19

Unbiased estimators

In Chapter 17 we saw that a dataset can be modeled as a realization of a random sample from a probability distribution and that quantities of interest correspond to features of the model distribution. One of our tasks is to use the dataset to estimate a quantity of interest. We shall mainly deal with the situation where it is modeled as one of the parameters of the model distribution or as a certain function of the parameters. We will first discuss what we mean exactly by an *estimator* and then introduce the notion of *unbiasedness* as a desirable property for estimators. We end the chapter by providing unbiased estimators for the expectation and variance of a model distribution.

19.1 Estimators

Consider the arrivals of packages at a network server. One is interested in the intensity at which packages arrive on a generic day and in the percentage of minutes during which no packages arrive. If the arrivals occur completely at random in time, the arrival process can be modeled by a Poisson process. This would mean that the number of arrivals during one minute is modeled by a random variable having a Poisson distribution with (unknown) parameter μ. The intensity of the arrivals is then modeled by the parameter μ itself, and the percentage of minutes during which no packages arrive is modeled by the probability of zero arrivals: $\mathrm{e}^{-\mu}$. Suppose one observes the arrival process for a while and gathers a dataset $x_1, x_2, \ldots, x_n$, where x_i represents the number of arrivals in the ith minute. Our task will be to estimate, based on the dataset, the parameter μ and a function of the parameter: $\mathrm{e}^{-\mu}$.

This example is typical for the general situation in which our dataset is modeled as a realization of a random sample $X_1, X_2, \ldots, X_n$ from a probability distribution that is completely determined by one or more parameters. The parameters that determine the model distribution are called the *model parameters*. We focus on the situation where the quantity of interest corresponds to a feature of the model distribution that can be described by the model parameters themselves or by some function of the model parameters. This distribution feature is referred to as the *parameter of interest*. In discussing this general setup we shall denote the parameter of interest by the Greek letter θ. So, for instance, in our network server example, μ is the model parameter. When we are interested in the arrival intensity, the role of θ is played by the parameter μ itself, and when we are interested in the percentage of idle minutes the role of θ is played by $\mathrm{e}^{-\mu}$.

Whatever method we use to estimate the parameter of interest θ, the result depends only on our dataset.

> ESTIMATE. An *estimate* is a value t that only depends on the dataset $x_1, x_2, \ldots, x_n$, i.e., t is some function of the dataset only:
>
> $$t = h(x_1, x_2, \ldots, x_n).$$

This description of *estimate* is a bit formal. The idea is, of course, that the value t, computed from our dataset $x_1, x_2, \ldots, x_n$, gives some indication of the "true" value of the parameter θ. We have already met several estimates in Chapter 17; see, for instance, Table 17.2. This table illustrates that the value of an estimate can be anything: a single number, a vector of numbers, even a complete curve.

Let us return to our network server example in which our dataset $x_1, x_2, \ldots, x_n$ is modeled as a realization of a random sample from a $Pois(\mu)$ distribution. The intensity at which packages arrive is then represented by the parameter μ. Since the parameter μ is the expectation of the model distribution, the law of large numbers suggests the sample mean $\bar{x}_n$ as a natural estimate for μ. On the other hand, the parameter μ also represents the variance of the model distribution, so that by a similar reasoning another natural estimate is the sample variance s_n^2.

The percentage of idle minutes is modeled by the probability of zero arrivals. Similar to the reasoning in Section 13.4, a natural estimate is the relative frequency of zeros in the dataset:

$$\frac{\text{number of } x_i \text{ equal to zero}}{n}.$$

On the other hand, the probability of zero arrivals can be expressed as a function of the model parameter: $\mathrm{e}^{-\mu}$. Hence, if we estimate μ by $\bar{x}_n$, we could also estimate $\mathrm{e}^{-\mu}$ by $\mathrm{e}^{-\bar{x}_n}$.

