Foundations of statistical estimation: Consistency, bias and variance

Statistical Computing and Empirical Methods
Unit EMATM0061, Data Science MSc

Rihuan Ke

rihuan.ke@bristol.ac.uk

Teaching Block 1, 2024

University of BRISTOL

What we will cover today

We will view sample statistics as estimators of parameters of interest.

We will discuss the concept of statistical consistency

We will also introduce the ideas of statistical bias and bias-variance decomposition.

We will also consider the concept of a minimum variance unbiased estimator.

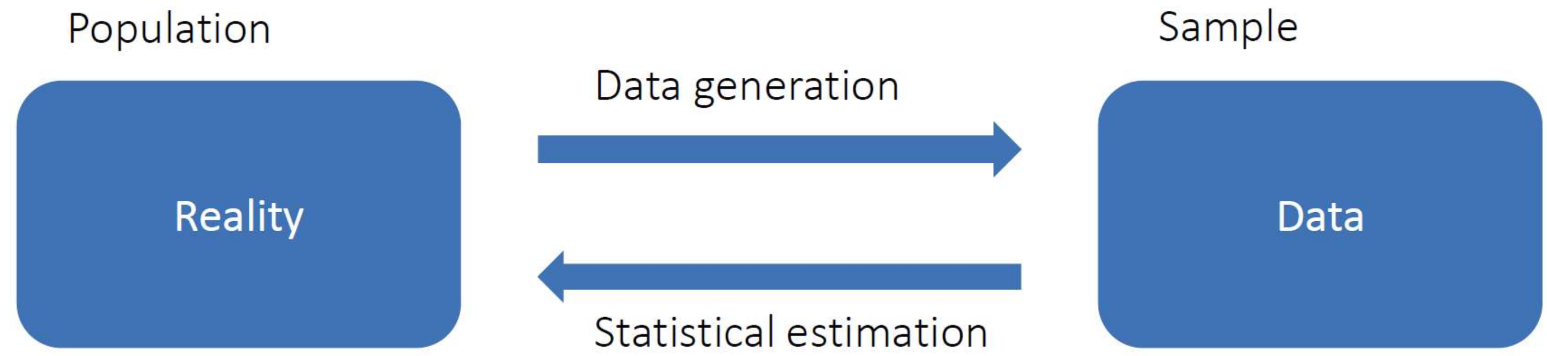
Samples and populations

We attempt to answer questions about the population by looking at data (random samples).

Sample (e.g., Palmer Penguins Dataset)

Population (e.g., all penguins in the areas of interest)

Statistical estimation and probability



A distribution P_X for a random variable X

A sample of independent copies $X_1, \cdots, X_n \sim P_X$.

	Population statistics	Sample statistics
Mean	$\mu = \mathbb{E}(X)$	$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$
Variance	$\sigma^2 = \mathbb{E}\left[(X-\mu)^2\right]$	$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$

1. Probabilistic model

It is often useful to model our data as being generated by a probabilistic model $\mathbb{P}_\theta$ (parametrized by θ).

Example 1. Suppose we have a sequence $(X_1, \cdots, X_n)$ in $\{0,1\}^n$ corresponding to pass or fail for a driving test.

We can model $(X_1, \cdots, X_n)$ as a sequence of independent and identically distributed Bernoulli random variables

$$X_1, \cdots, X_n \sim \mathcal{B}(q), \qquad \theta = q$$

Example 2. Suppose we have a sequence $(X_1, \cdots, X_n)$ in $\mathbb{R}^n$ corresponding to the height of a penguin.

We can model $(X_1, \cdots, X_n)$ as a sequence of independent and identically distributed Gaussian Random variables

$$X_1, \cdots, X_n \sim \mathcal{N}(\mu, \sigma^2), \qquad \theta = (\mu, \sigma^2)$$

2. Statistical estimation

We often use prior knowledge to choose the form of our model e.g. Gaussian, Bernoulli, $\cdots$, and perform tasks based on the probabilistic model $\mathbb{P}_\theta$.

Tasks. We need to estimate the parameters θ in our model based upon a sample $X_1, \cdots, X_n \sim \mathbb{P}_\theta$.

