| | |
|---|---|
| Course: | Mathematical Statistics |
| Term: | Fall 2018 |
| Instructor: | Gordan Žitković |

# Lecture 9
## Estimators

We defined an estimator as any function of the data which is not allowed to depend on the value of the unknown parameters. Such a broad definition allows for very silly examples. To rule out bad estimators and to find the ones that will provide us with the most "bang for the buck", we need to discuss, first, what it means for an estimator to be "good". There is no one answer to this question, and a large part of entire discipline (decision theory) tries to answer it. Some desirable properties of estimators are, however, hard to argue with, so we start from those.

## 9.1 Unbiased estimators

**Definition 9.1.1.** We say that an estimator $\hat{\theta}$ is **unbiased for the parameter** $\theta$ if

$$\mathbb{E}[\hat{\theta}] = \theta.$$

We also define the **bias** $\mathbb{E}[\hat{\theta}] - \theta$ of $\hat{\theta}$, and denote it by $\text{bias}(\hat{\theta})$.

In words, $\hat{\theta}$ is unbiased if we do not expect $\hat{\theta}$ to be systematically above or systematically below $\theta$. Clearly, in order to talk about the bias of an estimator, we need to specify what that estimator is trying to estimate. We will see below that same estimator can be unbiased as an estimator for one parameter, but biased when used to estimate another parameter.

Another important question that needs to be asked about Definition 9.1.1 above is: what does it mean to take the expected value of $\hat{\theta}$, when we do not know what its distribution is. Indeed, $\hat{\theta}$ is a function of the random sample from an *unknown* distribution $D$. In order to take the expected value of $\hat{\theta}$, we need to know $D$, but we do not. What we are really doing is computing the expected value of $\hat{\theta}$ *as a function of* $\theta$. It would, in fact, be more accurate to write $\mathbb{E}^{\theta}[\hat{\theta}]$, because this expected value depends on $\theta$, and the result will be an expression that features $\theta$ somewhere in it. So what do we mean by $\mathbb{E}[\hat{\theta}] = \theta$, then? It means that $\mathbb{E}^{\theta}[\hat{\theta}] = \theta$ for *each possible value* of the parameter $\theta$. In other words, we need a guarantee that the expected value of $\hat{\theta}$ is $\theta$, no matter what $\theta$ nature throws at us.

**Example 9.1.2.** Let $(Y_1, \dots, Y_n)$ be a random sample from $N(\mu, \sigma)$, with both $\mu$ and $\sigma > 0$ unknown, and let

$$\hat{\mu} = \bar{Y} = \frac{Y_1 + \dots + Y_n}{n}.$$

Let us think of $\hat{\mu}$ as an estimator for $\mu$ (that is why we named it $\hat{\mu}$). As a function of the parameter $\theta = (\mu, \sigma)$, the expected value of $\hat{\mu}$ is given by

$$\begin{aligned}\mathbb{E}[\hat{\mu}] = \mathbb{E}[\bar{Y}] &= \tfrac{1}{n}\Big(\mathbb{E}[Y_1] + \mathbb{E}[Y_2] + \dots + \mathbb{E}[Y_n]\Big)\\ &= \tfrac{1}{n}(\mu + \mu + \dots + \mu) = \mu,\end{aligned}$$

which means that $\hat{\mu}$ is unbiased for $\mu$. It is clearly <u>not unbiased</u> if we interpret it as an estimator for $\sigma$. Indeed, its bias in that case is given by

$$\mathrm{bias}(\hat{\mu}) = \mathbb{E}[\hat{\mu}] - \sigma = \mu - \sigma,$$

and this quantity is not equal to 0 for all possible values of the parameters $\mu$ and $\sigma$.

Computing the expected value of an estimator can sometimes be done without knowing its distribution (like in the example above). In general, one needs to know exactly how $\hat{\theta}$ is distributed:

**Definition 9.1.3.** The **sampling distribution** of an estimator $\hat{\theta}$ is its probability distribution (expressed as a function of the unknown parameter(s)).

