

## Efficiency and mean squared error

In the previous chapter we introduced the notion of unbiasedness as a desirable property of an estimator. If several unbiased estimators for the same parameter of interest exist, we need a criterion for comparison of these estimators. A natural criterion is some measure of spread of the estimators around the parameter of interest. For unbiased estimators we will use variance. For arbitrary estimators we introduce the notion of *mean squared error* (MSE), which combines variance and bias.

### 20.1 Estimating the number of German tanks

In this section we come back to the problem of estimating German war production as discussed in Section 1.5. We consider serial numbers on tanks, recoded to numbers running from 1 to some unknown largest number  $N$ . Given is a subset of  $n$  numbers of this set. The objective is to estimate the total number of tanks  $N$  on the basis of the observed serial numbers.

Denote the observed distinct serial numbers by  $x_1, x_2, \dots, x_n$ . This dataset can be modeled as a realization of random variables  $X_1, X_2, \dots, X_n$  representing  $n$  draws *without replacement* from the numbers  $1, 2, \dots, N$  with equal probability. Note that in this example our dataset is *not* a realization of a random sample, because the random variables  $X_1, X_2, \dots, X_n$  are *dependent*. We propose two unbiased estimators. The first one is based on the sample mean

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

and the second one is based on the sample maximum

$$M_n = \max\{X_1, X_2, \dots, X_n\}.$$

### An estimator based on the sample mean

To construct an unbiased estimator for  $N$  based on the sample mean, we start by computing the expectation of  $\bar{X}_n$ . The linearity-of-expectations rule also applies to dependent random variables, so that

$$E[\bar{X}_n] = \frac{E[X_1] + E[X_2] + \cdots + E[X_n]}{n}.$$

In Section 9.3 we saw that the marginal distribution of each  $X_i$  is the same:

$$P(X_i = k) = \frac{1}{N} \quad \text{for } k = 1, 2, \dots, N.$$

Therefore the expectation of each  $X_i$  is given by

$$\begin{aligned} E[X_i] &= 1 \cdot \frac{1}{N} + 2 \cdot \frac{1}{N} + \cdots + N \cdot \frac{1}{N} = \frac{1 + 2 + \cdots + N}{N} \\ &= \frac{\frac{1}{2}N(N+1)}{N} = \frac{N+1}{2}. \end{aligned}$$

It follows that

$$E[\bar{X}_n] = \frac{E[X_1] + E[X_2] + \cdots + E[X_n]}{n} = \frac{N+1}{2}.$$

This directly implies that

$$T_1 = 2\bar{X}_n - 1$$

is an unbiased estimator for  $N$ , since the change-of-units rule yields that

$$E[T_1] = E[2\bar{X}_n - 1] = 2E[\bar{X}_n] - 1 = 2 \cdot \frac{N+1}{2} - 1 = N.$$

**QUICK EXERCISE 20.1** Suppose we have observed tanks with (recoded) serial numbers

$$61 \quad 19 \quad 56 \quad 24 \quad 16.$$

Compute the value of the estimator  $T_1$  for the total number of tanks.

### An estimator based on the sample maximum

To construct an unbiased estimator for  $N$  based on the maximum, we first compute the expectation of  $M_n$ . We start by computing the probability that  $M_n = k$ , where  $k$  takes the values  $n, \dots, N$ . Similar to the combinatorics used in Section 4.3 to derive the binomial distribution, the number of ways to draw  $n$  numbers without replacement from  $1, 2, \dots, N$  is  $\binom{N}{n}$ . Hence each combination has probability  $1/\binom{N}{n}$ . In order to have  $M_n = k$ , we must have one number equal to  $k$  and choose the other  $n-1$  numbers out of the numbers  $1, 2, \dots, k-1$ . There are  $\binom{k-1}{n-1}$  ways to do this. Hence for the possible values  $k = n, n+1, \dots, N$ ,

$$\begin{aligned} P(M_n = k) &= \frac{\binom{k-1}{n-1}}{\binom{N}{n}} = \frac{(k-1)!}{(k-n)!(n-1)!} \cdot \frac{(N-n)!n!}{N!} \\ &= n \cdot \frac{(k-1)!(N-n)!}{(k-n)!N!}. \end{aligned}$$