Approach. We estimate our parameters based on sample statistics, which are functions of the samples

$$\hat{\theta} = g(X_1, \cdots, X_n) \text{ that don't depend on } \theta.$$

Goal. Find sample statistics $\hat{\theta} = g(X_1, \cdots, X_n)$ approximating θ.

Note that sample statistics depend on the sample, so are themselves random variables.

Statistical estimation

Goal. Find sample statistics $\hat{\theta} = g(X_1, \cdots, X_n)$ approximating θ.

Example 1. Suppose that $X_1, \cdots, X_n \sim \mathcal{B}(q)$ are i.i.d. with a single parameter $\theta = q$.

We estimate the population mean $q = \mathbb{E}(X_i)$ with the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. So $g(x_1, x_2, \cdots, x_n) = \frac{x_1 + \cdots + x_n}{n}$.

Example 2. Suppose $X_1, \cdots, X_n \sim \mathcal{N}\left(\mu, \sigma^2\right)$ are i.i.d. with parameters $\theta = \left(\mu, \sigma^2\right)$.

We estimate the population mean μ with the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

We estimate $\sigma^2 = \mathrm{Var}(X_i)$ with the sample variance $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$. So $g(x_1, \cdots, x_n) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \sum_{j=1}^{n} x_j/n)^2$.

Example 1

Example 1. Suppose that $X_1, \cdots, X_n \sim \mathcal{B}(q)$ are i.i.d. with a single parameter $\theta = q$.

We estimate the population mean $q = \mathbb{E}(X_i)$ with the sample mean $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$.

Simulation 1. Fix $n = 30$ and simulate the distribution of $\bar{X}$:

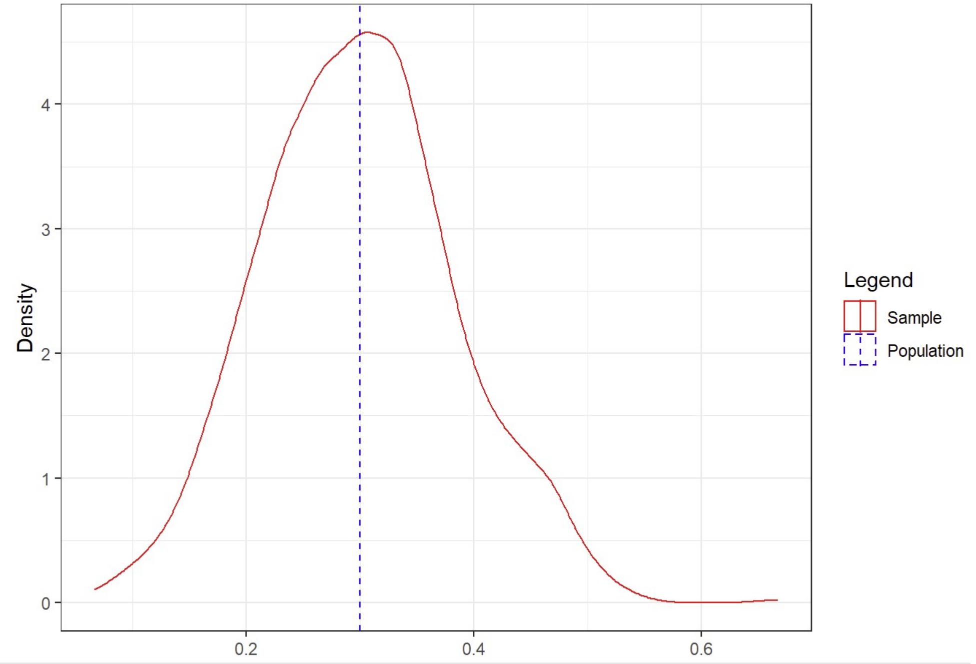
```
set.seed(0)
num_trials <- 1000
sample_size <- 30
q <- 0.3 # True parameter q

# 1. create data-frame (trial=1,2,...,num_trials)
df <- data.frame(trial=seq(num_trials))
# 2. generate samples for sequences of Bernoulli random variables
df <- mutate(df, simulation=map(.x=trial, .f=~rbinom(sample_size,1,q)))
#3. Compute the sample means
simulation_df <- mutate(df, sample_mean=map_dbl(.x=simulation, .f=mean))
```

