Problems due: 1,2

1. Suppose p is the unknown probability of an event A, and we estimate p by the sample proportion $\hat{p}$ based on an i.i.d. sample of size n.

   (a) Design and implement the following simulation study to verify this behaviour. For $p = 0.01$, 0.1, 0.25, 0.5, 0.75, 0.9, and 0.99,

   (i) Simulate 1000 values of $\hat{p}$ with $n = 500$.

   (ii) Simulate 1000 values of $\hat{p}$ with n chosen according to the formula derived above.

   In each case, you can think of the 1000 values as i.i.d. samples from the distribution of $\hat{p}$, and use the sample standard deviation as an estimate of $SD[\hat{p}]$. Plot the estimated values of $SD(\hat{p})$ against p for both choices of n.

2. 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.

3. Biologists use a technique called "capture-recapture" to estimate the size of the population of a species that cannot be directly counted.

   Suppose the unknown population size is N, and fifty members of the species are selected and given an identifying mark. Sometime later a sample of size twenty is taken from the population, and it is found to contain X of the twenty previously marked. Equating the proportion of marked members in the second sample and the population, we have $\frac{X}{20} = \frac{50}{N}$, giving an estimate of $\hat{N} = \frac{1000}{X}$.

   (a) Show that the distribution of X has a hypergeometric distribution that involves N as a parameter.

   (b) Using the function `rhyper`. For each $N = 50$, 100, 200, 300, 400, and 500, simulate 1000 values of $\hat{N}$ and use them to estimate $E[\hat{N}]$ and $Var[\hat{N}]$. Plot these estimates as a function of N.

4. Suppose p is the unknown probability of an event A, and we estimate p by the sample proportion $\hat{p}$ based on an i.i.d. sample of size n.

   (a) Write $Var[\hat{p}]$ and $SD[\hat{p}]$ as functions of n and p.

   (b) Using the relations derived above, determine the sample size n, as a function of p, that is required to acheive $SD(\hat{p}) = 0.01$. How does this required value of n vary with p?