Thus the expectation of  $M_n$  is given by

$$\begin{aligned} E[M_n] &= \sum_{k=n}^N kP(M_n = k) = \sum_{k=n}^N k \cdot n \cdot \frac{(k-1)!(N-n)!}{(k-n)!N!} \\ &= \sum_{k=n}^N n \cdot \frac{k!}{(k-n)!} \frac{(N-n)!}{N!} \\ &= n \cdot \frac{(N-n)!}{N!} \sum_{k=n}^N \frac{k!}{(k-n)!}. \end{aligned}$$

How to continue the computation of  $E[M_n]$ ? We use a trick: we start by rearranging

$$1 = \sum_{j=n}^N P(M_n = j) = \sum_{j=n}^N n \cdot \frac{(j-1)!(N-n)!}{(j-n)!N!},$$

finding that

$$\sum_{j=n}^N \frac{(j-1)!}{(j-n)!} = \frac{N!}{n(N-n)!}. \quad (20.1)$$

This holds for any  $N$  and any  $n \leq N$ . In particular we could replace  $N$  by  $N+1$  and  $n$  by  $n+1$ :

$$\sum_{j=n+1}^{N+1} \frac{(j-1)!}{(j-n-1)!} = \frac{(N+1)!}{(n+1)(N-n)!}.$$

Changing the summation variable to  $k = j - 1$ , we obtain

$$\sum_{k=n}^N \frac{k!}{(k-n)!} = \frac{(N+1)!}{(n+1)(N-n)!}. \quad (20.2)$$

This is exactly what we need to finish the computation of  $E[M_n]$ . Substituting (20.2) in what we obtained earlier, we find

$$\begin{aligned} E[M_n] &= n \cdot \frac{(N-n)!}{N!} \sum_{k=n}^N \frac{k!}{(k-n)!} \\ &= n \cdot \frac{(N-n)!}{N!} \cdot \frac{(N+1)!}{(n+1)(N-n)!} = n \cdot \frac{N+1}{n+1}. \end{aligned}$$

QUICK EXERCISE 20.2 Choosing  $n = N$  in this formula yields  $E[M_N] = N$ . Can you argue that this is the right answer without doing any computations?

With the formula for  $E[M_n]$  we can derive immediately that

$$T_2 = \frac{n + 1}{n} M_n - 1$$

is an unbiased estimator for  $N$ , since by the change-of-units rule,

$$E[T_2] = E\left[\frac{n + 1}{n} M_n - 1\right] = \frac{n + 1}{n} E[M_n] - 1 = \frac{n + 1}{n} \cdot \frac{n(N + 1)}{n + 1} - 1 = N.$$

QUICK EXERCISE 20.3 Compute the value of estimator  $T_2$  for the total number of tanks on basis of the observed numbers from Quick exercise 20.1.

## 20.2 Variance of an estimator

In the previous section we saw that we can construct two completely different estimators for the total number of tanks  $N$  that are *both* unbiased. The obvious question is: which of the two is better? To answer this question, we investigate how both estimators vary around the parameter of interest  $N$ . Although we could in principle compute the distributions of  $T_1$  and  $T_2$ , we carry out a small simulation study instead. Take  $N = 1000$  and  $n = 10$  fixed. We draw 10 numbers, without replacement, from  $1, 2, \dots, 1000$  and compute the value of the estimators  $T_1$  and  $T_2$ . We repeat this two thousand times, so that we have 2000 values for both estimators. In Figure 20.1 we have displayed the histogram of the 2000 values for  $T_1$  on the left and the histogram of the 2000 values for  $T_2$  on the right. From the histograms, which reflect the probability