Example 1

Simulation 1. Fix $n = 30$ and simulate the distribution of $\bar{X}$:

```
# 4.1 kernel density plot of sample means
plot_obj <- ggplot() + labs(x="Mean", y="Density") + theme_bw() +
  geom_density(data=simulation_df, aes(x=sample_mean,color="Sample", linetype="Sample"))
# 4.2 vertical line displaying population mean
plot_obj <- plot_obj + geom_vline(aes(xintercept=q,color="Population", linetype="Population") )
# 4.3 Legends
plot_obj + scale_color_manual(name = "Legend",values=c("Sample"="red", "Population"="blue")) +
  scale_linetype_manual(name="Legend", values=c("Sample"="solid", "Population"="dashed") )
```

We estimate the population mean $q = \mathbb{E}(X_i)$ with the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.



Example 1

Recall that: We estimate the population mean $q = \mathbb{E}(X_i)$ with the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Simulation 2. Compare $\bar{X}$ for different sample sizes (n from 1 to 10000).

```
# A sequence of samples with sizes varying from 1 and 10000

set.seed(0)
num_trials_per_sample_size<-10
max_sample_size<-10000
q<-0.3 # True parameter q

# 1. create a data frame containing all pairs of sample_size and trial
df <- crossing(trial=seq(num_trials_per_sample_size),
               sample_size=seq(to=sqrt(max_sample_size), by=0.1)**2)
# 2. For each pair, simulate a sequence of Bernoulli random variables
df <- mutate(df, simulation=pmap(.l=list(trial,sample_size), .f=~rbinom(.y,1,q)))
# 3. compute the sample mean of each sequence
sim_by_n_df <- mutate(df, sample_mean=map_dbl(.x=simulation, .f=mean))
```

Example 1

Recall that: We estimate the population mean $q = \mathbb{E}(X_i)$ with the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Simulation 2. Compare $\bar{X}$ for different sample sizes (n from 1 to 10000).

```r
# 4.1 scatter plot of sample means (for different sample sizes)
plot_obj <- ggplot() + labs(x="Sample size",y="Mean") + theme_bw() +
  geom_point(data=sim_by_n_df,
             aes(x=sample_size,y=sample_mean,color="Sample",linetype="Sample"),
             size=0.1)
```

