Problem Set 1
PHY670: Astro Statistics
August - November 2025
IISER Mohali

1. Given the Bernouli distribution, plot the probability distribution for $p = 0.5$ and $N = 5$, 10, 20, and 100. Over plot the Gaussian/Normal distribution with the same average and standard deviation. Compute the probability of values outside the range of average $\pm$ standard deviation for each case and compare with the expectation from the Normal distribution.

2. Repeat the above problem for $p = 0.2$.

3. Repeat the above problem for $p = 0.01$ and 0.05. Also over plot the Poisson distribution in this case.

4. Use a pseudo random number generator that generates numbers uniformly distributed in the range $[0 - 1]$. Calculate the first four moments numerically and plot these as a function of the number of trials (length of the sequence of random numbers). Also plot the expected values of these moments on the same plots. Use the range $0 - 10^n$, with $n = 2$, 3, 4 and 6. What do we learn from these plots?

5. Use pseudo random numbers to simulate tossing of a fair coin. Plot the ratio of the number of heads to tails as a function of the total number of trials for $(10 \leq N \leq 10000$. Do this for an ensemble of tossing, essentially repeating the numerical experiment many times. What do you need to ensure that you get a fresh sequence of random numbers each time? The ensemble should have a size of at least 100. Plot all cases of the ensemble on the same plot.

6. Repeat the last problem but for a coin with probability of heads being 0.49 At what N can you clearly distinguish between the two sets of ensembles (last problem vs this problem)? Can you think of a simple graphical representation? How can we arrive at a similar deduction analytically?

7. You are provided a file with historical climate data for two cities: Delhi and Hyderabad. Compute the average quantities for each month for each decade (minimum temperature, maximum temperature, rainfall per day). For each decade, also calculate the standard deviation in each of these quantities as well as extreme values. Which distribution describes each of these quantities well?

8. Using the data from the last problem, do you see a systematic trend in any of the quantities between $1951 - 2020$?

9. You are given a file containing voltages recorded in observations by the Ooty radio telescope. The voltages are recorded for the same direction and at the same time by two halves of the telescope (North and South). Using the data:

   - Plot the distribution of voltages V_N, V_S recorded by the each part of the telescope separately. Which distribution describes this well?
   - Now plot the distribution for $V_N - V_S$, Clearly, any contribution from a source is eliminated in this. Use this to estimate the characteristics of noise in the radio telescope.
   - Average the data over 25, 100 and 400 data points and recompute the distribution. Interpret.