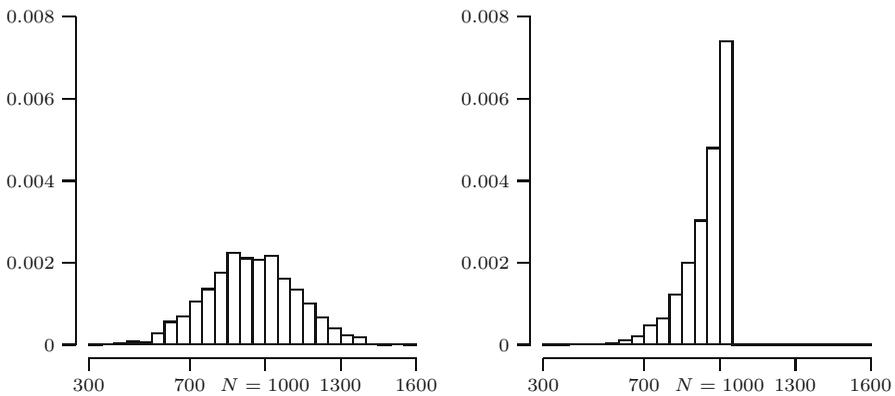


Fig. 20.1. Histograms of two thousand values for  $T_1$  (left) and  $T_2$  (right).

mass functions of both estimators, we see that the distributions of  $T_1$  and  $T_2$  are of completely different types. As can be expected from the fact that both estimators are unbiased, the values vary around the parameter of interest  $N = 1000$ . The most important difference between the histograms is that the variation in the values of  $T_2$  is *less* than the variation in the values of  $T_1$ . This suggests that estimator  $T_2$  estimates the total number of tanks more efficiently than estimator  $T_1$ , in the sense that it produces estimates that are more concentrated around the parameter of interest  $N$  than estimates produced by  $T_1$ . Recall that the variance measures the spread of a random variable. Hence the previous discussion motivates the use of the variance of an estimator to evaluate its performance.

**EFFICIENCY.** Let  $T_1$  and  $T_2$  be two unbiased estimators for the same parameter  $\theta$ . Then estimator  $T_2$  is called *more efficient* than estimator  $T_1$  if  $\text{Var}(T_2) < \text{Var}(T_1)$ , irrespective of the value of  $\theta$ .

Let us compare  $T_1$  and  $T_2$  using this criterion. For  $T_1$  we have

$$\text{Var}(T_1) = \text{Var}(2\bar{X}_n - 1) = 4\text{Var}(\bar{X}_n).$$

Although the  $X_i$  are not independent, it is true that all pairs  $(X_i, X_j)$  with  $i \neq j$  have the *same* distribution (this follows in the same way in which we showed on page 122 that all  $X_i$  have the same distribution). With the variance-of-the-sum rule for  $n$  random variables (see Exercise 10.17), we find that

$$\text{Var}(X_1 + \cdots + X_n) = n\text{Var}(X_1) + n(n-1)\text{Cov}(X_1, X_2).$$

In Exercises 9.18 and 10.18, we computed that

$$\text{Var}(X_1) = \frac{1}{12}(N-1)(N+1), \quad \text{Cov}(X_1, X_2) = -\frac{1}{12}(N+1).$$

We find therefore that

$$\begin{aligned} \text{Var}(T_1) &= 4\text{Var}(\bar{X}_n) = \frac{4}{n^2}\text{Var}(X_1 + \cdots + X_n) \\ &= \frac{4}{n^2} \left[ n \cdot \frac{1}{12}(N-1)(N+1) - n(n-1) \cdot \frac{1}{12}(N+1) \right] \\ &= \frac{1}{3n}(N+1)[N-1 - (n-1)] \\ &= \frac{(N+1)(N-n)}{3n}. \end{aligned}$$