**Example 9.1.4.** Let $(Y_1, \dots, Y_n)$ be a random sample from $N(\mu, \sigma)$, just like in the previous example, and let $\hat{\mu} = \bar{Y}$ (the sample mean). Since each $Y_i$ is normally distributed with parameters $\mu$ and $\sigma > 0$, the sum $Y_1 + \dots + Y_n$ is also normal, and so is the average $\hat{\mu} = \frac{1}{n}(Y_1 + \dots + Y_n)$. The two parameters of this normal distribution are obtained by computing the mean and the standard deviation of $\hat{\mu}$. We have already computed the mean, namely $\mu$, and, in order to compute the standard deviation, we first compute its variance:

$$\begin{aligned}\mathrm{Var}[\hat{\mu}] = \mathrm{Var}[\tfrac{1}{n}(Y_1 + \dots + Y_n)] &= \tfrac{1}{n^2}\Big(\mathrm{Var}[Y_1] + \dots + \mathrm{Var}[Y_n]\Big)\\ &= \tfrac{1}{n^2}\Big(\sigma^2 + \sigma^2 + \dots + \sigma^2\Big) = \tfrac{n\sigma^2}{n^2} = \tfrac{\sigma^2}{n}.\end{aligned}$$

Therefore $\mathrm{sd}[\hat{\mu}] = \frac{\sigma}{\sqrt{n}}$, and the sampling distribution of $\hat{\mu}$ is $N(\mu, \frac{\sigma}{\sqrt{n}})$.

**Example 9.1.5.** Sampling distributions are sometimes best understood when illustrated by simulations. Consider the elections example, and suppose that the true value of the parameter $p$ is 0.4. The pollster does not know this, all she sees is the data collected from a sample (of size $n = 10$, to fit on the page)

$$A\ A\ B\ A\ B\ A\ B\ A\ B\ B$$

She uses the estimator $\hat{p}$ which computes the sample proprtion of $A$. In this case $\hat{p} = 5/10 = 0.5$. Another pollster, or, if you want to let your imagination run wild, the same pollster in a parallel universe, collected a different sample

$$B\ B\ B\ A\ B\ B\ B\ A\ B\ A$$

Her estimate $\hat{p}$ equals $3/10 = 0.3$. We keep repeating the same for 13 more statisticians, and we get the following list of estimates

$$0.5,\ 0.3,\ 0.1,\ 0.6,\ 0.4,\ 0.5,\ 0.5,\ 0.2,\ 0.3,\ 0.2,\ 0.2,\ 0.3,\ 0.4,\ 0.8,\ 0.2$$

Some statisticians were right on the nose, but some of them were far off (think of the unlucky pollster who got 0.8 - she would predict that the candidate $A$ would win by a landslide, while he or she will, in fact, lose). The point is, of course, that most of the statisticians were relatively close.
We increase the number of statisticians to 2000 and repeat the same procedure. The list of 2000 estimates is too big to be reproduced in these notes, but we can summarize it quite well by a histogram:

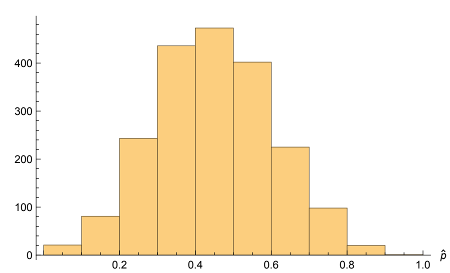


**Figure 1.** Histogram of the values of the estimate $\hat{p}$ obtined by 2000 statisticians

Sometimes our intuition about what is biased and what is not can be a little bit off. Let us start with the case where our intuition works just fine:

**Example 9.1.6.** Let $(Y_1, \dots, Y_n)$ be a random sample from $N(78, \sigma)$. In this case, we know the mean - it happens to be 78 - but not the standard deviation of our observations. Such examples occur when we try to calibrate measuring instruments by using them to measure a known quantity.

We are interested in an unbiased estimator of $\sigma^2$. A natural idea is to use the **sample variance**

$$S^2 = \tfrac{1}{n} \sum_{k=1}^{n} (Y_k - 78)^2.$$

We compute the expected value of $S^2$:

$$\mathbb{E}[S^2] = \tfrac{1}{n} \sum_{k=1}^{n} \mathbb{E}[(Y_k - 78)^2] = \tfrac{1}{n} \sum_{k=1}^{n} \sigma^2 = \sigma^2,$$

where we used the fact that $Y_k \sim N(78, \sigma)$ so that $\mathbb{E}[(Y_k - 78)^2] = \text{Var}[Y_k] = \sigma^2$. Therefore, $S^2$ is unbiased for $\sigma^2$.