QUICK EXERCISE 19.1 Suppose we estimate the probability of zero arrivals $\mathrm{e}^{-\mu}$ by the relative frequency of x_i equal to zero. Deduce an estimate for μ from this.

The preceding examples illustrate that one can often think of several estimates for the parameter of interest. This raises questions like

- When is one estimate better than another?
- Does there exist a best possible estimate?

For instance, can we say which of the values $\bar{x}_n$ or s_n^2 computed from the dataset is closer to the "true" parameter μ? The answer is *no*. The measurements and the corresponding estimates are subject to randomness, so that we cannot say anything with certainty about which of the two is closer to μ. One of the things we can say for each of them is *how likely* it is that they are within a given distance from μ. To this end, we consider the random variables that correspond to the estimates. Because our dataset $x_1, x_2, \ldots, x_n$ is modeled as a realization of a random sample $X_1, X_2, \ldots, X_n$, the estimate t is a realization of a random variable T.

ESTIMATOR. Let $t = h(x_1, x_2, \ldots, x_n)$ be an estimate based on the dataset $x_1, x_2, \ldots, x_n$. Then t is a realization of the random variable

$$T = h(X_1, X_2, \ldots, X_n).$$

The random variable T is called an *estimator*.

The word *estimator* refers to the method or device for estimation. This is distinguished from *estimate*, which refers to the actual value computed from a dataset. Note that estimators are special cases of sample statistics. In the remainder of this chapter we will discuss the notion of *unbiasedness* that describes to some extent the behavior of estimators.

19.2 Investigating the behavior of an estimator

Let us continue with our network server example. Suppose we have observed the network for 30 minutes and we have recorded the number of arrivals in each minute. The dataset is modeled as a realization of a random sample $X_1, X_2, \ldots, X_n$ of size $n = 30$ from a $Pois(\mu)$ distribution. Let us concentrate on estimating the probability p_0 of zero arrivals, which is an unknown number between 0 and 1. As motivated in the previous section, we have the following possible estimators:

$$S = \frac{\text{number of } X_i \text{ equal to zero}}{n} \quad \text{and} \quad T = \mathrm{e}^{-\bar{X}_n}.$$

Our first estimator S can only attain the values $0, \frac{1}{30}, \frac{2}{30}, \ldots, 1$, so that in general it *cannot* give the exact value of p_0. Similarly for our second estimator T, which can only attain the values $1, \mathrm{e}^{-1/30}, \mathrm{e}^{-2/30}, \ldots$. So clearly, we cannot expect our estimators always to give the exact value of p_0 on basis of 30 observations. Well, then what *can* we expect from a reasonable estimator?

To get an idea of the behavior of both estimators, we pretend we know μ and we simulate the estimation process in the case of $n = 30$ observations. Let us choose $\mu = \ln 10$, so that $p_0 = e^{-\mu} = 0.1$. We draw 30 values from a Poisson distribution with parameter $\mu = \ln 10$ and compute the value of estimators S and T. We repeat this 500 times, so that we have 500 values for each estimator. In Figure 19.1 a frequency histogram[1] of these values for estimator S is displayed on the left and for estimator T on the right. Clearly, the values of both estimators vary around the value 0.1, which they are supposed to estimate.

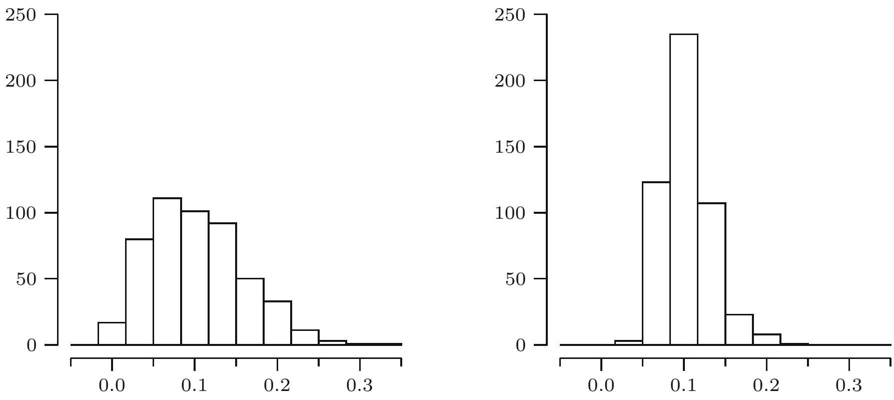

Fig. 19.1. Frequency histograms of 500 values for estimators S (left) and T (right) of $p_0 = 0.1$.