Obtaining the variance of  $T_2$  is a little more work. One can compute the variance of  $M_n$  in a way that is very similar to the way we obtained  $E[M_n]$ . The result is (see Remark 20.1 for details)

$$\text{Var}(M_n) = \frac{n(N+1)(N-n)}{(n+2)(n+1)^2}.$$

**Remark 20.1 (How to compute this variance).** The trick is to compute not  $E[M_n^2]$  but  $E[M_n(M_n + 1)]$ . First we derive an identity from Equation (20.1) as before, this time replacing  $N$  by  $N + 2$  and  $n$  by  $n + 2$ :

$$\sum_{j=n+2}^{N+2} \frac{(j-1)!}{(j-n-2)!} = \frac{(N+2)!}{(n+2)(N-n)!}.$$

Changing the summation variable to  $k = j - 2$  yields

$$\sum_{k=n}^N \frac{(k+1)!}{(k-n)!} = \frac{(N+2)!}{(n+2)(N-n)!}.$$

With this formula one can obtain:

$$E[M_n(M_n + 1)] = \sum_{k=n}^N k(k+1) \cdot n \frac{(k-1)! (N-n)!}{(k-n)! N!} = \frac{n(N+1)(N+2)}{n+2}.$$

Since we know  $E[M_n]$ , we can determine  $E[M_n^2]$  from this, and subsequently the variance of  $M_n$ .

With the expression for the variance of  $M_n$ , we derive

$$\text{Var}(T_2) = \text{Var}\left(\frac{n+1}{n}M_n - 1\right) = \frac{(n+1)^2}{n^2} \text{Var}(M_n) = \frac{(N+1)(N-n)}{n(n+2)}.$$

We see that  $\text{Var}(T_2) < \text{Var}(T_1)$  for all  $N$  and  $n \geq 2$ . Hence  $T_2$  is always more efficient than  $T_1$ , except when  $n = 1$ . In this case the variances are equal, simply because the estimators are the same—they both equal  $X_1$ .

The quotient  $\text{Var}(T_1)/\text{Var}(T_2)$ , is called the *relative efficiency* of  $T_2$  with respect to  $T_1$ . In our case the relative efficiency of  $T_2$  with respect to  $T_1$  equals

$$\frac{\text{Var}(T_1)}{\text{Var}(T_2)} = \frac{(N+1)(N-n)}{3n} \cdot \frac{n(n+2)}{(N+1)(N-n)} = \frac{n+2}{3}.$$

Surprisingly, this quotient does not depend on  $N$ , and we see clearly the advantage of  $T_2$  over  $T_1$  as the sample size  $n$  gets larger.

**QUICK EXERCISE 20.4** Let  $n = 5$ , and let the sample be

$$7 \quad 3 \quad 10 \quad 45 \quad 15.$$

Compute the value of the estimator  $T_1$  for  $N$ . Do you notice anything strange?

The self-contradictory behavior of  $T_1$  in Quick exercise 20.4 is not rare: this phenomenon will occur for up to 50% of the samples if  $n$  and  $N$  are large. This gives another reason to prefer  $T_2$  over  $T_1$ .

**Remark 20.2 (The Cramér-Rao inequality).** Suppose we have a random sample from a continuous distribution with probability density function  $f_\theta$ , where  $\theta$  is the parameter of interest. Under certain smoothness conditions on the density  $f_\theta$ , the variance of an unbiased estimator  $T$  for  $\theta$  always has to be larger than or equal to a certain positive number, the so-called Cramér-Rao lower bound:

$$\text{Var}(T) \geq \frac{1}{n\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\ln f_\theta(X)\right)^2\right]} \quad \text{for all } \theta.$$

Here  $n$  is the size of the sample and  $X$  a random variable whose density function is  $f_\theta$ . In some cases we can find unbiased estimators attaining this bound. These are called *minimum variance unbiased estimators*. An example is the sample mean for the expectation of an exponential distribution. (We will consider this case in Exercise 20.3.)

