- Properties of Estimators: MSE
- Bias and Unbiasedness
- Asymptotic Unbiasedness
- Bias-Variance Decomposition of MSE

Best MSE Estimators

Now, having the idea of the Error of estimation for an Estimator, we can try to find the best Estimator in the sense of MSE, i.e., we can try to solve

$$\text{Find } \hat{\theta} \text{ among all Estimators that minimizes } MSE(\hat{\theta}, \theta).$$

Unfortunately, this Minimization problem has no solution in general.

So we move forward considering other Properties of Estimators making them useful in the estimation process.

Bias, Biased and Unbiased Estimators

Definition: The **Bias** of Estimator $\hat{\theta}$ of θ is

$$Bias(\hat{\theta}, \theta) = \mathbb{E}_\theta(\hat{\theta} - \theta) = \mathbb{E}_\theta(\hat{\theta}) - \theta.$$

Definition: We say that our Estimator $\hat{\theta}$ (of the parameter θ) is **Unbiased**, if

$$Bias(\hat{\theta}, \theta) = 0 \qquad \text{for any} \qquad \theta \in \Theta.$$

If $\hat{\theta}$ is not Unbiased, then we say that it is **Biased**.

In other words, the Estimator $\hat{\theta}$ is **Unbiased**, if

$$\mathbb{E}_\theta\left(\hat{\theta}\right) = \theta \qquad \text{for any} \qquad \theta \in \Theta.$$

Note: We require the above hold **for any parameter value** since, if, say, it is correct for *some* values of θ, then it can happen that the true value of our unknown θ is exactly that value, for whihc we do not have the equality.

Bias and Unbiasedness

- The idea of Unbiased Estimator is the following: if we will calculate Estimates many-many times using our Unbiased Estimator, and then average the obtained Estimates, we will obtain the (almost exact, exact when many-many$\rightarrow \infty$) value of our Parameter.
- So we can say about an Unbiased Estimator as: **In average, it is Exact**
- Bias can be interpreted, in some sense, as the *accuracy* of the Estimator.

Example

Example: Assume we have a Sample of size n from some Population, and we want to estimate the Mean of that Population. Let $\mathcal{F}_\mu$ be the Distribution of our Population, and μ be the Mean (Expectation) of that Distribution. So our unknown Parameter here is μ.

We use size n Random Sample to Model the situation:

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\mu.$$

2) $\hat{\mu}_3 = \frac{X_1+X_4}{10}$ - Biased.

Bias $(\hat{\mu}_3, \mu) = \mathbb{E}(\hat{\mu}_3 - \mu) = \mathbb{E}\left(\frac{X_1+X_4}{10} - \mu\right) =$
$= \frac{\mathbb{E}(X_1)+\mathbb{E}(X_4)}{10} - \mu = \frac{\mu+\mu}{10} - \mu = -\frac{8}{10}\mu = -\frac{4}{5}\mu$

Now, we choose several Estimators to estimate μ:

$$\hat{\mu}_1 = X_1, \qquad \hat{\mu}_2 = \frac{X_1 + X_3}{2}, \qquad \hat{\mu}_3 = \frac{X_1 + X_4}{10},$$

$$\hat{\mu}_4 = \overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}.$$

Problem: Check if each Estimator is Biased or Unbiased.

Solution: OTB

Example

Example: Let's do an experiment with Biased and Unbiased Estimators, with **R**.

UnBiased Estimator Case

We consider the Poisson Model:

$$X_1, X_2, ..., X_{10} \sim Pois(\lambda)$$

and we want to estimate λ. We consider the following Estimator:

$$\hat{\lambda} = \frac{X_1 + X_2 + ... + X_{10}}{10}.$$

Easy to se that $\hat{\lambda}$ is an Unbiased Estimator for λ (OTB!).

Biased Estimator Case

Say, let us consider the Exponential Model:

$$X_1, X_2, ..., X_{10} \sim Exp(\lambda)$$

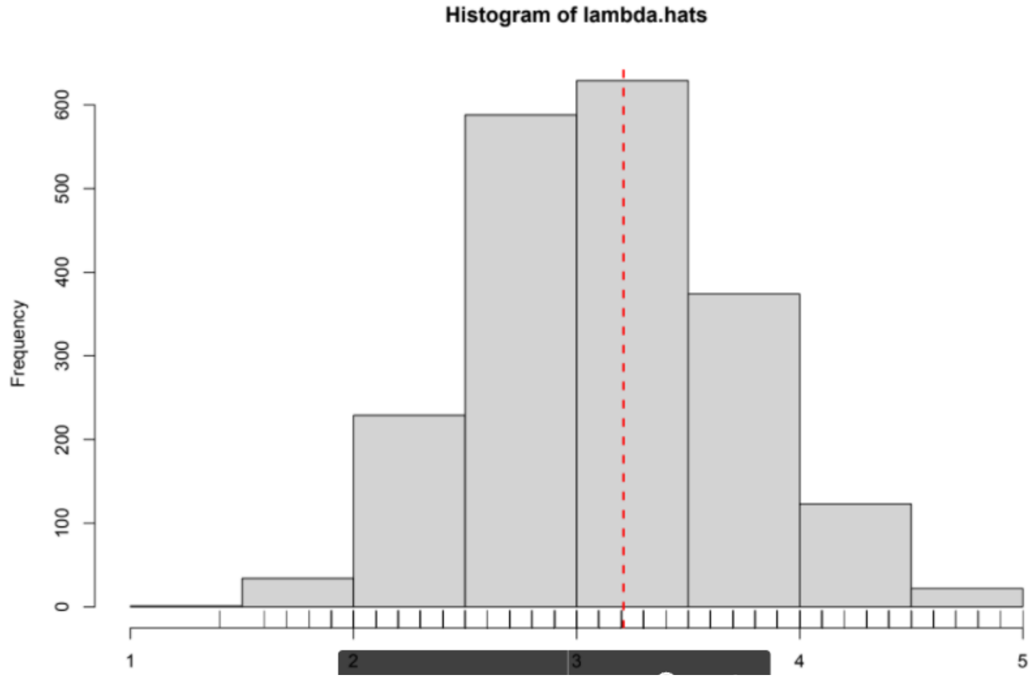
and we want to estimate λ. We consider the following Estimator:

$$\hat{\lambda} = \frac{X_1 + X_2 + ... + X_{10}}{10}.$$

Easy to see that $\hat{\lambda}$ is an Biased Estimator for λ (OTB!).

With a Histogram:

```
hist(lambda.hats)
rug(lambda.hats)
abline(v = lambda, col="red", lwd = 2, lty = 2)
```



Now, the code

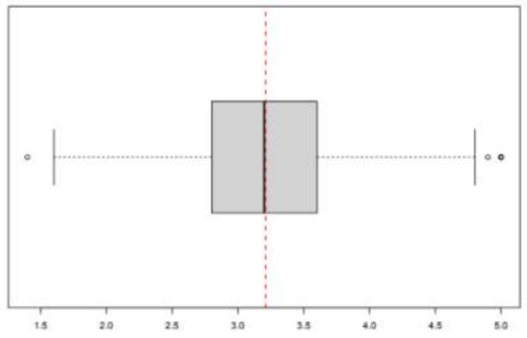
- observing once: generating a Sample just once and calculating one Estimate:

```
lambda <- 3.21
x <- rpois(10, lambda = lambda)
lambda.hat <- mean(x)
lambda.hat
```

```
## [1] 2.5
```

- observing many times: generating Samples many times, calculating Estimates, and then averaging:

```
lambda <- 3.21; n <- 10; m <- 2000
x <- rpois(n*m, lambda = lambda)
x <- as.data.frame(matrix(x, ncol = m))
lambda.hats <- sapply(x, mean)
boxplot(lambda.hats, horizontal = T);
abline(v = lambda, col="red", lwd = 2, lty = 2)
```

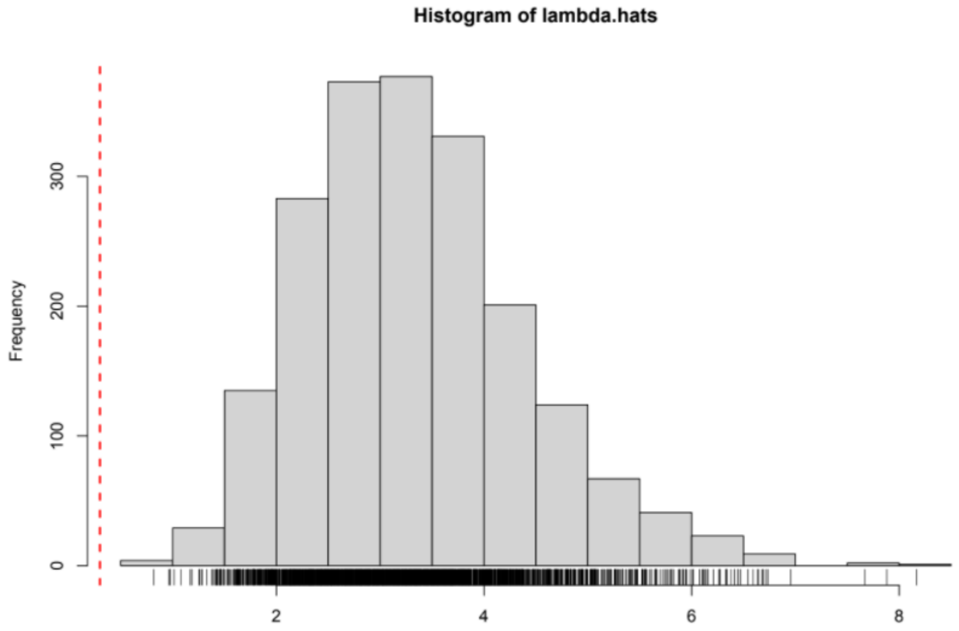



```
mean(lambda.hats)
```

```
## [1] 3.1901
```