19.3 The sampling distribution and unbiasedness

We have just seen that the values generated for estimator S fluctuate around $p_0 = 0.1$. Although the value of this estimator is not always equal to 0.1, it is desirable that on average, S is on target, i.e., $\mathrm{E}[S] = 0.1$. Moreover, it is desirable that this property holds no matter what the actual value of p_0 is, i.e.,

$$\mathrm{E}[S] = p_0$$

irrespective of the value $0 < p_0 < 1$. In order to find out whether this is true, we need the probability distribution of the estimator S. Of course this

[1] In a frequency histogram the height of each vertical bar equals the frequency of values in the corresponding bin.

is simply the distribution of a random variable, but because estimators are constructed from a random sample $X_1, X_2, \ldots, X_n$, we speak of the sampling distribution.

THE SAMPLING DISTRIBUTION. Let $T = h(X_1, X_2, \ldots, X_n)$ be an estimator based on a random sample $X_1, X_2, \ldots, X_n$. The probability distribution of T is called the *sampling distribution* of T.

The sampling distribution of S can be found as follows. Write

$$S = \frac{Y}{n},$$

where Y is the number of X_i equal to zero. If for each i we label $X_i = 0$ as a success, then Y is equal to the number of successes in n independent trials with p_0 as the probability of success. Similar to Section 4.3, it follows that Y has a $Bin(n, p_0)$ distribution. Hence the sampling distribution of S is that of a $Bin(n, p_0)$ distributed random variable divided by n. This means that S is a discrete random variable that attains the values k/n, where $k = 0, 1, \ldots, n$, with probabilities given by

$$p_S\left(\frac{k}{n}\right) = \mathrm{P}\left(S = \frac{k}{n}\right) = \mathrm{P}(Y = k) = \binom{n}{k} p_0^k (1 - p_0)^{n-k}.$$

The probability mass function of S for the case $n = 30$ and $p_0 = 0.1$ is displayed in Figure 19.2. Since $S = Y/n$ and Y has a $Bin(n, p_0)$ distribution, it follows that

$$\mathrm{E}[S] = \frac{\mathrm{E}[Y]}{n} = \frac{np_0}{n} = p_0.$$

So, indeed, the estimator S for p_0 has the property $\mathrm{E}[S] = p_0$. This property reflects the fact that estimator S has no systematic tendency to produce

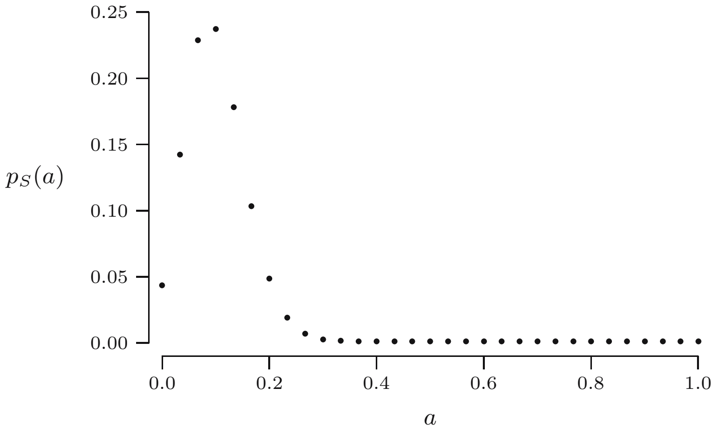

Fig. 19.2. Probability mass function of S.

estimates that are larger than p_0, and no systematic tendency to produce estimates that are smaller than p_0. This is a desirable property for estimators, and estimators that have this property are called unbiased.

