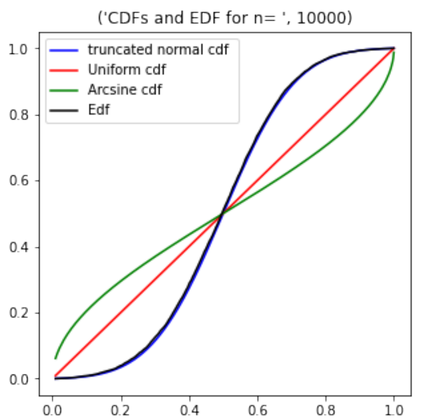

('CDFs and EDF for n= ', 1000)
truncated normal cdf
Uniform cdf
Arcsine cdf
Edf

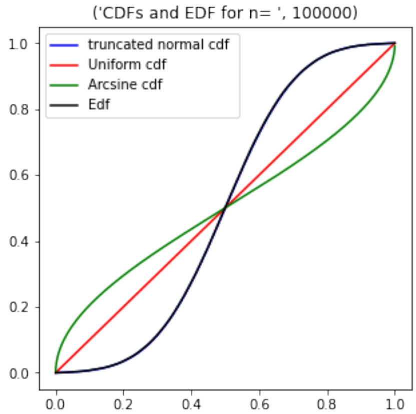

('CDFs and EDF for n= ', 10000)
truncated normal cdf
Uniform cdf
Arcsine cdf
Edf


('CDFs and EDF for n= ', 100000)
truncated normal cdf
Uniform cdf
Arcsine cdf
Edf