Let us now repeat the exercise, but in a slightly different situation:

**Example 9.1.7.** This time, we are using our measuring device to measure an unknown quantity. It is brand new and we have never used it before, so we do not know anything about its error, and we do not have the resources to calibrate it first. Therefore, we need to be able to get some information about both $\mu$ and $\sigma$ at the same time, from the random sample. In other words, we are in the situation where $(Y_1, \dots, Y_n)$ is a random sample from $N(\mu, \sigma)$, with *both* $\mu$ and $\sigma$ unknown. We cannot use the estimator from Example 9.1.6 anymore for the simple reason that we knew that $\mu = 78$ then, but that may not be the case anymore. We cannot replace 78 by the symbol $\mu$ either - $S^2$ would no longer be an estimator. A way out is to first estimate $\mu$ and the use *the estimated value* in its place when computing the sample variance. We already know that $\bar{Y}$ is an unbiased estimator for $\mu$, so we may define[1]

$$S'^2 = \tfrac{1}{n} \sum_{k=1}^{n} (Y_k - \bar{Y})^2. \tag{9.1.1}$$

Let us check whether $S^2$ is an unbiased estimator of $\sigma^2$. We expand the squares in its definition and use the fact that $Y_1 + \dots + Y_n = n\bar{Y}$ in the following computation:

$$\begin{aligned}
\mathbb{E}[S'^2] &= \tfrac{1}{n}\sum_{k=1}^n \mathbb{E}[(Y_k-\bar{Y})^2] = \tfrac{1}{n}\sum_{k=1}^n \mathbb{E}[Y_k^2] - 2\tfrac{1}{n}\sum_{k=1}^n \mathbb{E}[Y_k\bar{Y}] + \tfrac{1}{n}\sum_{k=1}^n \mathbb{E}[\bar{Y}^2] \\
&= \tfrac{1}{n}\sum_{k=1}^n \mathbb{E}[Y_k^2] - 2\mathbb{E}[\tfrac{1}{n}\sum_{k=1}^n Y_k\bar{Y}] + \tfrac{1}{n}n\mathbb{E}[\bar{Y}^2] \\
&= \tfrac{1}{n}\sum_{k=1}^n \mathbb{E}[Y_k^2] - 2\mathbb{E}[\bar{Y}^2] + \mathbb{E}[\bar{Y}^2] = \tfrac{1}{n}\sum_{k=1}^n \mathbb{E}[Y_k^2] - \mathbb{E}[\bar{Y}^2].
\end{aligned}$$

Since $Y_k \sim N(\mu,\sigma)$, we have $\mathbb{E}[Y_k^2] = \mathrm{Var}[Y_k] + (\mathbb{E}[Y_k])^2 = \sigma^2+\mu^2$. On the other hand, $\bar{Y} \sim N(\mu, \sigma/\sqrt{n})$, so $\mathbb{E}[\bar{Y}^2] = \mu^2 + \sigma^2/n$, and it follows that

$$\mathbb{E}[S'^2] = \tfrac{1}{n}n(\mu^2+\sigma^2) - (\mu^2+\sigma^2/n) = (1-\tfrac{1}{n})\sigma^2.$$

Thus, $S'^2$ is *not unbiased*. In fact, since $(1-1/n) < 1$ its bias $\mathbb{E}[S'^2] - \sigma^2 = -\frac{1}{n}\sigma^2$ is always negative.

[1] why we use the notation $S'^2$ and not $S^2$ will be explained shortly.

To find an unbiased estimator all we have to do is multiply the original expression (9.1.1) by $n/(n-1)$, which leads to the following estimator for $\sigma^2$:

$$S^2 = \tfrac{1}{n-1}\sum_{k=1}^n (Y_k-\bar{Y})^2.$$

This, unbiased, version of the sample variance turns out to be even more important than the "naive" one. Therefore, we choose to denote it by $S^2$ and use the longer notation $S'^2$ for the less important version. We do not care to distinguish notationally between the known-$\mu$ and the unknown-$\mu$ case - these will never be used in the same context.

The examples above features some of the most important estimators in statistics; let us list them again in one place:

| **estimator** | **formula** | **notation** |
|---|---|---|
| sample mean | $\frac{Y_1+\dots+Y_n}{n}$ | $\bar{Y}$ |
| sample variance (known mean) | $\frac{1}{n}\sum_{k=1}^n (Y_k-\mu)^2$ | $S^2$ |
| sample variance (unknown mean, biased) | $\frac{1}{n}\sum_{k=1}^n (Y_k-\bar{Y})$ | $S'^2$ |
| sample variance (unknown mean, unbiased) | $\frac{1}{n-1}\sum_{k=1}^n (Y_k-\bar{Y})^2$ | $S^2$ |

## 9.2 Mean-square errors and UMVUE

While being unbiased is a nice property, it alone does not guarantee that the estimator is any good. We would like it to be "close" to the true parameter "most of the time", and not just "right on average". This leads us to following definition