## 20.3 Mean squared error

In the last section we compared two unbiased estimators by considering their spread around the value to be estimated, where the spread was measured by the variance. Although unbiasedness is a desirable property, the performance of an estimator should mainly be judged by the way it spreads around the parameter  $\theta$  to be estimated. This leads to the following definition.

**DEFINITION.** Let  $T$  be an estimator for a parameter  $\theta$ . The *mean squared error* of  $T$  is the number  $\text{MSE}(T) = \mathbb{E}[(T - \theta)^2]$ .

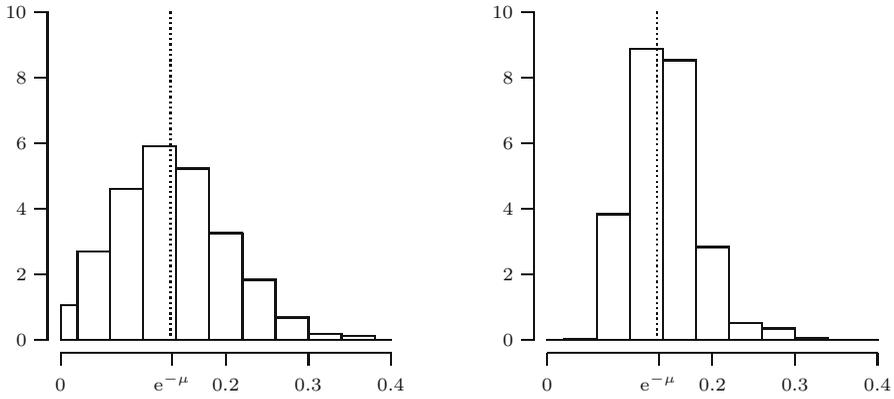
According to this criterion, an estimator  $T_1$  performs better than an estimator  $T_2$  if  $\text{MSE}(T_1) < \text{MSE}(T_2)$ . Note that

$$\begin{aligned} \text{MSE}(T) &= \mathbb{E}[(T - \theta)^2] \\ &= \mathbb{E}[(T - \mathbb{E}[T] + \mathbb{E}[T] - \theta)^2] \\ &= \mathbb{E}[(T - \mathbb{E}[T])^2] + 2\mathbb{E}[T - \mathbb{E}[T]](\mathbb{E}[T] - \theta) + (\mathbb{E}[T] - \theta)^2 \\ &= \text{Var}(T) + (\mathbb{E}[T] - \theta)^2. \end{aligned}$$

So the MSE of  $T$  turns out to be the variance of  $T$  plus the square of the bias of  $T$ . In particular, when  $T$  is unbiased, the MSE of  $T$  is just the variance of  $T$ . This means that we already used mean squared errors to compare the estimators  $T_1$  and  $T_2$  in the previous section. We extend the notion of efficiency by saying that estimator  $T_2$  is more *efficient* than estimator  $T_1$  (for the same parameter of interest), if the MSE of  $T_2$  is smaller than the MSE of  $T_1$ .

### Unbiasedness and efficiency

A biased estimator with a small variance may be more useful than an unbiased estimator with a large variance. We illustrate this with the network server

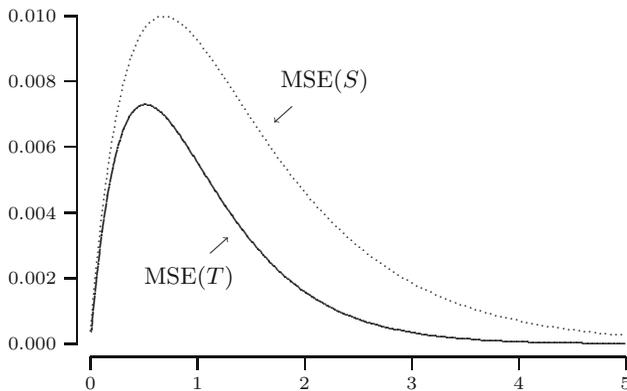


**Fig. 20.2.** Histograms of a thousand values for  $S$  (left) and  $T$  (right).

example from Section 19.2. Recall that our goal was to estimate the probability  $p_0 = e^{-\mu}$  of zero arrivals (of packages) in a minute. We did have two promising candidates as estimators:

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

In Figure 20.2 we depict histograms of one thousand simulations of the values of  $S$  and  $T$  computed for random samples of size  $n = 25$  from a  $Pois(\mu)$  distribution, where  $\mu = 2$ . Considering the way the values of the (biased!) estimator  $T$  are more concentrated around the true value  $e^{-\mu} = e^{-2} = 0.1353$ , we would be inclined to prefer  $T$  over  $S$ . This choice is strongly supported by the fact that  $T$  is more efficient than  $S$ :  $MSE(T)$  is always smaller than  $MSE(S)$ , as illustrated in Figure 20.3.



**Fig. 20.3.** MSEs of  $S$  and  $T$  as a function of  $\mu$ .

## 20.4 Solutions to the quick exercises

**20.1** We have  $\bar{x}_5 = (61 + 19 + 56 + 24 + 16)/5 = 176/5 = 35.2$ . Therefore  $t_1 = 2 \cdot 35.2 - 1 = 69.4$ .

**20.2** When  $n = N$ , we have drawn *all* the numbers. But then the largest number  $M_N$  is  $N$ , and so  $E[M_N] = N$ .

**20.3** We have  $t_2 = (6/5) \cdot 61 - 1 = 72.2$ .

**20.4** Since 45 is in the sample,  $N$  has to be at least 45. Adding the numbers yields  $7 + 3 + 10 + 15 + 45 = 80$ . So  $t_1 = 2\bar{x}_n - 1 = 2 \cdot 16 - 1 = 31$ . What is strange about this is that the estimate for  $N$  is far smaller than the number 45 in the sample!

## 20.5 Exercises

**20.1** Given is a random sample  $X_1, X_2, \dots, X_n$  from a distribution with finite variance  $\sigma^2$ . We estimate the expectation of the distribution with the sample mean  $\bar{X}_n$ . Argue that the larger our sample, the more efficient our estimator. What is the relative efficiency  $\text{Var}(\bar{X}_n) / \text{Var}(\bar{X}_{2n})$  of  $\bar{X}_{2n}$  with respect to  $\bar{X}_n$ ?

**20.2**  $\boxplus$  Given are two estimators  $S$  and  $T$  for a parameter  $\theta$ . Furthermore it is known that  $\text{Var}(S) = 40$  and  $\text{Var}(T) = 4$ .

- Suppose that we know that  $E[S] = \theta$  and  $E[T] = \theta + 3$ . Which estimator would you prefer, and why?
- Suppose that we know that  $E[S] = \theta$  and  $E[T] = \theta + a$  for some positive number  $a$ . For each  $a$ , which estimator would you prefer, and why?

**20.3**  $\boxplus$  Suppose we have a random sample  $X_1, \dots, X_n$  from an  $\text{Exp}(\lambda)$  distribution. Suppose we want to estimate the mean  $1/\lambda$ . According to Section 19.4 the estimator

$$T_1 = \bar{X}_n = \frac{1}{n} (X_1 + X_2 + \dots + X_n)$$

is an unbiased estimator of  $1/\lambda$ . Let  $M_n$  be the minimum of  $X_1, X_2, \dots, X_n$ . Recall from Exercise 8.18 that  $M_n$  has an  $\text{Exp}(n\lambda)$  distribution. In Exercise 19.5 you have determined that

$$T_2 = nM_n$$

is another unbiased estimator for  $1/\lambda$ . Which of the estimators  $T_1$  and  $T_2$  would you choose for estimating the mean  $1/\lambda$ ? Substantiate your answer.