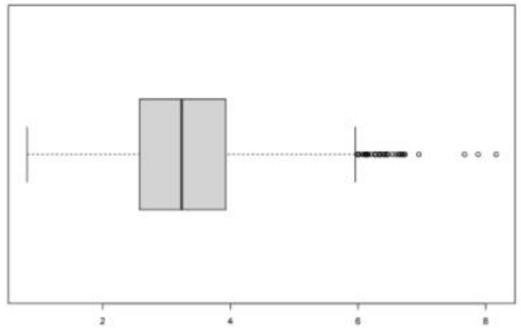
With a Histogram:

```r
hist(lambda.hats)
rug(lambda.hats)
abline(v = lambda, col="red", lwd = 2, lty = 2)
```



- observing many times: generating Samples many times, calculating Estimates, and then averaging:

```r
lambda <- 0.3; n <- 10; m <- 2000
x <- rexp(n*m, rate = lambda)
x <- as.data.frame(matrix(x, ncol = m))
lambda.hats <- sapply(x, mean)
boxplot(lambda.hats, horizontal = T);
abline(v = lambda, col="red", lwd = 2, lty = 2)
```




```r
mean(lambda.hats)
```

```
## [1] 3.331708
```

Example

Example: Assume we have a Random Sample for a some Distribution with the Mean μ and Variance σ^2:

$$X_1, X_2, ..., X_n \sim \mathcal{F}_{\mu,\sigma^2},$$

and we want to estimate the Parameters μ and σ^2.

We consider the following Estimators:

$$\hat{\mu} = \overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}$$

and

$$\widehat{\sigma^2} = \frac{\sum_{k=1}^{n}(X_k - \overline{X}_n)^2}{n} \qquad \text{and} \qquad \widehat{\sigma^2} = S^2 = \frac{\sum_{k=1}^{n}(X_k - \overline{X}_n)^2}{n-1}$$

Let us see (OTB) which ones are Biased and which ones are not.

Asymptotic Unbiasedness

Some years ago Unbiasedness was a very important, desirable property from any good Estimator. Nowadays, it is enough to have *Asymptotic Unbiasedness*:

Definition: Estimator $\hat{\theta}_n$ is called **Asymptotically Unbiased** for θ, if

$$Bias(\hat{\theta}_n, \theta) \to 0, \qquad \text{for any } \theta \in \Theta.$$

Idea: If the Sample size is very large, then the behaviour of our Asymptotic Unbiased Estimator is close to an Unbiased one, $Bias(\hat{\theta}_n, \theta) \approx 0$

Example: Say, for the Mean μ of the Population,

$$\hat{\mu} = \frac{X_1 + X_2 + ... + X_n}{n+1}$$

is a Biased, but Asymptotically Unbiased Estimator. OTB, please!

Bias-Variance Decomposition

This is an important result:

Theorem(Bias-Variance Decomposition of the MSE): If $\hat{\theta}$ is an Estimator for θ, then

$$MSE(\hat{\theta}, \theta) = \left(Bias(\hat{\theta}, \theta)\right)^2 + Var_{\theta}(\hat{\theta}).$$

Proof: OTB

Bias-Variance Decomposition

Again, let's recall our BVD:

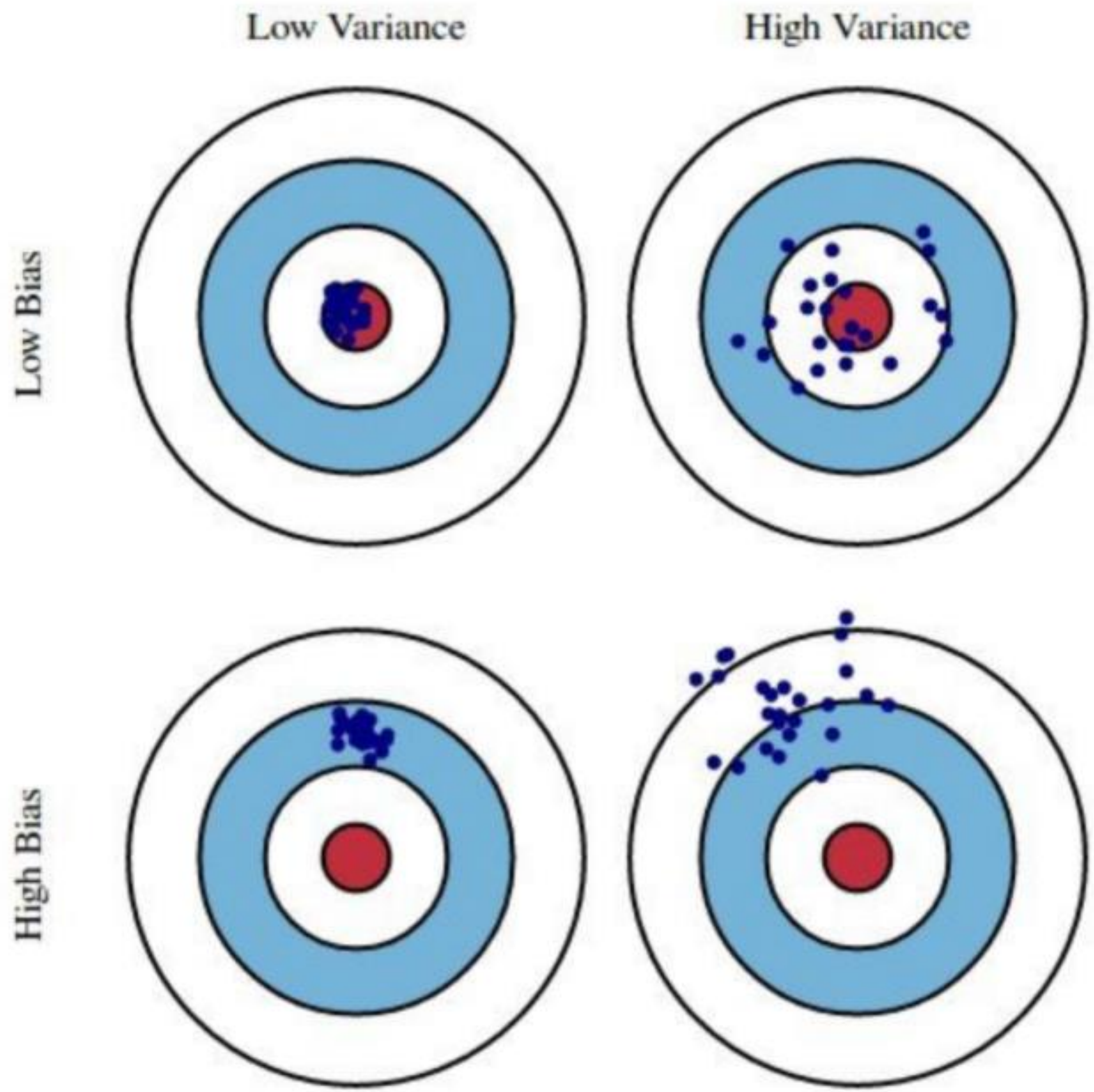
$$MSE(\hat{\theta}, \theta) = \left(Bias(\hat{\theta}, \theta)\right)^2 + Var_\theta(\hat{\theta}).$$

We interpret the RHS:

- Bias is the **Accuracy** of our Estimator $\hat{\theta}$
- Variance is the **Precision** of our Estimator $\hat{\theta}$

Nice Graphical Interpretation: Link, see also the next slide.

Bias-Variance Decomposition/Tradeoff



Standard Error and Estimated Standard Error

Definition: The Standard Deviation of the Estimator is called the **Standard Error** of the Estimator $\hat{\theta}$ and is denoted by

$$SE(\hat{\theta}) = SD(\hat{\theta}) = \sqrt{Var_{\theta}(\hat{\theta})}.$$

Usually, the Standard Error will depend on the unknown value of the Parameter θ. If we use the Estimator $\hat{\theta}$, then the **Estimated Standard Error** of $\hat{\theta}$, $\widehat{SE}(\hat{\theta})$ is the Standard Error, where after calculation we plug $\hat{\theta}$ instead of θ.

And statisticians, when reporting the Estimate, usually report also the Estimated Standard Error, as a measure how precise is the result. If the Standard Error is small (and we are using a nice Estimator, say, it is Unbiased), then this is a sign that the result is close to real/actual one.

Example

Example: Assume we are facing an election with Parties A and B, and we want to estimate the percentage of voters for A in advance. So we do a poll, asking 10 persons to give their preferences. Let the result be:

$$A, B, B, B, A, B, B, A, B, B.$$

Problem: Estimate the percentage of voters for the Party A, and give the Estimated Standard Error.

Solution: OTB.