**Definition 9.2.1.** Let $\hat{\theta}$ be an estimator of the parameter $\theta$. Then

1. the **error** of $\hat{\theta}$ is $\hat{\theta} - \theta$,
2. the **absolute error** of $\hat{\theta}$ is $|\hat{\theta} - \theta|$
3. the **relative error** of $\hat{\theta}$ is $\left|\frac{\hat{\theta}-\theta}{\theta}\right|$,
4. the **squared error** of $\hat{\theta}$ is $(\hat{\theta} - \theta)^2$, and
5. the **mean-squared error** of $\hat{\theta}$ is $\text{MSE}(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2]$.
6. the **standard error**[1] of $\hat{\theta}$ is $\text{se}(\hat{\theta}) = \sqrt{\text{Var}(\hat{\theta})}$.

[1]even though this is really just the standard deviation of the estimator $\hat{\theta}$, it is given its own name so that it does not get confused with the population standard deviation.

All of the above quantify the discrepancy between the true value of the parameter $\theta$ and its estimate $\hat{\theta}$. Unfortunately, none of them are available to the statistician, even after the data are collected - they all depend on the parameter $\theta$. Indeed, if we knew the error exactly, there would be no need for statistics - we would simply adjust the value of $\hat{\theta}$ by that error and get $\theta$ exactly.

In some cases, however, we can say a great deal about the mean-squared error[1] and use it to find good estimators. Just like the other quantities defined above, $\text{MSE}(\hat{\theta})$ is not a number - it is a function of the parameter $\hat{\theta}$.

**Example 9.2.2.** Consider the elections example, where $(Y_1, \dots, Y_n)$ is a random sample from the $B(p)$-distribution, with an unknown parameter $p \in (0, 1)$. We consider two estimators

$$\hat{p}_1 = \bar{Y} = Y/n, \text{ and } \hat{p}_2 = \tfrac{1+Y}{n+2},$$

where $Y = \sum_{i=1}^{n} Y_i$ is the number of votes for candidate $A$. To compute the mean-square errors of $\hat{p}_1$ and $\hat{p}_2$ we note first that $Y \sim b(n, p)$, so that $\mathbb{E}[Y] = np$ and $\mathbb{E}[Y^2] = \text{Var}[Y] + (\mathbb{E}[Y])^2 = np(1-p) + n^2p^2$.

[1]the reason we are using squared error $(\hat{\theta} - \theta)^2$ instead of the absolute error $|\hat{\theta} - \theta|$ is the same reason we use the standard deviation and not the absolute deviation - mathematics is much simpler.

Therefore

$$\begin{aligned}\mathrm{MSE}(\hat{p}_1) &= \mathbb{E}[(\tfrac{Y}{n} - p)^2] = \tfrac{1}{n^2}\mathbb{E}[Y^2] - \tfrac{2p}{n}\mathbb{E}[Y] + p^2 \\ &= (\tfrac{p(1-p)}{n} + p^2) - 2p^2 + p^2 = \tfrac{p(1-p)}{n}.\end{aligned}$$

Similarly, but with a bit more algebra,

$$\begin{aligned}\mathrm{MSE}(\hat{p}_2) &= \mathbb{E}[(\tfrac{1+Y}{n+2} - p)^2] = \tfrac{1}{(n+2)^2}\mathbb{E}[(Y+1-p(n+2))^2] \\ &= \tfrac{1}{(n+2)^2}\Big(\mathbb{E}[Y^2] + 2(1-p(n+2))\mathbb{E}[Y] + (1-p(n+2))^2\Big) \\ &= \tfrac{1+p(1-p)(n-4)}{(n+2)^2}\end{aligned}$$

These two expression do not mean much, but if we plot them on the same graph, we can easily compare the two estimators:

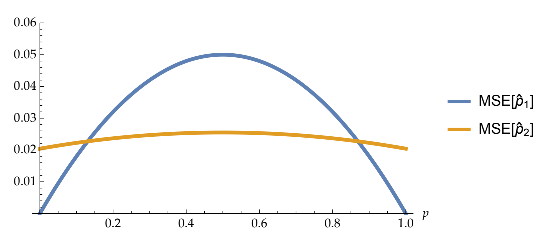


**Figure 2.** Mean-square errors for $\hat{p}_1$ and $\hat{p}_2$ with $n = 5$.

Indeed, the first one is better for extreme values of $p$, and the second one for $p$ closed to $1/2$. For larger values of $n$, the difference is not so dramatic:

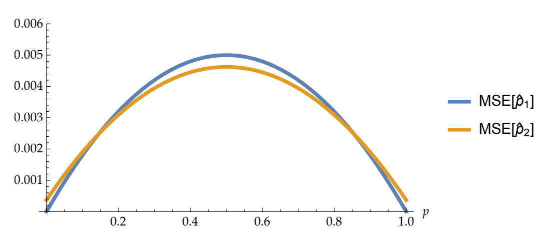


**Figure 3.** Mean-square errors for $\hat{p}_1$ and $\hat{p}_2$ with $n = 50$.

Without any knowledge about the true value of $p$, it is hard to decide which estimator to use. Sometimes, however, a single estimator beats all others, no matter what the value of the parameter is.

**Definition 9.2.3.** An estimator $\hat{\theta}$ for $\theta$ is said to be the **uniformly minimum variance unbiased estimator (UMVUE)** if

1. $\hat{\theta}$ is unbiased, and
2. $\text{MSE}(\hat{\theta}) \leq \text{MSE}(\hat{\theta}')$ for all $\theta$, and all unbiased estimators $\hat{\theta}'$ of $\theta$.

The reason for restricting to unbiased estimators only in the previous definition is simple. One can always construct biased (and very bad) estimators which outperform any other estimator for a particular value of the parameter. Simply set $\hat{\theta} = 0.8$, no matter what the data say. If it happens that $\theta = 0.8$, $\hat{\theta}$ provides a perfect estimate (*Even a broken clock is right twice a day.*) Also, once an unbiased estimator is chosen, some of the computations simplify significantly:

**Proposition 9.2.4.** *We have*

$$\text{MSE}(\hat{\theta}) = (\text{se}(\hat{\theta}))^2 + (\text{bias}(\hat{\theta}))^2.$$

*In particular, if $\hat{\theta}$ is unbiased,* $\text{MSE}(\hat{\theta}) = (\text{se}(\hat{\theta}))^2$.

*Proof.* By writing $(\hat{\theta} - \theta)^2 = (\hat{\theta} - \mathbb{E}[\hat{\theta}] + \mathbb{E}[\hat{\theta}] - \theta)^2$, and then expanding the square

$$(\hat{\theta} - \theta)^2 = (\hat{\theta} - \mathbb{E}[\hat{\theta}])^2 + 2(\hat{\theta} - \mathbb{E}[\hat{\theta}])(\mathbb{E}[\hat{\theta}] - \theta) + (\mathbb{E}[\hat{\theta}] - \theta)^2,$$

we can write $\text{MSE}(\hat{\theta}) = \mathbb{E}(\hat{\theta} - \theta)^2$ as a sum of three terms

$$\text{MSE}(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2] + 2\mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])(\mathbb{E}[\hat{\theta}] - \theta)] + \mathbb{E}[(\mathbb{E}[\hat{\theta}] - \theta)^2].$$

The first term is the variance of $\hat{\theta}$, i.e., $\text{se}(\hat{\theta})^2$. The expression inside the expectation in the third term is not random, so the entire term equals $(\mathbb{E}[\hat{\theta}] - \theta)^2$, which equals $\text{bias}(\hat{\theta})^2$. Finally, the expression inside the expectation in the second term is the product the constant $(\mathbb{E}[\hat{\theta}] - \theta)$ and the random variable $(\hat{\theta} - \mathbb{E}[\theta])$. The constant can be pulled out of the expectation:

$$\mathbb{E}[(\hat{\theta} - \mathbb{E}[\theta])(\mathbb{E}[\hat{\theta}] - \theta)] = (\mathbb{E}[\hat{\theta}] - \theta)\mathbb{E}[\hat{\theta} - \mathbb{E}[\hat{\theta}]] = 0,$$

since $\mathbb{E}[\hat{\theta} - \mathbb{E}[\hat{\theta}]] = \mathbb{E}[\hat{\theta}] - \mathbb{E}[\hat{\theta}] = 0$. □

We have not developed all the mathematics needed to establish the UMVUE property of some of the popular estimators (we need the notions of sufficiency and completeness, and the Rao-Blackwell theorem), but here are some examples without proof:

**Example 9.2.5.** The following are UMVUEs:

1. $\bar{Y}$, for $\mu$, when $(Y_1, \dots, Y_n)$ is a random sample from $N(\mu, 1)$.
2. $\bar{Y}$, for $p$, when $(Y_1, \dots, Y_n)$ is a random sample from $B(p)$
3. $S^2 = \frac{1}{n-1} \sum_{k=1}^{n} (Y_i - \bar{Y})^2$, for $\sigma$, when $(Y_1, \dots, Y_n)$ is a random sample from $N(\mu, \sigma)$

We can say more about minimum-variance operators in some restricted classes:

**Definition 9.2.6.** An estimator $\hat{\mu}$ is said to be **linear** if it is a linear function of the data, i.e., if it of the form

$$\hat{\mu} = \alpha_1 Y_1 + \alpha_2 Y_2 + \cdots + \alpha_n Y_n,$$

for some constants $\alpha_1, \dots, \alpha_n$.