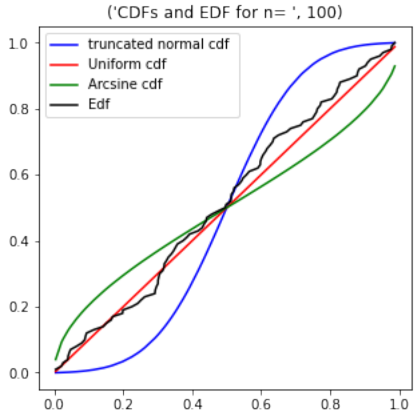

('CDFs and EDF for n= ', 100)
truncated normal cdf
Uniform cdf
Arcsine cdf
Edf

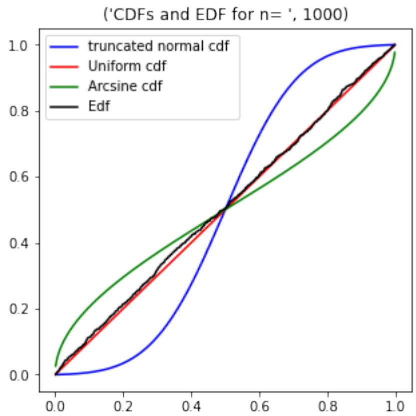

('CDFs and EDF for n= ', 1000)
truncated normal cdf
Uniform cdf
Arcsine cdf
Edf

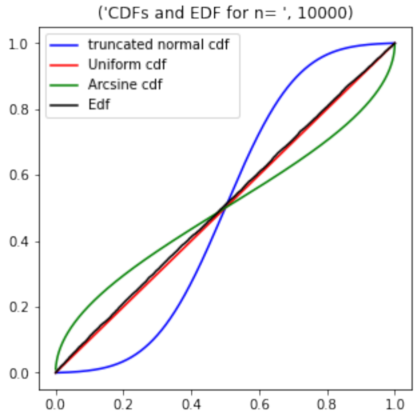

('CDFs and EDF for n= ', 10000)
truncated normal cdf
Uniform cdf
Arcsine cdf
Edf

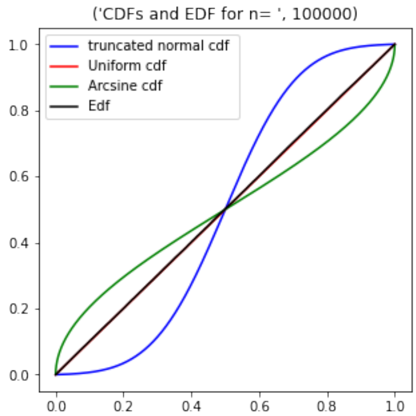


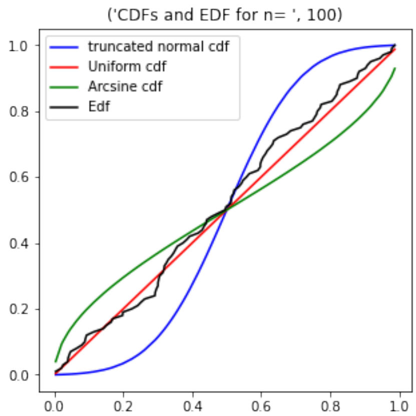
Samples X are drawn from Uniform Distribution $U(0,1)$ and EDF is plotted along with CDFs corresponding to X assuming it to be drawn from Truncated Normal, Uniform and Arcsine distribution.Here we tested for 3 trials, and see that the EDF(empirical distribution function) nearly close to Uniform CDF.Hence most likely the drawn samples are from Uniform Distribution.

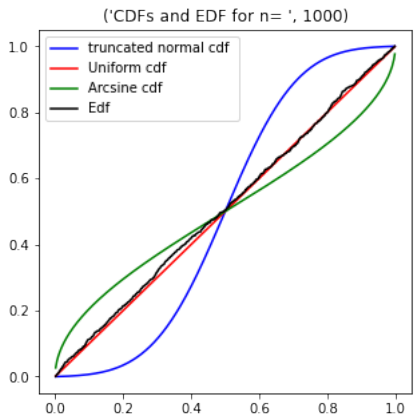


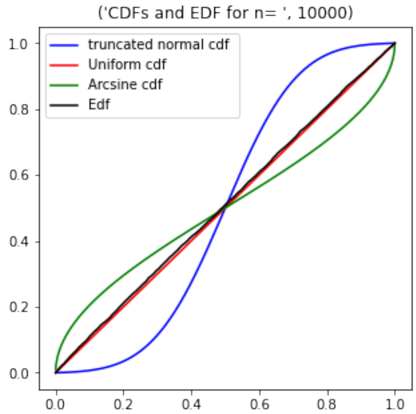


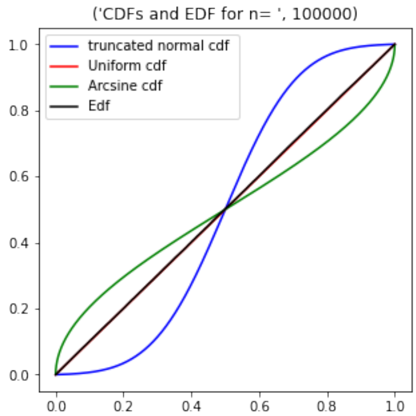

('CDFs and EDF for n= ', 10000)
truncated normal cdf
Uniform cdf
Arcsine cdf
Edf

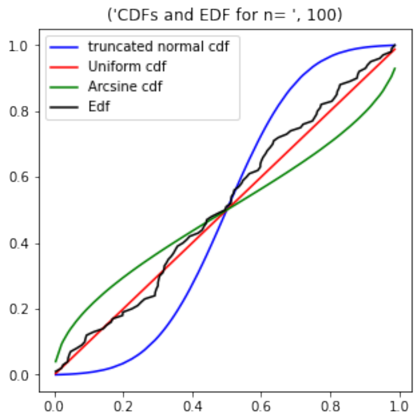

('CDFs and EDF for n= ', 100000)
truncated normal cdf
Uniform cdf
Arcsine cdf
Edf

