Download the template Jupyter notebook `HW1_Template.ipynb` from Canvas and work from that template. For Problem 1, you can write your solutions down either 1) in Jupyter using markdown and Latex or 2) handwritten on paper and scanned/submitted to Canvas. You should be able to complete Problems 2 – 4 using only `numpy` and `matplotlib`, but feel free to use whatever you are comfortable with (e.g., `scipy` and `pandas`). We will cover `pandas` and make more extensive use of it in the coming weeks.

1. (4 pts) For a uniform distribution $p_U(x; \mu, W)$ of mean μ and width W.

$$p_U(x; \mu, W) = \begin{cases} 1/W & \text{if } |x - \mu| \leq W/2 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

   calculate the variance σ^2, skewness Σ, and kurtosis K.

2. **CODING:** (5 pts) Verify the results of Problem 1 by generating an array of random deviates and computing the various moments.

   (a) Write a function that returns an array of random deviates uniformly distributed according to the probability distribution function $p_U(x; \mu, W)$.

   (b) Write a function that computes the mean, variance, skewness, and kurtosis of a given array of values.

   (c) Use these two functions to verify that for sufficiently large N, the empirical values closely match the theoretical values you calculated.

3. **CODING:** (7 pts) The datafile `SATscores.csv` contains the SAT scores of 1987 high school students in a hypothetical school district. The columns are `ID` and `score`.

   (a) Write a function that calculates the mean and variance of the scores.

   (b) Make a plot of the histogram of the scores in bins of 50 points (0-50, 51-100, ... 1551-1600).

   (c) On the histogram from part b), overplot an appropriately normalized (by eye) gaussian distribution function with the mean and variance found in part a). Based on your visual comparison, can you conclude with confidence that the observed distribution is gaussian? We will approach this more quantitatively later in the course.

4. **CODING:** (4 pts) This problem demonstrates empirically several properties of counting statistics, which obeys the Poisson distribution. We will approximate a Poisson process by generating random values between 0 and 10, but only looking at the ones between 0 and 1. Technically, this is a multinomial distribution, but for large N and at least moderately large number of bins, it is a good approximation to a Poisson process.

   (a) Generate $N = 10^6$ uniform random deviates between 0 and 10. Count the number of deviates that fall in each of 100 equal-sized bins between 0 and 1 (0.00-0.01, 0.01-0.02, etc). Calculate the mean μ and variance σ^2 of the counts per bin. Make sure that the mean is almost exactly what you expect: $\bar{n} \approx N/100 = 1000$.

   (b) Repeat part a) for 10 and 1000 equally-sized bins. This empirically shows that $\mu = \sigma^2 = \bar{n}$, which is an important property of the Poisson distribution.

(c) **EXTRA CREDIT:** (2 pts) Now generate $M = 100$ realizations of part a) and calculate the mean and variance of the counts per bin each time using 100 equally-sized bins. Make sure to not use the same random seed each time. You should now have $M = 100$ values of both the mean and variance. Let's call these μ_i and σ_i^2, respectively, where $i = 1, 2, ..., M$. Now calculate the mean and variance of μ_i and σ_i^2. This demonstrates that the general property that the variance (2nd moment) is much more difficult to measure accurately than the mean (1st moment).