**20.4**  $\square$  Consider the situation of this chapter, where we have to estimate the parameter  $N$  from a sample  $x_1, \dots, x_n$  drawn without replacement from the numbers  $\{1, \dots, N\}$ . To keep it simple, we consider  $n = 2$ . Let  $M = M_2$  be the maximum of  $X_1$  and  $X_2$ . We have found that  $T_2 = 3M/2 - 1$  is a good unbiased estimator for  $N$ . We want to construct a new unbiased estimator  $T_3$  based on the *minimum*  $L$  of  $X_1$  and  $X_2$ . In the following you may use that the random variable  $L$  has the same distribution as the random variable  $N + 1 - M$  (this follows from symmetry considerations).

- Show that  $T_3 = 3L - 1$  is an unbiased estimator for  $N$ .
- Compute  $\text{Var}(T_3)$  using that  $\text{Var}(M) = (N + 1)(N - 2)/18$ . (The latter has been computed in Remark 20.1.)
- What is the relative efficiency of  $T_2$  with respect to  $T_3$ ?

**20.5** Someone is proposing two unbiased estimators  $U$  and  $V$ , with the *same* variance  $\text{Var}(U) = \text{Var}(V)$ . It therefore appears that we would not prefer one estimator over the other. However, we could go for a third estimator, namely  $W = (U + V)/2$ . Note that  $W$  is unbiased. To judge the quality of  $W$  we want to compute its variance. Lacking information on the joint probability distribution of  $U$  and  $V$ , this is impossible. However, we should prefer  $W$  in any case! To see this, show by means of the variance-of-the-sum rule that the relative efficiency of  $U$  with respect to  $W$  is equal to

$$\frac{\text{Var}((U + V)/2)}{\text{Var}(U)} = \frac{1}{2} + \frac{1}{2}\rho(U, V).$$

Here  $\rho(U, V)$  is the correlation coefficient. Why does this result imply that we should use  $W$  instead of  $U$  (or  $V$ )?

**20.6** A geodesic engineer measures the three unknown angles  $\alpha_1, \alpha_2$ , and  $\alpha_3$  of a triangle. He models the uncertainty in the measurements by considering them as realizations of three independent random variables  $T_1, T_2$ , and  $T_3$  with expectations

$$E[T_1] = \alpha_1, \quad E[T_2] = \alpha_2, \quad E[T_3] = \alpha_3,$$

and all three with the same variance  $\sigma^2$ . In order to make use of the fact that the three angles must add to  $\pi$ , he also considers new estimators  $U_1, U_2$ , and  $U_3$  defined by

$$\begin{aligned} U_1 &= T_1 + \frac{1}{3}(\pi - T_1 - T_2 - T_3), \\ U_2 &= T_2 + \frac{1}{3}(\pi - T_1 - T_2 - T_3), \\ U_3 &= T_3 + \frac{1}{3}(\pi - T_1 - T_2 - T_3). \end{aligned}$$

(Note that the “deviation”  $\pi - T_1 - T_2 - T_3$  is evenly divided over the three measurements and that  $U_1 + U_2 + U_3 = \pi$ .)

- a. Compute  $E[U_1]$  and  $\text{Var}(U_1)$ .
- b. What does he gain in efficiency when he uses  $U_1$  instead of  $T_1$  to estimate the angle  $\alpha_1$ ?
- c. What kind of estimator would you choose for  $\alpha_1$  if it is known that the triangle is isosceles (i.e.,  $\alpha_1 = \alpha_2$ )?

**20.7**  $\square$  (Exercise 19.7 continued.) 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  $\text{Bin}(n, p_1)$  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  $\text{Bin}(n, p_2)$  distribution, where  $p_2 = \frac{1}{4}\theta$ . Consider the following two estimators for  $\theta$ :

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

In Exercise 19.7 you showed that both  $T_1$  and  $T_2$  are unbiased estimators for  $\theta$ . Which estimator would you prefer? Motivate your answer.