**Example 9.2.7.** The sample mean $\bar{Y}$ is linear since it equals $\alpha_1 Y_1 + \cdots + \alpha_n Y_n$ with $\alpha_1 = \alpha_2 = \cdots = \alpha_n = \frac{1}{n}$.

The sample variance $S^2$ is not linear, as its formula

$$S^2 = \frac{1}{n-1} \sum_{k=1}^{n} (Y_k - \bar{Y})^2,$$

involves terms like $Y_k^2$ and $Y_k Y_l$.

**Definition 9.2.8.** A linear estimator $\hat{\theta}$ is said to be the **best linear unbiased estimator (BLUE)** if

1. $\hat{\theta}$ is unbiased, and
2. $\text{MSE}(\hat{\theta}) \leq \text{MSE}(\hat{\theta}')$ for all $\theta$, and all unbiased <u>linear</u> estimators $\hat{\theta}'$ of $\theta$.

In words, BLUE is the best estimator (measured in mean-square error), but only among all linear estimators. It may happen that a better estimator exists, but then it cannot be linear.

**Example 9.2.9.** Consider a random sample $(Y_1, \dots, Y_n)$ from the $N(\mu, \sigma)$-distribution with unknown $\mu$ and $\sigma$. For a linear estimator

$$\hat{\theta} = \alpha_1 Y_1 + \dots + \alpha_n Y_n,$$

we have

$$\mathbb{E}[\hat{\theta}] = \sum_{k=1}^{n} \alpha_k \mathbb{E}[Y_k] = \sum_{k=1}^{n} \alpha_k \mu = \mu\Big(\alpha_1 + \dots + \alpha_n\Big),$$

and we conclude that $\hat{\theta}$ is unbiased if and only if $\alpha_1 + \dots + \alpha_n = 1$. Assuming that that is the case, and remembering that the MSE equals variance for unbiased estimators, we have

$$\begin{aligned} \mathrm{MSE}(\hat{\theta}) &= \mathrm{Var}[\alpha_1 Y_1 + \dots + \alpha_n Y_n] = \alpha_1^2 \mathrm{Var}[Y_1] + \dots + \alpha_n^2 \mathrm{Var}[Y_n] \\ &= \sigma^2(\alpha_1^2 + \dots + \alpha_n^2). \end{aligned}$$

At this point, we are faced with the following minimization problem:

minimize the sum $\alpha_1^2 + \dots + \alpha_n^2$, over all $\alpha_1, \dots, \alpha_n$ such that $\sum_{k=1}^{n} \alpha_k = 1$.

Intuitively, the sum of squares is going to be minimized when all $\alpha_i$ are the same. To show that that is indeed the case, we start from $n = 2$. In this case $\alpha_1 + \alpha_2 = 1$, and, so

$$\alpha_1^2 + \alpha_2^2 = \alpha_1^2 + (1 - \alpha_1)^2 = 2\alpha_1^2 - 2\alpha_1 + 1.$$

We differentiate the right-hand side in $\alpha_1$ and set the result to 0 to get $4\alpha_1 - 2 = 0$, i.e., $\alpha_1 = 1/2$. It is not hard to check that this is, indeed, the unique minimum.

The general case follows from the well-known **quadratic-arithmetic** inequality, which holds for all $\alpha_1, \dots, \alpha_n \geq 0$:

$$\sqrt{\tfrac{\alpha_1^2 + \dots + \alpha_n^2}{n}} \geq \tfrac{\alpha_1 + \dots + \alpha_n}{n}. \tag{9.2.1}$$

It is not very hard to prove - simply square both sides, multiply by $n^2$ and subtract the right-hand side from the left-hand side. You will get (can you see why?)

$$(\alpha_1 - \alpha_2)^2 + (\alpha_1 - \alpha_3)^2 + \dots + (\alpha_{n-1} - \alpha_n)^2.$$

As a sum of squares, this quantity is always nonnegative; therefore the left-hand side of (9.2.1) is always larger than its right-hand side. Moreover, they are only equal if all the expressions inside the squares are 0, i.e., if all $\alpha_i$ are equal.