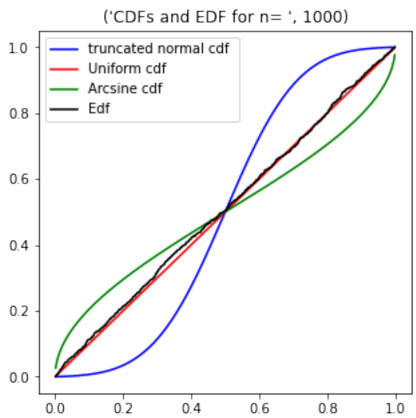

('CDFs and EDF for n= ', 100)
truncated normal cdf
Uniform cdf
Arcsine cdf
Edf

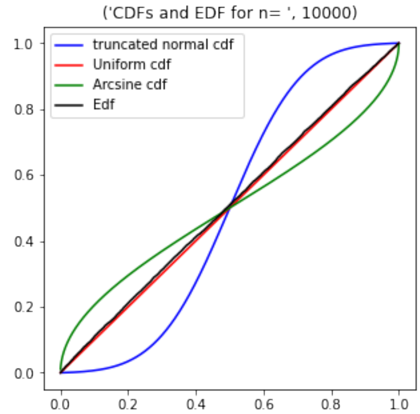

('CDFs and EDF for n= ', 1000)
truncated normal cdf
Uniform cdf
Arcsine cdf
Edf

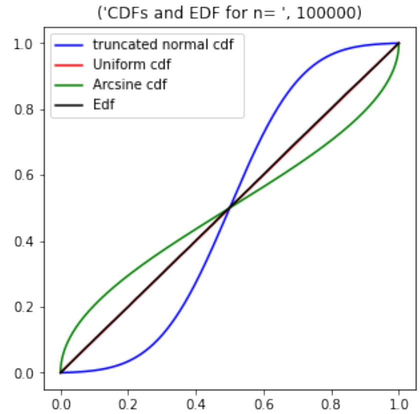

('CDFs and EDF for n= ', 10000)
truncated normal cdf
Uniform cdf
Arcsine cdf
Edf
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0


('CDFs and EDF for n= ', 100000)
truncated normal cdf
Uniform cdf
Arcsine cdf
Edf
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

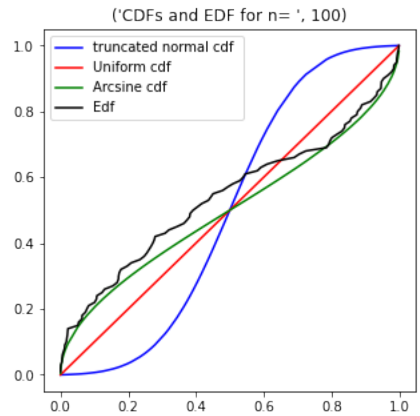

('CDFs and EDF for n= ', 100)
truncated normal cdf
Uniform cdf
Arcsine cdf
Edf
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

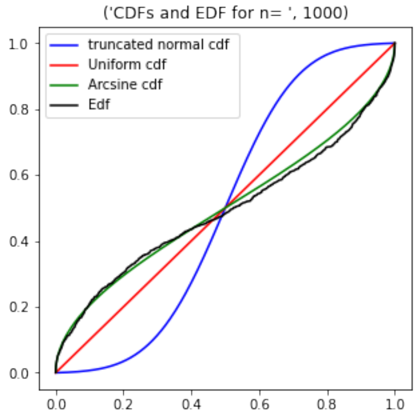

('CDFs and EDF for n= ', 1000)
truncated normal cdf
Uniform cdf
Arcsine cdf
Edf
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