**20.8**  $\boxplus$  Let  $\bar{X}_n$  and  $\bar{Y}_m$  be the sample means of two independent random samples of size  $n$  (resp.  $m$ ) from the same distribution with mean  $\mu$ . We combine these two estimators to a new estimator  $T$  by putting

$$T = r\bar{X}_n + (1 - r)\bar{Y}_m,$$

where  $r$  is some number between 0 and 1.

- a. Show that  $T$  is an unbiased estimator for the mean  $\mu$ .
- b. Show that  $T$  is most efficient when  $r = n/(n + m)$ .

**20.9** Given is a random sample  $X_1, X_2, \dots, X_n$  from a  $\text{Ber}(p)$  distribution. One considers the estimators

$$T_1 = \frac{1}{n}(X_1 + \dots + X_n) \quad \text{and} \quad T_2 = \min\{X_1, \dots, X_n\}.$$

- a. Are  $T_1$  and  $T_2$  unbiased estimators for  $p$ ?
- b. Show that

$$\text{MSE}(T_1) = \frac{1}{n}p(1 - p), \quad \text{MSE}(T_2) = p^n - 2p^{n+1} + p^2.$$

- c. Which estimator is more efficient when  $n = 2$ ?

**20.10** Suppose we have a random sample  $X_1, \dots, X_n$  from an  $\text{Exp}(\lambda)$  distribution. We want to estimate the expectation  $1/\lambda$ . According to Section 19.4,

$$\bar{X}_n = \frac{1}{n} (X_1 + X_2 + \cdots + X_n)$$

is an unbiased estimator of  $1/\lambda$ . Let us consider more generally estimators  $T$  of the form

$$T = c \cdot (X_1 + X_2 + \cdots + X_n),$$

where  $c$  is a real number. We are interested in the MSE of these estimators and would like to know whether there are choices for  $c$  that yield a smaller MSE than the choice  $c = 1/n$ .

- a. Compute  $\text{MSE}(T)$  for each  $c$ .
- b. For which  $c$  does the estimator perform best in the MSE sense? Compare this to the unbiased estimator  $\bar{X}_n$  that one obtains for  $c = 1/n$ .

**20.11**  $\square$  In Exercise 17.9 we modeled diameters of black cherry trees with the linear regression model (without intercept)

$$Y_i = \beta x_i + U_i$$

for  $i = 1, 2, \dots, n$ . As usual, the  $U_i$  here are independent random variables with  $E[U_i] = 0$ , and  $\text{Var}(U_i) = \sigma^2$ .

We considered three estimators for the slope  $\beta$  of the line  $y = \beta x$ : the so-called least squares estimator  $T_1$  (which will be considered in Chapter 22), the average slope estimator  $T_2$ , and the slope of the averages estimator  $T_3$ . These estimators are defined by:

$$T_1 = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}, \quad T_2 = \frac{1}{n} \sum_{i=1}^n \frac{Y_i}{x_i}, \quad T_3 = \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n x_i}.$$

In Exercise 19.8 it was shown that all three estimators are unbiased. Compute the MSE of all three estimators.

Remark: it can be shown that  $T_1$  is always more efficient than  $T_3$ , which in turn is more efficient than  $T_2$ . To prove the first inequality one uses a famous inequality called the Cauchy Schwartz inequality; for the second inequality one uses Jensen's inequality (can you see how?).

**20.12** Let  $X_1, X_2, \dots, X_n$  represent  $n$  draws *without replacement* from the numbers  $1, 2, \dots, N$  with equal probability. The goal of this exercise is to compute the distribution of  $M_n$  in a way other than by the combinatorial analysis we did in this chapter.

- a. Compute  $P(M_n \leq k)$ , by using, as in Section 8.4, that:

$$P(M_n \leq k) = P(X_1 \leq k, X_2 \leq k, \dots, X_n \leq k).$$

b. Derive that

$$P(M_n = n) = \frac{n!(N-n)!}{N!}.$$

c. Show that for  $k = n + 1, \dots, N$

$$P(M_n = k) = n \cdot \frac{(k-1)!}{(k-n)!} \frac{(N-n)!}{N!}.$$