Exercise 1: Let Y be a $Gamma(n, \lambda)$-distributed random variable with $n \geq 2$. Consequently, Y has the density

$$f(x) = \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} 1_{(0,\infty)}(x).$$

Furthermore, let $X_1, \ldots, X_n$ be a random sample with $X_1 \sim Exp(\lambda)$.

a) Find $E\left[\frac{1}{Y}\right]$

b) Find the Expectation of $\overline{X}$ as well as the expectation of $\frac{1}{\overline{X}}$. Is $\frac{1}{\overline{X}}$ an unbiased estimator for λ? (Hint: Sheet 5 Exercise 2.)

c) For which a_n is $\frac{a_n}{\overline{X}}$ an unbiased estimator for λ?

Solution:

a) We can show that

$$\begin{aligned} E\left[\frac{1}{Y}\right] &= \int_0^\infty \frac{1}{y} \lambda^n y^{n-1} e^{-\lambda y} \frac{1}{(n-1)!} dy \\ &= \frac{\lambda}{n-1} \int_0^\infty \lambda^{n-1} y^{n-2} e^{-\lambda y} \frac{1}{(n-2)!} dy \\ &= \frac{\lambda}{n-1} \end{aligned}$$

b) For the expectation of $\overline{X}$ we have

$$E[\overline{X}] = \frac{1}{n} \sum_{i=1}^n E[X_i] = \frac{1}{\lambda}.$$

We know that $\sum_{i=1}^n X_i \sim Gamma(n, \lambda)$. Thus,

$$E\left[\frac{1}{\overline{X}}\right] = nE\left[\frac{1}{\sum_{i=1}^n X_i}\right] = \frac{n}{n-1}\lambda$$

and $\frac{1}{\overline{X}}$ is not an unbiased estimator for λ.

c) With $a_n = \frac{n-1}{n}$ we know that $\frac{a_n}{\overline{X}}$ is an unbiased estimator for λ.

Exercise 2: We consider the random sample (X_1, X_2, X_3).

a) Find expressions for

  a) $Y_1 = \frac{1}{4}(X_1 + 3X_2)$

  b) $Y_2 = \frac{1}{3}(X_1 + X_2 + X_3)$

  c) $Y_3 = \frac{1}{9}(2X_1 + 3X_2 + 4X_3)$

  only using the variance an the expectation of X_1.

b) Which of the statistics Y_1, Y_2, Y_3 are unbiased estimators for the expectation and the population?

c) Which estimator you would use if you had to estimate the expectation?

Solution:

a) a) $E[Y_1] = E[X_1]$ and $\text{Var}(Y_1) = \frac{10}{16}\text{Var}(X_1)$

   b) $E[Y_2] = E[X_1]$ and $\text{Var}(Y_2) = \frac{1}{3}\text{Var}(X_1)$

   c) $E[Y_3] = E[X_1]$ and $\text{Var}(Y_3) = \frac{29}{81}\text{Var}(X_1)$

b) all

c) Y_2 (smallest variance = lowest risk)

Exercise 3: Like in Exercise 2 we consider a random sample (X_1, X_2, X_3) and let Y_2 be like in Exercise 2 ($Y_2 = \frac{1}{3}(X_1 + X_2 + X_3)$). Compute the variance and the expectation of

a) $S_1 = \frac{1}{3}\{(X_1 - Y_2)^2 + (X_2 - Y_2)^2 + (X_3 - Y_2)^2\}$

b) $S_2 = \frac{1}{2}\{(X_1 - Y_2)^2 + (X_2 - Y_2)^2 + (X_3 - Y_2)^2\}$

c) Which of the estimators S_1 and S_2 is an unbiased estimator for $\text{Var}(X_1)$?

Solution:

a) $E[S_1] = \frac{2}{3}\text{Var}(X_1)$

b) $E[S_2] = \text{Var}(X_1)$

c) S_2

You do not have to compute the variances.

Exercise 4: The quality control of a company checks n products for their functionality.

a) Based on an appropriate sample $(X_1, \ldots, X_n)$ give an estimator for the fraction of faultily products of the whole production of the company.

b) Verify whether the chosen estimator is unbiased and calculate the variance.

Solution:

a) $\hat{\pi} = Y/n$ where Y is the number of faultily products in the sample

b)

$$E[\hat{\pi}] = \frac{1}{n} np = p$$

p is the fraction of faultily products in the whole production. $Y \sim B(n, p)$

$$\text{Var}(\hat{\pi}) = \frac{1}{n^2}\text{Var(Y)} = \frac{p(1-p)}{n}$$

since $Y \sim B(n, p)$ and therefore $\text{Var}(Y) = np(1-p)$

Exercise 5: Assume you want to determine the number of fishes in a lake, denoted as N. In order to do so, you first angle k fishes, mark them and throw them back in the lake. Now you fish until you catch a marked fish. Let X_1 be the number of fishes (after having caught the fishes you throw them back in the lake.) You repeat this procedure n-times which yields the results $X_2, \ldots, X_n$. Assume that the probability p for catching a marked fish is $p = k/N$.

a) Using only X_1 find an unbiased estimator for N and compute its variance.

b) Now using $X_1, \ldots, X_n$ find an unbiased estimator for N and compute its variance.

c) Which estimator is better? a) or b)?

Solution:

a) Due to the experimental set-up we know that $X_1, X_2, \ldots, X_n \sim Geo(p)$ are independent and identically distributed. Since $N = k/p$ we know that $k \cdot X_1$ is an unbiased estimator for N because

$$E[k \cdot X_1] = kE[X_1] = \frac{k}{p} = N.$$

The variance of the estimator is

$$\text{Var}(kX_1) = k^2 \text{Var}(X_1) = \frac{k^2(1-p)}{p^2}$$

b) $k\frac{X_1+\ldots+X_n}{n}$ is also an unbiased estimator for N since

$$E\left[k\frac{X_1 + \ldots + X_n}{n}\right] = \frac{k}{n}\sum_{i=1}^{n} E[X_i] = \frac{k}{p} = N.$$

Due to the independency, the variance of this estimator is given by

$$\text{Var}\left(k\frac{X_1 + \ldots + X_n}{n}\right) = \frac{k^2}{n^2}\sum_{i=1}^{n} \text{Var}(X_i) = \frac{k^2(1-p)}{np^2}$$

c) Both estimators are unbiased but the second one is the better one because its variance is smaller.