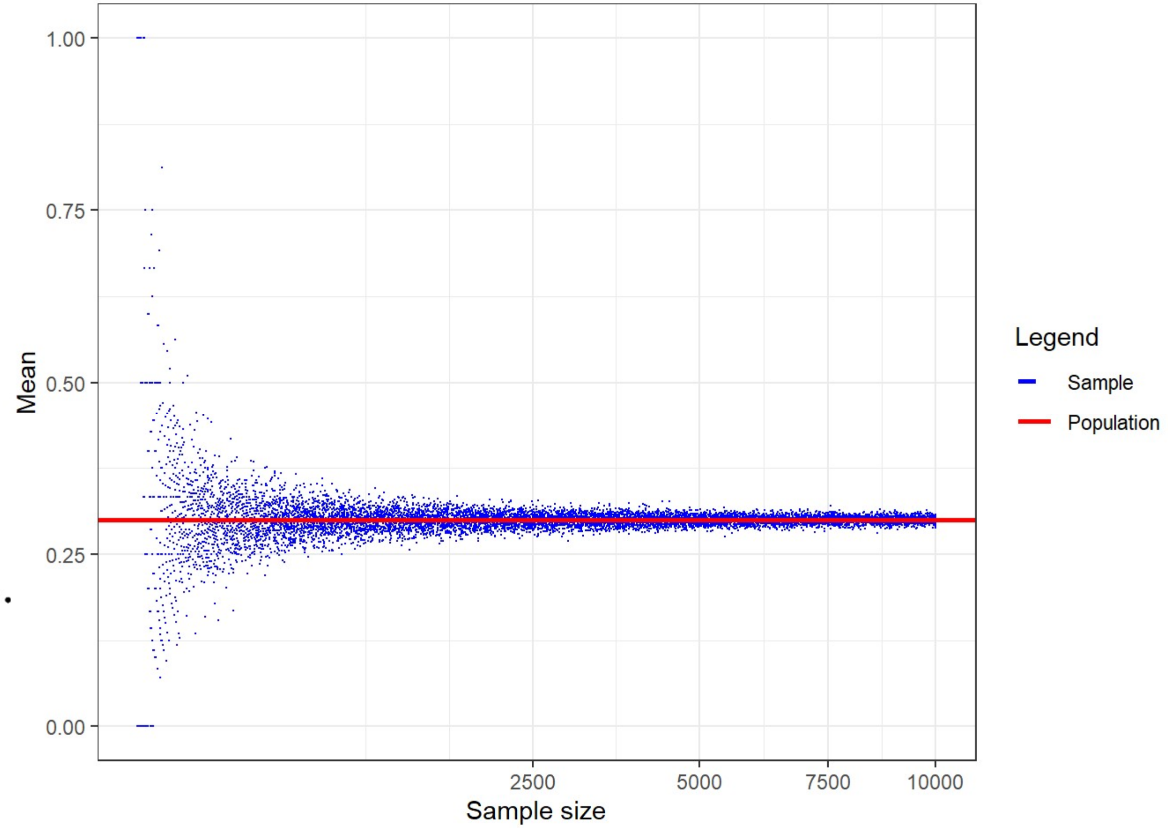
Example 1

Simulation 2. Compare $\bar{X}$ for different sample sizes (n from 1 to 10000).

```
# 4.2 horizontal line displaying population mean
plot_obj <- plot_obj +
  geom_hline(aes(yintercept=q,color="Population", linetype="Population"), size=1)
# 4.3 add Legends
plot_obj +  scale_color_manual(name = "Legend",values=c("Sample"="blue", "Population"="red")) +
  scale_linetype_manual(name="Legend",values=c("Sample"="dashed", "Population"="solid")) +
  scale_x_sqrt()
```

We estimate the population mean $q = \mathbb{E}(X_i)$ with the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Conclusion: the sample mean approaches the population mean for large n.



Example 2.

Example 2. Suppose $X_1, \cdots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ are i.i.d. with parameters $\theta = (\mu, \sigma^2)$.

We estimate $\sigma^2 = \mathrm{Var}(X_i)$ with the sample variance $s^2 := \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$.

Simulation 1. Fix $n = 30$ and simulate the distribution of $\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$:

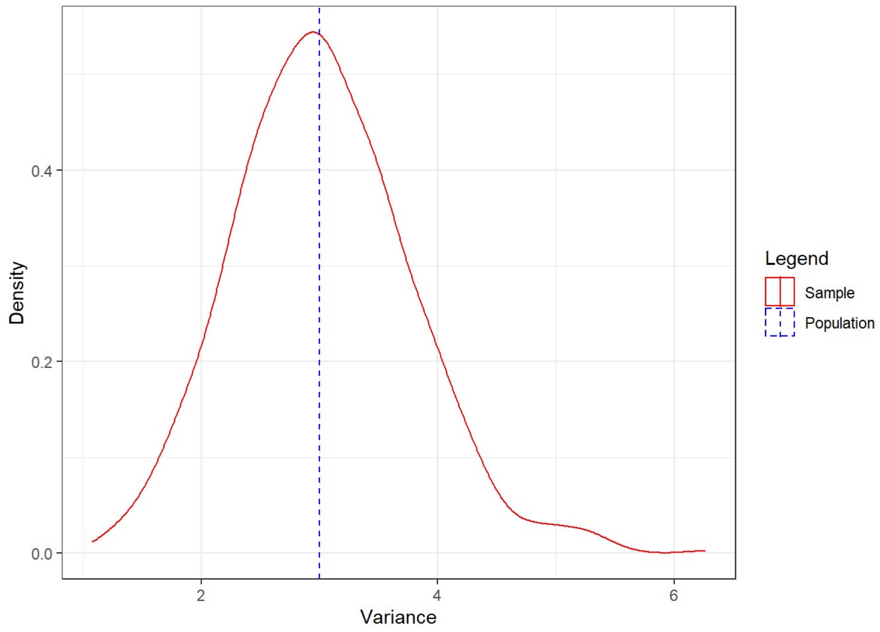
```
set.seed(0)
num_trials <- 1000
sample_size <- 30
mu <- 1  # True mu
sigma_sqr <- 3 # True sigma^2

# 1. Create a data frame (trial=1,2,...,num_trials)
df <- data.frame(trial=seq(num_trials))
# 2. generate num_trials samples with rnorm()
df <- mutate(df, simulation=map(.x=trial,
                                .f=~rnorm(sample_size,mean=mu, sd=sqrt (sigma_sqr))))
# 3. compute the sample variances (estimated sigma^2)
simulation_df <- mutate(df, sample_var=map_dbl(.x=simulation, .f=var))
```

