Due: March 14th, 2024,10pm
Problems Due: 2(b), 2(c), 3(b), 3(d)

1. Consider Poisson λ distribution.
   (a) Show that both the sample mean and the sample variance of a sample obtained from the Poisson(λ) distribution will be unbiased estimators of λ.
   (b) For $\lambda = 10, 20, 50$ simulate $100, 500, 1000$ random observations from the Poisson(λ) distribution for various values of λ using the inbuilt function `rpois`.
   (c) Explore the behaviour of the two estimates for each λ as well as three sample sizes.

2. Suppose $a, b \in \mathbb{R}$ and $f : [a, b] \to \mathbb{R}$ is a continuous function. We wish to compute the
$$\int_a^b f(x)dx$$
.
   (a) Suppose we generate samples of $X_1, X_2, \ldots X_n$ i.i.d. Uniform (a, b). Justify, using the weak law of large numbers, that for $\epsilon > 0$, we have
$$P(|\ \frac{1}{n}\sum_{i=1}^{n} f(X_i) - \frac{1}{b-a}\int_a^b f(x)dx\ | > \epsilon) \to 0,$$
   as $n \to \infty$.
   (b) Write an `R`-code that takes in an input n and:
      i. Generate n samples $X_1, X_2, \ldots, X_n$ of Uniform $(0, 7)$
      ii. Computes $I_n = \frac{1}{n}\sum_{i=1}^{n} f(X_i)$, when $f(x) = \frac{16+\sin(x)}{x^2+4}$
      iii. Suppose $n = 50$ in, simluate 100 samples of I_n and use the mean of these 100 simulations as an estimate of $\int_0^7 \frac{16+\sin(x)}{x^2+4}dx$.
   (c) Use the `integrate` command in `R` to evaluate the integral. Compare the answer with the estimate obtained above.

3. A Biologist decides to use a technique called "capture-recapture" to estimate the number of fish in a pond. Suppose the unknown population size is N, and fifty fish are selected and an identifying mark is placed on them. Sometime later a random sample of size twenty is taken from the population, and the sample is found to contain X of the twenty previously marked.
   (a) Find the distribution of X (it will involve N as a parameter).
   (b) In `R`, using the appropriate inbuilt function `r...` generate 100 samples of X given $N = 50$.
   (c) The biologist plans to equate the proportion of marked members in the sample and the population sample, to get an estimate $\hat{N}$ of N. Write $\hat{N}$ as a function of X.
   (d) For each $N = 50$, 100, 200, 300, 400, and 500, simulate 1000 values of X and compute $\hat{N}$ for each trial. Using these provide an estimate for $E[\hat{N}]$ and $Var[\hat{N}]$. Plot these estimates as a function of N.

Book-Keeping Exercises

1. Let N, r, and m be positive integers for which $r < N$ and $m < N$. Let X be a random variable taking values in the integers between $\max\{0, m-(N-r)\}$ and $\min\{m, r\}$ inclusive with probability mass function
$$P(X = k) = \frac{\binom{r}{k}\binom{N-r}{m-k}}{\binom{N}{m}}$$
The random variable X is called hypergeoemtric with parameters N, r, and m. Find $E[X]$ and $Var[X]$.

2. Let $X_1, X_2, X_3, \ldots, X_n$ be i.i.d. random variables with finite expectation μ, finite variance σ^2, and finite $\gamma = E[X_1 - \mu]^4$. Compute $Var(S^2)$ in terms of μ, σ^2, and γ and show that $Var(S^2) \to 0$ as $n \to \infty$.

3. Let $X_1, X_2, X_3, \ldots, X_n$ be i.i.d. random variables with finite expectation μ and finite variance σ^2. let $S = \sqrt{S^2}$, the non-negative root of the sample variance. The quantity S is called the "sample standard deviation". Although $E[S^2] = \sigma^2$, it is not true that $E[S] = \sigma$. In other words, S is not an unbiased estimator for σ. Follow the steps below to see why.
   (a) Let Z be a random variable with finite mean and finite variance. Prove that $E[Z^2] \geq E[Z]^2$ and give an example to show that equality may not hold. (Hint: Consider how these quantities relate to the variance of Z).
   (b) Use (a) to explain why $E[S] \leq \sigma$ and give an example to show that equality may not hold.

Apple Bread Challenge[1]

Challenge 1: *Due date: March 16th, 2024, 10pm (via email, in pdf format)*

A random walk on $\mathbb{Z}^2$, can be thought of as a position of particle, at time n, denoted by S_n which is defined inductively below. Let $x \in \mathbb{Z}^2, n \geq 1$.

- Let $S_0 = x$
- Suppose we have defined $S_1, S_2, \ldots S_{n-1}$.
- At time n, the particle chooses one of the four neighbours of the vertex S_{n-1} at random and moves to that neighbour which we denote by S_n.

Simulate and plot the range of the random walk on Z^2 for:

(a) $n = 10$ and $x = 0$ (origin)

(b) $n = 100$ and $x = 0$ (origin)

(c) $x = 0$ and till the first time the walk hits a point at a distance 55 from the origin.

(d) $x = (55, 0)$ and till the first time the walk hits the origin.

[1]This will be a multi-part challenge each with a specific due date. If I get a complete solution to the multi-part challenge [From 5 students] by due dates then I will bake apple bread for everyone in class who has 90% attendance.