In our special case, when $\sum_i \alpha_i = 1$, the quadratic-arithmetic inequality becomes

$$\sqrt{\tfrac{1}{n}(\alpha_1^2 + \cdots + \alpha_2^2)} \geq \tfrac{1}{n}, \text{ i.e. } \alpha_1^2 + \cdots + \alpha_2^2 \geq \tfrac{1}{n},$$

and the equality is attained when all $\alpha_1$ are the same, i.e., when $\alpha^i = \frac{1}{n}$ for each $i$.

Therefore, the sample mean $\bar{Y}$ is the unique BLUE in this model.

## 9.3 Problems

**Problem 9.3.1.** Which of the following estimators is **not** unbiased for $\mu$ if $Y_1, \ldots, Y_n$ is a random sample from the normal distribution $N(\mu, \sigma)$:

(a) $Y_n$ (b) $\frac{1}{2}(Y_1 + Y_2)$ (c) $Y_1 - Y_2 + Y_3$ (d) $\bar{Y}$ (e) all of the above are unbiased

**Problem 9.3.2.** Let $Y_1, \ldots, Y_n$ be a random sample from a uniform distribution $U(0, \theta)$, with the parameter $\theta > 0$. The quantity

$$\hat{\theta} = \tfrac{c}{n} \sum_{i=1}^{n} Y_i^2$$

is an unbiased estimator for $\theta$ when

(a) $c = 3$ (b) $c = 3/\theta^2$ (c) $c = 2/\theta^2$ (d) $c = 1$ (e) none of the above

**Problem 9.3.3.** Let $Y_1, \ldots, Y_n$ be a random sample from $U(0, \theta)$, with an unknown $\theta > 0$. For what value of the constant $c$ is the estimator $\hat{\theta} = c \sum_{i=1}^{n} Y_i$ unbiased for $\theta$?

(a) 1 (b) $1/n$ (c) $2/n$ (d) $n$ (e) none of the above

**Problem 9.3.4.** Let $(Y_1, \ldots, Y_n)$ be a random sample from the normal distribution with mean $\mu \in \mathbb{R}$ and standard deviation $\sigma > 0$.

1. Is $\hat{\mu} = \frac{Y_1 + Y_n}{2}$ unbiased for $\mu$? What is it mean square error?
2. Repeat the above but consider $\hat{\mu}$ to be an estimator for $\sigma$.

**Problem 9.3.5.** Let $Y_1, \ldots, Y_n$ be a random sample from the uniform distribution on $[0, \theta]$, where $\theta > 0$ is an unknown parameter. We consider the estimator

$$\hat{\theta} = c \, \frac{1}{n} \sum_{i=1}^{n} Y_i^2.$$

where $c$ is a constant (not dependent on $\theta$ or on $Y_1, \ldots, Y_n$).

1. For what value of the constant $c$ will $\hat{\theta}$ be an unbiased estimator for $\theta^2$? Is there such a value if $\hat{\theta}$ is used as an estimator for $\theta$ instead of $\theta^2$?

2. Using the value of $c$ obtained above, compute the mean squared error of $\hat{\theta}$ (when interpreted as an estimator of $\theta^2$).

**Problem 9.3.6.** Let $Y_1, \dots, Y_n$ be a random sample from the uniform distribution $U(0, \theta)$, with parameter $\theta > 0$. The MSE (mean-squared error) of the estimator $\hat{\theta} = c\bar{Y}$ for $\theta$ is the smallest when the constant $c$ equals

(a) $\frac{1}{2}$ (b) 2 (c) $\frac{6n}{3n+1}$ (d) $\frac{3n}{6n+1}$ (e) none of the above

**Problem 9.3.7.** Let $Y_1, \dots, Y_n$ be a random sample from the normal distribution $N(\mu, 1)$, with the unknown parameter $\mu$ and a known variance, equal to 1. The MSE (mean-squared error) of the estimator $\hat{\theta} = \bar{Y} + c$ for $\mu$ is the smallest when the constant $c$ equals

(a) $-\mu$ (b) 0 (c) 1 (d) $\frac{1}{n}$ (e) none of the above

**Problem 9.3.8.** Let $Y_1, \dots, Y_n$ be a random sample of size $n \geq 2$, from $N(\mu, \sigma)$ and let the estimators $\hat{\mu}_1$, $\hat{\mu}_2$ and $\hat{\mu}_3$, for $\mu$, be given by

$$\hat{\mu}_1 = Y_1, \hat{\mu}_2 = \tfrac{1}{2}(Y_1 + Y_2) \text{ and } \hat{\mu}_3 = \bar{Y}.$$