Example 2

Simulation 1. Fix $n = 30$ and simulate the distribution of $\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$:

```r
# 4.1 kernel density plot of sample variances
plot_obj <- ggplot() + labs(x="Variance",y="Density") + theme_bw() +
  geom_density(data=simulation_df, aes(x=sample_var,color="Sample", linetype="Sample"))
# 4.2 vertical line displaying population mean
plot_obj <- plot_obj + geom_vline(aes(xintercept=sigma_sqr,
                                      color="Population",linetype="Population"))
# 4.3 add legend
plot_obj + scale_color_manual(name = "Legend",
                              values=c("Sample"="red", "Population"="blue")) +
  scale_linetype_manual(name="Legend",
                        values=c("Sample"="solid", "Population"="dashed"))
```

We estimate $\sigma^2 = \mathrm{Var}(X_i)$ with the sample variance $\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$.



Example 2

Simulation 2. Compare $\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$ for different sample sizes (n from 1 to 10000).

```
set.seed(0)
num_trials_per_sample_size <- 10
max_sample_size <- 10000
mu <- 1 # True parameter mu
sigma_sqr <- 3 # sigma^2

# 1. create data frame of all pairs of sample_size and trial
df<-crossing(trial=seq(num_trials_per_sample_size),
               sample_size=seq(to=sqrt(max_sample_size), by=0.1)**2)
# 2. For each pair, simulate a sequence of Gaussian random variables
df<-mutate(df, simulation=pmap(.l=list(trial,sample_size),
                                .f=~rnorm(.y,mean=mu,sd=sqrt(sigma_sqr))))
# 3. For each sequence, compute the sample variance
sim_by_n_df<-mutate(df, sample_var=map_dbl(.x=simulation, .f=var))
```

Example 2

Simulation 2. Compare $\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$ for different sample sizes (n from 1 to 10000).