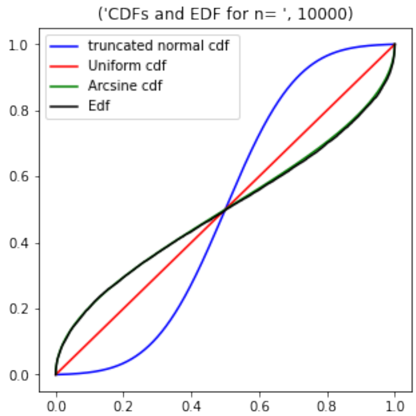


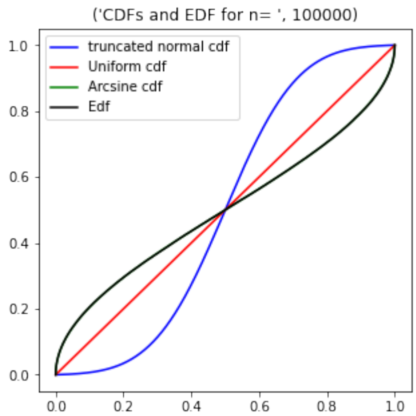


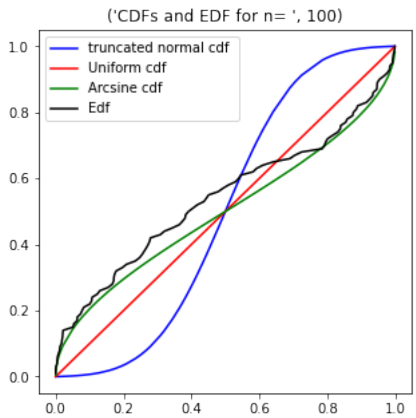
Samples X are drawn from Arcsine Distribution $Beta(1/2, 1/2)$ and EDF is plotted along with CDFs corresponding to X assuming it to be drawn from Truncated Normal, Uniform and Arcsine distribution.Here we tested for 3 trials, and see that the EDF(empirical distribution function) nearly close to Arcsine CDF.Hence most likely the drawn samples are from Arcsine Distribution.

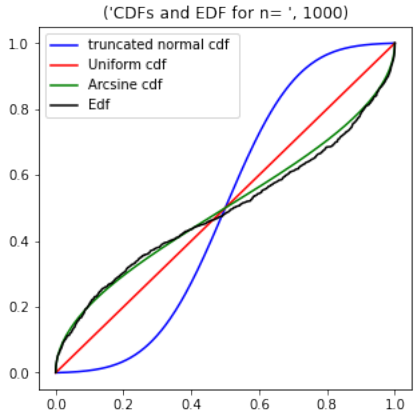


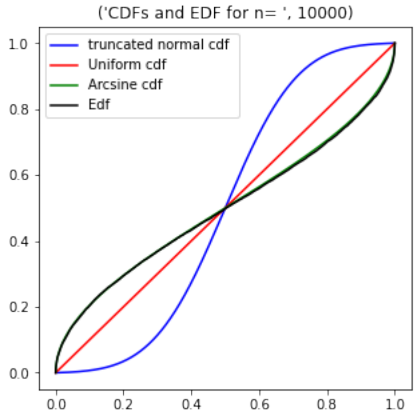

('CDFs and EDF for n= ', 1000)
truncated normal cdf
Uniform cdf
Arcsine cdf
Edf
0.0
0.2
0.4
0.6
0.8
1.0

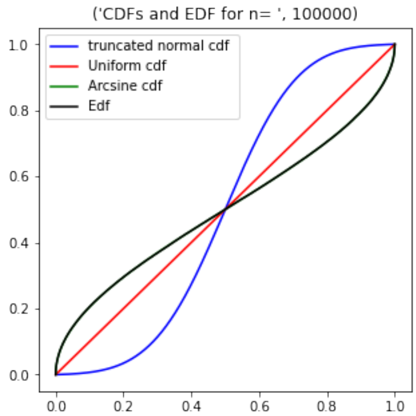

('CDFs and EDF for n= ', 10000)
truncated normal cdf
Uniform cdf
Arcsine cdf
Edf
0.0
0.2
0.4
0.6
0.8
1.0