Then, no matter what $\mu$ and $\sigma$ are, we always have

(a) $\text{MSE}(\hat{\mu}_1) \leq \text{MSE}(\hat{\mu}_2) \leq \text{MSE}(\hat{\mu}_3)$

(b) $\text{MSE}(\hat{\mu}_3) \leq \text{MSE}(\hat{\mu}_2) \leq \text{MSE}(\hat{\mu}_1)$

(c) $\text{MSE}(\hat{\mu}_3) \leq \text{MSE}(\hat{\mu}_1) \leq \text{MSE}(\hat{\mu}_2)$

(d) $\text{MSE}(\hat{\mu}_1) \leq \text{MSE}(\hat{\mu}_3) \leq \text{MSE}(\hat{\mu}_2)$

(e) none of the above

**Problem 9.3.9.** Let $Y_1, \dots, Y_n$ be a random sample from the normal distribution with mean 0 and an unknown standard deviation $\sigma > 0$. The mean-squared error (MSE) of the estimator $\hat{\theta} = \frac{1}{n}\sum_{i=1}^n Y_i^2$ for $\sigma^2$ is

(a) $\frac{\sigma^4}{2}$ (b) $\frac{\sigma^2}{2}$ (c) $\frac{2\sigma^4}{n}$ (d) $\frac{\sigma^4}{2n}$
(e) none of the above

(*Hint:* What is the distribution of $\frac{1}{\sigma^2}\sum_i Y_i^2$?)

**Problem 9.3.10.** Let $(Y_1, Y_2)$ be a random sample (of size $n = 2$) from the uniform distribution $U(0, \theta)$, with $\theta > 0$ unknown.

1. Find constants $c_1, c_2$ and $c_3$ such that the following estimators

$$\hat{\theta}_1 = c_1 Y_1,\ \hat{\theta} = c_2 Y_2 \text{ and } \hat{\theta}_3 = c_3 \max(Y_1, Y_2),$$

   are unbiased. (*Hint:* For $\hat{\theta}_3$, integrate the function $\max(y_1, y_2)$ multiplied by the joint density of $Y_1, Y_2$. Split the integral over $[0, \theta] \times [0, \theta]$ into two parts - one where $y_1 \geq y_2$ and the other where $y_1 < y_2$ and note that $\max(y_1, y_2) = y_1 \mathbf{1}_{\{y_1 \geq y_2\}} + y_2 \mathbf{1}_{\{y_1 < y_2\}}$.)

2. With values $c_1, c_2$ and $c_3$ as above, compute mean-square errors $\mathrm{MSE}(\hat{\theta}_1)$, $\mathrm{MSE}(\hat{\theta}_2)$ and $\mathrm{MSE}(\hat{\theta}_3)$ of $\hat{\theta}_1$, $\hat{\theta}_2$ and $\hat{\theta}_3$.

3. Sketch the graphs of $\mathrm{MSE}(\hat{\theta}_1)$, $\mathrm{MSE}(\hat{\theta}_2)$ and $\mathrm{MSE}(\hat{\theta}_3)$ as a functions of $\theta$. Is one of the three clearly better (in the mean-square sense) than the others?

**Problem 9.3.11.** A random variable with pdf

$$f_Y(y) = \tfrac{1}{2\tau} e^{-|y|/\tau},$$

is called a **double exponential distribution**, with parameter $\tau > 0$.

1. Compute $\mathbb{E}[Y]$ and $\mathbb{E}[|Y|]$ if $Y$ is double exponential with paramter $\tau$. (*Hint:* Use the fundamental formula to get an integral for $\mathbb{E}[Y]$ and note that the function under the integral is odd in the case of $\mathbb{E}[Y]$, and even in the case of $\mathbb{E}[|Y|]$. That makes the computation of $\mathbb{E}[Y]$ immediate, and reduces the computation of $\mathbb{E}[|Y|]$ to (twice) a simple integral which does not contain the absolute value. )

Let now $Y_1, \dots, Y_n$ be a random sample from the double exponential distribution with an unknown parameter $\tau > 0$, and let the estimator $\hat{\tau}$ be defined by

$$\hat{\tau} = \tfrac{1}{n} \sum_{i=1}^{n} |X_i| \, .$$

2. Is $\hat{\tau}$ unbiased?

3. Compute $\mathrm{MSE}(\hat{\tau})$?