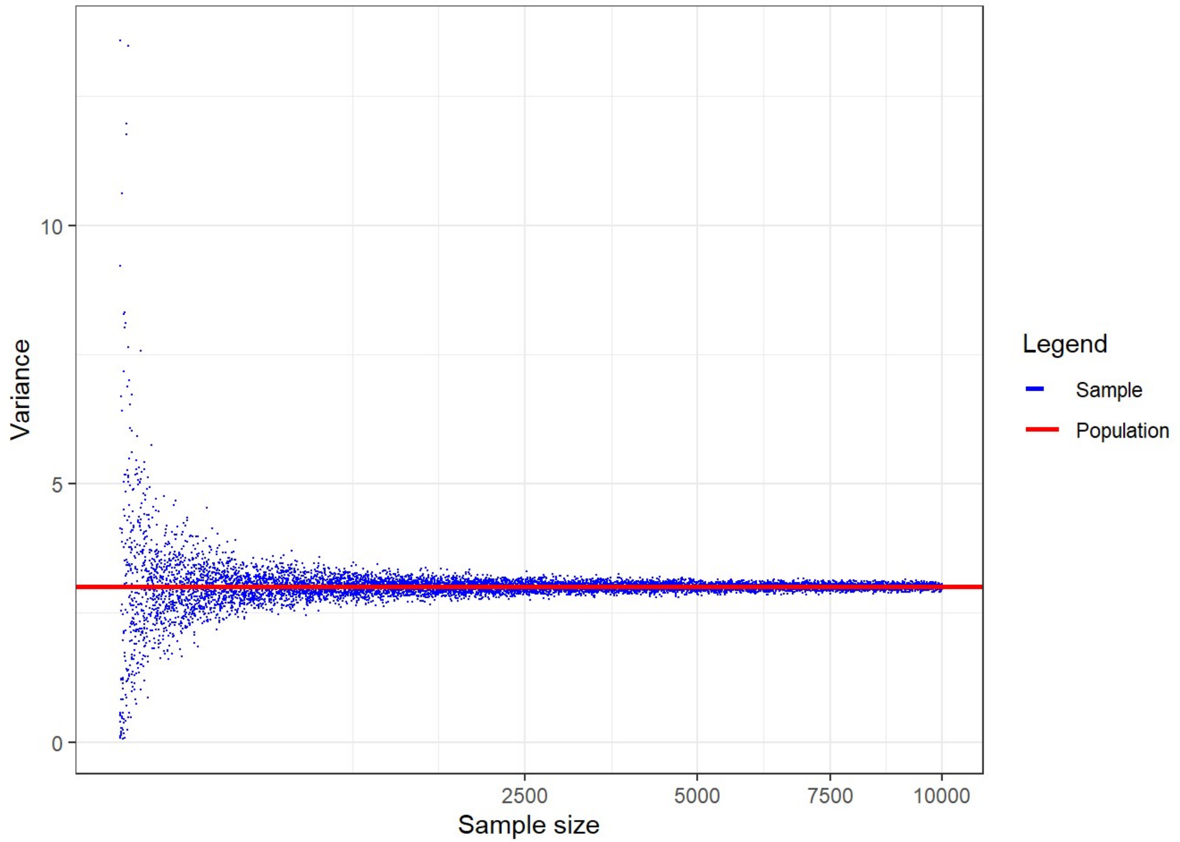
```r
# 4.1 create a scatter plot of sample variance
plot_obj <- ggplot() + labs(x="Sample size", y="Variance") + theme_bw() +
  geom_point(data=sim_by_n_df,
             aes(x=sample_size, y=sample_var,color="Sample",
                 linetype="Sample"), size=0.1)
```

Example 2

```
# 4.2 add a horizontal line displaying population variance
plot_obj <- plot_obj + geom_hline(aes(yintercept=sigma_sqr,color="Population",
                                  linetype="Population"), size=1)
# 4.3 add Legends
plot_obj+
  scale_color_manual(name = "Legend",values=c("Sample"="blue", "Population"="red") )+
  scale_linetype_manual(name="Legend", values=c("Sample"="dashed", "Population"="solid"))+
  scale_x_sqrt()
```

We estimate $\sigma^2 = \mathrm{Var}(X_i)$ with the sample variance $\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$.