DEFINITION. An estimator T is called an *unbiased* estimator for the parameter θ, if

$$\mathrm{E}[T] = \theta$$

irrespective of the value of θ. The difference $\mathrm{E}[T] - \theta$ is called the *bias* of T; if this difference is nonzero, then T is called *biased.*

Let us return to our second estimator for the probability of zero arrivals in the network server example: $T = \mathrm{e}^{-\bar{X}_n}$. The sampling distribution can be obtained as follows. Write

$$T = \mathrm{e}^{-Z/n},$$

where $Z = X_1 + X_2 + \cdots + X_n$. From Exercise 12.9 we know that the random variable Z, being the sum of n independent $Pois(\mu)$ random variables, has a $Pois(n\mu)$ distribution. This means that T is a discrete random variable attaining values $\mathrm{e}^{-k/n}$, where $k = 0, 1, \ldots$ and the probability mass function of T is given by

$$p_T\left(\mathrm{e}^{-k/n}\right) = \mathrm{P}\left(T = \mathrm{e}^{-k/n}\right) = \mathrm{P}(Z = k) = \frac{\mathrm{e}^{-n\mu}(n\mu)^k}{k!}.$$

The probability mass function of T for the case $n = 30$ and $p_0 = 0.1$ is displayed in Figure 19.3. From the histogram in Figure 19.1 as well as from the probability mass function in Figure 19.3, you may get the impression that T is also an unbiased estimator. However, this not the case, which follows immediately from an application of Jensen's inequality:

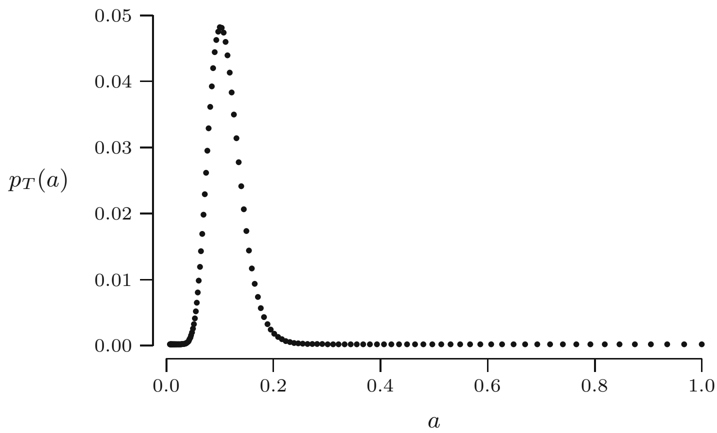

Fig. 19.3. Probability mass function of T.

$$\mathrm{E}[T] = \mathrm{E}\left[\mathrm{e}^{-\bar{X}_n}\right] > \mathrm{e}^{-\mathrm{E}\left[\bar{X}_n\right]},$$

where we have a strict inequality because the function $g(x) = \mathrm{e}^{-x}$ is strictly convex ($g''(x) = \mathrm{e}^{-x} > 0$). Recall that the parameter μ equals the expectation of the *Pois*(μ) model distribution, so that according to Section 13.1 we have $\mathrm{E}\left[\bar{X}_n\right] = \mu$. We find that

$$\mathrm{E}[T] > \mathrm{e}^{-\mu} = p_0,$$

which means that the estimator T for p_0 has positive bias. In fact we can compute $\mathrm{E}[T]$ exactly (see Exercise 19.9):

$$\mathrm{E}[T] = \mathrm{E}\left[\mathrm{e}^{-\bar{X}_n}\right] = \mathrm{e}^{-n\mu(1-\mathrm{e}^{-1/n})}.$$

Note that $n(1 - \mathrm{e}^{-1/n}) \to 1$, so that

$$\mathrm{E}[T] = \mathrm{e}^{-n\mu(1-\mathrm{e}^{-1/n})} \to \mathrm{e}^{-\mu} = p_0$$

as n goes to infinity. Hence, although T has positive bias, the bias decreases to zero as the sample size becomes larger. In Figure 19.4 the expectation of T is displayed as a function of the sample size n for the case $\mu = \ln(10)$. For $n = 30$ the difference between $\mathrm{E}[T]$ and $p_0 = 0.1$ equals 0.0038.