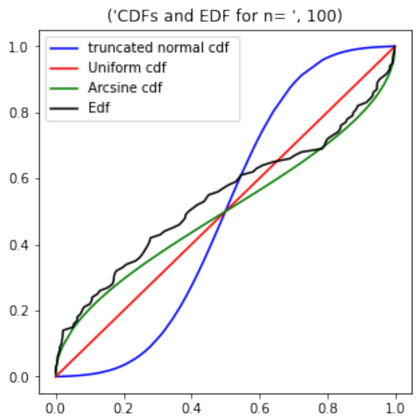

('CDFs and EDF for n= ', 100000)
truncated normal cdf
Uniform cdf
Arcsine cdf
Edf
0.0
0.2
0.4
0.6
0.8
1.0

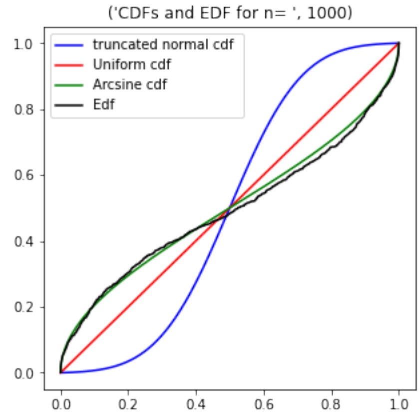

('CDFs and EDF for n= ', 100)
truncated normal cdf
Uniform cdf
Arcsine cdf
Edf
0.0
0.2
0.4
0.6
0.8
1.0

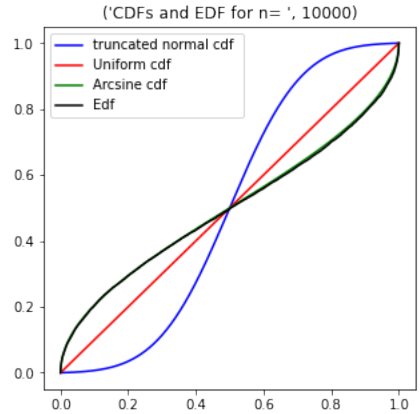

('CDFs and EDF for n= ', 1000)
truncated normal cdf
Uniform cdf
Arcsine cdf
Edf

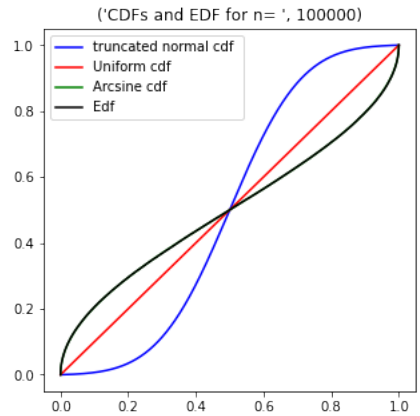

('CDFs and EDF for n= ', 10000)
truncated normal cdf
Uniform cdf
Arcsine cdf
Edf


('CDFs and EDF for n= ', 100000)
truncated normal cdf
Uniform cdf
Arcsine cdf
Edf


('CDFs and EDF for n= ', 100)
truncated normal cdf
Uniform cdf
Arcsine cdf
Edf







Here we use the power model relation between K-S statistics D and Sample size (n) given by

$$\epsilon(n) = C \times n^p$$

Where C and p are proportionality coefficient and power index respectively.
Taking log base 10 on both sides we get $\log_{10}(D) = \log_{10}(C) + p \times \log_{10}(n)$
Evaluating C and p based on different distribution using curve fitting,
For truncated normal distribution we got $\epsilon(n) = 0.07897475774365889 \times n^{-0.3066345554804011}$
For Uniform distribution we get $\epsilon(n) = 0.33203606128784174 \times n^{(-0.13316969540722856)}$.
For arcsine distribution we get $\epsilon(n) = 0.35788293117247916 \times n^{(-0.14863630206401224)}$

Plotting Average K-S statistics vs Sample size (n) on $\log_{10}$ scale for all 3 distributions, we can see that KS statistic decreases with the increasing value of sample size n.