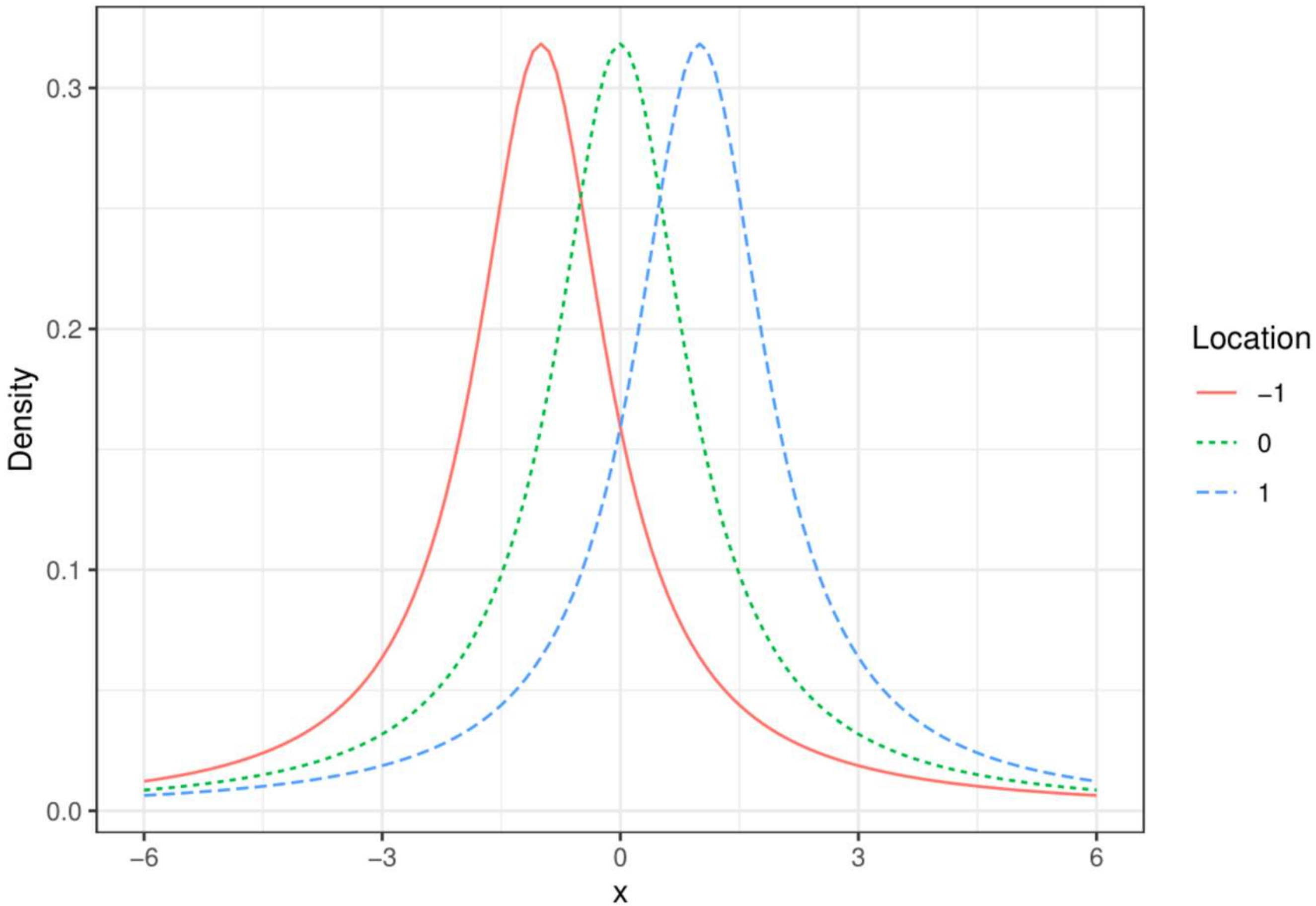
Conclusion: the sample variance approaches the population variance for large n.



The Cauchy distribution

A random variable X has a Cauchy distribution with location parameter θ if its density is

$$f_\theta(x) := \frac{1}{\pi \cdot (1 + (x-\theta)^2)}.$$



Example 3

Example 3. Suppose $X_1, \cdots, X_n$ are i.i.d. Cauchy random variables with parameters θ.

Attempt: We estimate the population mean with the sample mean $\frac{1}{n}\sum_{i=1}^{n} X_i$.

Simulation 1. Compare the sample mean $\frac{1}{n}\sum_{i=1}^{n} X_i$ for different sample sizes (n from 1 to 10000).

```
set.seed(0)
num_trials_per_sample_size <- 10
max_sample_size <- 10000
theta <- 1 # True parameter theta

# 1. create data frame consisting of all pairs of sample_size and trial
df <- crossing(trial=seq(num_trials_per_sample_size),
               sample_size=seq(to=sqrt(max_sample_size) , by=0.1)**2)
# 2. for each pair, simulate a sequence of Cauchy random variables
df <- mutate(df, simulation=pmap(.l=list(trial,sample_size),
                                 .f=~rcauchy(.y, location=theta) ))
# 3. for each sequence, compute its sample mean
sim_by_n_df <- mutate(df, sample_mean=map_dbl(.x=simulation, .f=mean))
```