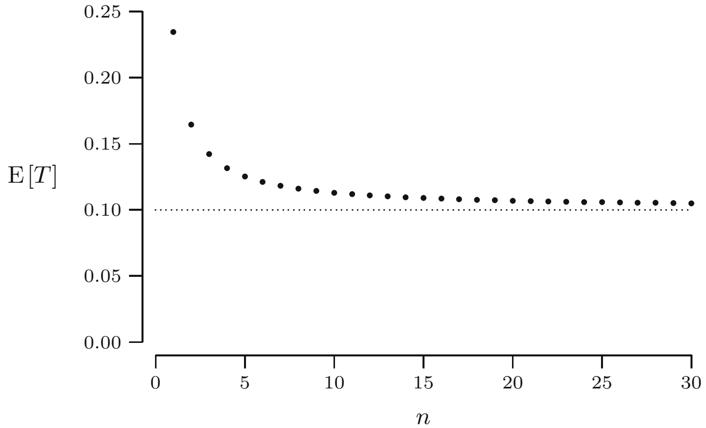

Fig. 19.4. $\mathrm{E}[T]$ as a function of n.

Quick exercise 19.2 If we estimate $p_0 = \mathrm{e}^{-\mu}$ by the relative frequency of zeros $S = Y/n$, then we could estimate μ by $U = -\ln(S)$. Argue that U is a biased estimator for μ. Is the bias positive or negative?

We conclude this section by returning to the estimation of the parameter μ. Apart from the (biased) estimator in Quick exercise 19.2 we also considered the sample mean $\bar{X}_n$ and sample variance S_n^2 as possible estimators for μ. These are both unbiased estimators for the parameter μ. This is a direct consequence of a more general property of $\bar{X}_n$ and S_n^2, which is discussed in the next section.

19.4 Unbiased estimators for expectation and variance

Sometimes the quantity of interest can be described by the expectation or variance of the model distribution, and is it irrelevant whether this distribution is of a parametric type. In this section we propose unbiased estimators for these distribution features.

UNBIASED ESTIMATORS FOR EXPECTATION AND VARIANCE. Suppose $X_1, X_2, \ldots, X_n$ is a random sample from a distribution with finite expectation μ and finite variance σ^2. Then

$$\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

is an *unbiased estimator for* μ and

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

is an *unbiased estimator for* σ^2.

The first statement says that $\mathrm{E}\left[\bar{X}_n\right] = \mu$, which was shown in Section 13.1. The second statement says $\mathrm{E}\left[S_n^2\right] = \sigma^2$. To see this, use linearity of expectations to write

$$\mathrm{E}\left[S_n^2\right] = \frac{1}{n-1} \sum_{i=1}^{n} \mathrm{E}\left[(X_i - \bar{X}_n)^2\right].$$

Since $\mathrm{E}\left[\bar{X}_n\right] = \mu$, we have $\mathrm{E}\left[X_i - \bar{X}_n\right] = \mathrm{E}\left[X_i\right] - \mathrm{E}\left[\bar{X}_n\right] = 0$. Now note that for any random variable Y with $\mathrm{E}\left[Y\right] = 0$, we have

$$\mathrm{Var}(Y) = \mathrm{E}\left[Y^2\right] - (\mathrm{E}\left[Y\right])^2 = \mathrm{E}\left[Y^2\right].$$

Applying this to $Y = X_i - \bar{X}_n$, it follows that

$$\mathrm{E}\left[(X_i - \bar{X}_n)^2\right] = \mathrm{Var}\big(X_i - \bar{X}_n\big).$$

Note that we can write

$$X_i - \bar{X}_n = \frac{n-1}{n} X_i - \frac{1}{n} \sum_{j \neq i} X_j.$$

Then from the rules concerning variances of sums of independent random variables we find that

$$\begin{aligned}\mathrm{Var}\big(X_i - \bar{X}_n\big) &= \mathrm{Var}\left(\frac{n-1}{n}X_i - \frac{1}{n}\sum_{j\neq i} X_j\right)\\ &= \frac{(n-1)^2}{n^2}\mathrm{Var}(X_i) + \frac{1}{n^2}\sum_{j\neq i}\mathrm{Var}(X_j)\\ &= \left[\frac{(n-1)^2}{n^2} + \frac{n-1}{n^2}\right]\sigma^2 = \frac{n-1}{n}\sigma^2.\end{aligned}$$

We conclude that

$$\begin{aligned}\mathrm{E}\left[S_n^2\right] &= \frac{1}{n-1}\sum_{i=1}^{n}\mathrm{E}\left[(X_i - \bar{X}_n)^2\right]\\ &= \frac{1}{n-1}\sum_{i=1}^{n}\mathrm{Var}\big(X_i - \bar{X}_n\big) = \frac{1}{n-1}\cdot n \cdot \frac{n-1}{n}\sigma^2 = \sigma^2.\end{aligned}$$

This explains why we divide by $n-1$ in the formula for S_n^2; only in this case S_n^2 is an unbiased estimator for the "true" variance σ^2. If we would divide by n instead of $n-1$, we would obtain an estimator with negative bias; it would systematically produce too-small estimates for σ^2.

Quick exercise 19.3 Consider the following estimator for σ^2:

$$V_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2.$$

Compute the bias $\mathrm{E}\left[V_n^2\right] - \sigma^2$ for this estimator, where you can keep computations simple by realizing that $V_n^2 = (n-1)S_n^2/n$.

Unbiasedness does not always carry over

We have seen that S_n^2 is an unbiased estimator for the "true" variance σ^2. A natural question is whether S_n is again an unbiased estimator for σ. This is not the case. Since the function $g(x) = x^2$ is strictly convex, Jensen's inequality yields that

$$\sigma^2 = \mathrm{E}\left[S_n^2\right] > \left(\mathrm{E}\left[S_n\right]\right)^2,$$

which implies that $\mathrm{E}\left[S_n\right] < \sigma$. Another example is the network arrivals, in which $\bar{X}_n$ is an unbiased estimator for μ, whereas $\mathrm{e}^{-\bar{X}_n}$ is positively biased with respect to $\mathrm{e}^{-\mu}$. These examples illustrate a general fact: unbiasedness *does not* always carry over, i.e., if T is an unbiased estimator for a parameter θ, then $g(T)$ *does not* have to be an unbiased estimator for $g(\theta)$.

However, there is one special case in which unbiasedness does carry over, namely if $g(T) = aT + b$. Indeed, if T is unbiased for θ: $\mathrm{E}[T] = \theta$, then by the change-of-units rule for expectations,

$$\mathrm{E}[aT + b] = a\mathrm{E}[T] + b = a\theta + b,$$

which means that $aT + b$ is unbiased for $a\theta + b$.

19.5 Solutions to the quick exercises

19.1 Write y for the number of x_i equal to zero. Denote the probability of zero by p_0, so that $p_0 = \mathrm{e}^{-\mu}$. This means that $\mu = -\ln(p_0)$. Hence if we estimate p_0 by the relative frequency y/n, we can estimate μ by $-\ln(y/n)$.

19.2 The function $g(x) = -\ln(x)$ is strictly convex, since $g''(x) = 1/x^2 > 0$. Hence by Jensen's inequality

$$\mathrm{E}[U] = \mathrm{E}[-\ln(S)] > -\ln(\mathrm{E}[S]).$$

Since we have seen that $\mathrm{E}[S] = p_0 = \mathrm{e}^{-\mu}$, it follows that $\mathrm{E}[U] > -\ln(\mathrm{E}[S]) = -\ln(\mathrm{e}^{-\mu}) = \mu$. This means that U has positive bias.

19.3 Using that $\mathrm{E}\left[S_n^2\right] = \sigma^2$, we find that

$$\mathrm{E}\left[V_n^2\right] = \mathrm{E}\left[\frac{n-1}{n}S_n^2\right] = \frac{n-1}{n}\mathrm{E}\left[S_n^2\right] = \frac{n-1}{n}\sigma^2.$$

We conclude that the bias of V_n^2 equals $\mathrm{E}\left[V_n^2\right] - \sigma^2 = -\sigma^2/n < 0$.

19.6 Exercises

19.1 ⊞ Suppose our dataset is a realization of a random sample $X_1, X_2, \ldots, X_n$ from a uniform distribution on the interval $[-\theta, \theta]$, where θ is unknown.

a. Show that

$$T = \frac{3}{n}(X_1^2 + X_2^2 + \cdots + X_n^2)$$

is an unbiased estimator for θ^2.

b. Is $\sqrt{T}$ also an unbiased estimator for θ? If not, argue whether it has positive or negative bias.

19.2 Suppose the random variables $X_1, X_2, \ldots, X_n$ have the same expectation μ.

a. Is $S = \frac{1}{2}X_1 + \frac{1}{3}X_2 + \frac{1}{6}X_3$ an unbiased estimator for μ?

b. Under what conditions on constants $a_1, a_2, \ldots, a_n$ is

$$T = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n$$

an unbiased estimator for μ?

19.3 ⊡ Suppose the random variables $X_1, X_2, \ldots, X_n$ have the same expectation μ. For which constants a and b is

$$T = a(X_1 + X_2 + \cdots + X_n) + b$$

an unbiased estimator for μ?

19.4 Recall Exercise 17.5 about the number of cycles to pregnancy. Suppose the dataset corresponding to the table in Exercise 17.5 **a** is modeled as a realization of a random sample $X_1, X_2, \ldots, X_n$ from a *Geo*(p) distribution, where $0 < p < 1$ is unknown. Motivated by the law of large numbers, a natural estimator for p is

$$T = 1/\bar{X}_n.$$

a. Check that T is a biased estimator for p and find out whether it has positive or negative bias.

b. In Exercise 17.5 we discussed the estimation of the probability that a woman becomes pregnant within three or fewer cycles. One possible estimator for this probability is the relative frequency of women that became pregnant within three cycles

$$S = \frac{\text{number of } X_i \leq 3}{n}.$$

Show that S is an unbiased estimator for this probability.

19.5 ⊡ Suppose a dataset is modeled as a realization of a random sample $X_1, X_2, \ldots, X_n$ from an *Exp*(λ) distribution, where $\lambda > 0$ is unknown. Let μ denote the corresponding expectation and let M_n denote the minimum of $X_1, X_2, \ldots, X_n$. Recall from Exercise 8.18 that M_n has an *Exp*$(n\lambda)$ distribution. Find out for which constant c the estimator

$$T = cM_n$$

is an unbiased estimator for μ.

19.6 ⊡ Consider the following dataset of lifetimes of ball bearings in hours.

6278	3113	5236	11584	12628	7725	8604	14266	6125	9350
3212	9003	3523	12888	9460	13431	17809	2812	11825	2398

Source: J.E. Angus. Goodness-of-fit tests for exponentiality based on a loss-of-memory type functional equation. *Journal of Statistical Planning and Inference*, 6:241-251, 1982; example 5 on page 249.

One is interested in estimating the minimum lifetime of this type of ball bearing. The dataset is modeled as a realization of a random sample $X_1, \ldots, X_n$. Each random variable X_i is represented as

$$X_i = \delta + Y_i,$$

where Y_i has an $Exp(\lambda)$ distribution and $\delta > 0$ is an unknown parameter that is supposed to model the minimum lifetime. The objective is to construct an unbiased estimator for δ. It is known that

$$\mathrm{E}[M_n] = \delta + \frac{1}{n\lambda} \quad \text{and} \quad \mathrm{E}\left[\bar{X}_n\right] = \delta + \frac{1}{\lambda},$$

where M_n = minimum of $X_1, X_2, \ldots, X_n$ and $\bar{X}_n = (X_1 + X_2 + \cdots + X_n)/n$.

a. Check that

$$T = \frac{n}{n-1}\left(\bar{X}_n - M_n\right)$$

is an unbiased estimator for $1/\lambda$.

b. Construct an unbiased estimator for δ.

c. Use the dataset to compute an estimate for the minimum lifetime δ. You may use that the average lifetime of the data is 8563.5.

19.7 Leaves are divided into four different types: starchy-green, sugary-white, starchy-white, and sugary-green. According to genetic theory, the types occur with probabilities $\frac{1}{4}(\theta + 2)$, $\frac{1}{4}\theta$, $\frac{1}{4}(1-\theta)$, and $\frac{1}{4}(1-\theta)$, respectively, where $0 < \theta < 1$. Suppose one has n leaves. Then the number of starchy-green leaves is modeled by a random variable N_1 with a $Bin(n, p_1)$ distribution, where $p_1 = \frac{1}{4}(\theta + 2)$, and the number of sugary-white leaves is modeled by a random variable N_2 with a $Bin(n, p_2)$ distribution, where $p_2 = \frac{1}{4}\theta$. The following table lists the counts for the progeny of self-fertilized heterozygotes among 3839 leaves.

Type	Count
Starchy-green	1997
Sugary-white	32
Starchy-white	906
Sugary-green	904

Source: R.A. Fisher. *Statistical methods for research workers.* Hafner, New York, 1958; Table 62 on page 299.

Consider the following two estimators for θ:

$$T_1 = \frac{4}{n}N_1 - 2 \quad \text{and} \quad T_2 = \frac{4}{n}N_2.$$

a. Check that both T_1 and T_2 are unbiased estimators for θ.

b. Compute the value of both estimators for θ.

19.8 ⊞ Recall the black cherry trees example from Exercise 17.9, modeled by a linear regression model without intercept

$$Y_i = \beta x_i + U_i \quad \text{for } i = 1, 2, \ldots, n,$$

where $U_1, U_2, \ldots, U_n$ are independent random variables with $\mathrm{E}[U_i] = 0$ and $\mathrm{Var}(U_i) = \sigma^2$. We discussed three estimators for the parameter β:

$$\begin{aligned} B_1 &= \frac{1}{n}\left(\frac{Y_1}{x_1} + \cdots + \frac{Y_n}{x_n}\right), \\ B_2 &= \frac{Y_1 + \cdots + Y_n}{x_1 + \cdots + x_n}, \\ B_3 &= \frac{x_1 Y_1 + \cdots + x_n Y_n}{x_1^2 + \cdots + x_n^2}. \end{aligned}$$

Show that all three estimators are unbiased for β.

19.9 Consider the network example where the dataset is modeled as a realization of a random sample $X_1, X_2, \ldots, X_n$ from a $Pois(\mu)$ distribution. We estimate the probability of zero arrivals $\mathrm{e}^{-\mu}$ by means of $T = \mathrm{e}^{-\bar{X}_n}$. Check that

$$\mathrm{E}[T] = \mathrm{e}^{-n\mu(1-\mathrm{e}^{-1/n})}.$$

Hint: write $T = \mathrm{e}^{-Z/n}$, where $Z = X_1 + X_2 + \cdots + X_n$ has a $Pois(n\mu)$ distribution.