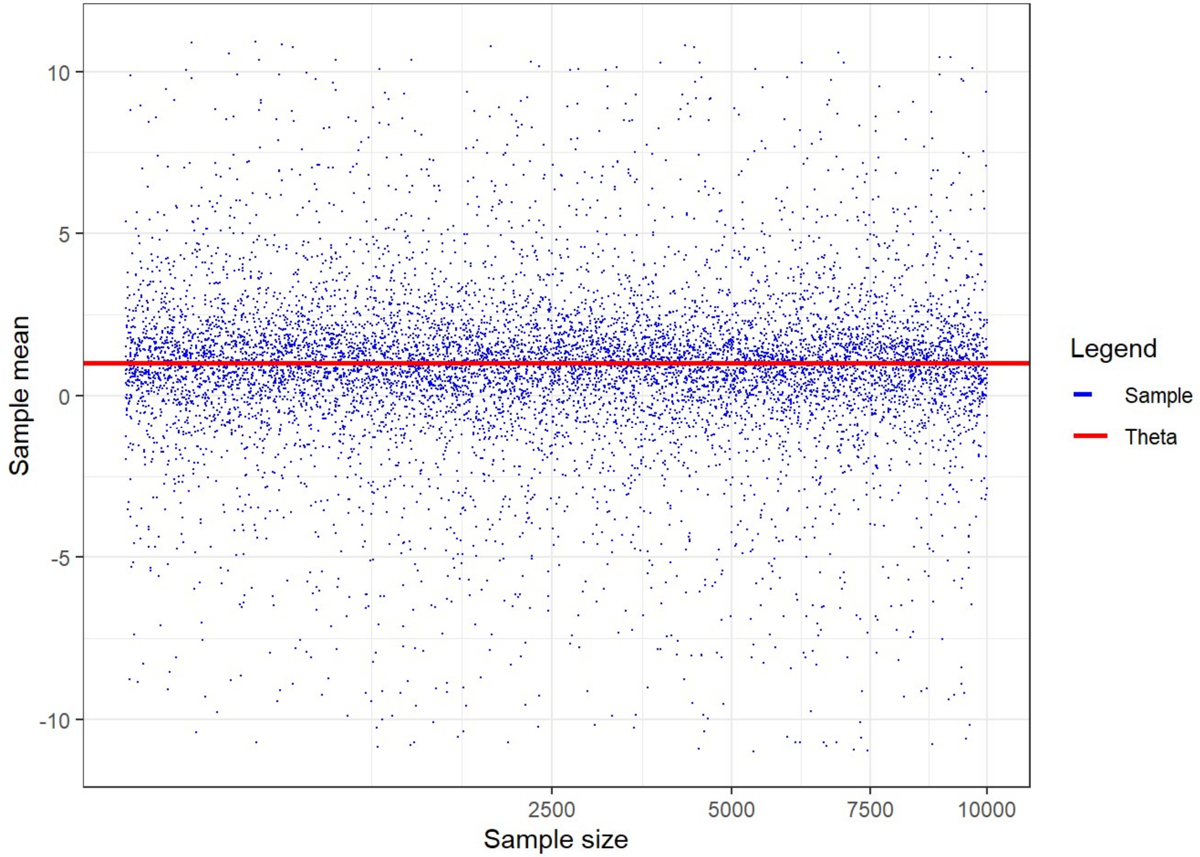
Example 3

Example 3. Suppose $X_1, \cdots, X_n$ are i.i.d. Cauchy random variables with parameters θ.

Attempt: We estimate the population mean with the sample mean $\frac{1}{n}\sum_{i=1}^{n} X_i$.

Simulation 1. Compare the sample mean $\frac{1}{n}\sum_{i=1}^{n} X_i$ for different sample sizes (n from 1 to 10000).

Conclusion: the sample mean does not converge!



The Cauchy distribution

The sample mean of a sequence of i.i.d. Cauchy random variables is a bad estimator as it does not converge when the sample size is arbitrarily large.

Recall the **weak law of large numbers**: let $X : \Omega \to \mathbb{R}$ be a random variable with a *well-defined expectation* $\mu = \mathbb{E}(X)$. Let $X_1, \cdots, X_n : \Omega \to \mathbb{R}$ be a sequence of independent copies of X. Then for all $\epsilon > 0$, $\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum X_i - \mu\right| \geq \epsilon\right) = 0$.

Since the sample mean of a sequence of i.i.d. Cauchy random variables does not converge, the weak law of large numbers implies that the Cauchy distribution does not have a well-defined expectation.

没有明确期望值

What are reasonable estimators in the setting of Cauchy random variables?

The Cauchy distribution

Recall that a random variable X has a Cauchy distribution with location parameter θ if its density is $f_\theta(x) := \frac{1}{\pi \cdot \left(1+(x-\theta)^2\right)}$.

A Cauchy random variable X has a cumulative distribution function

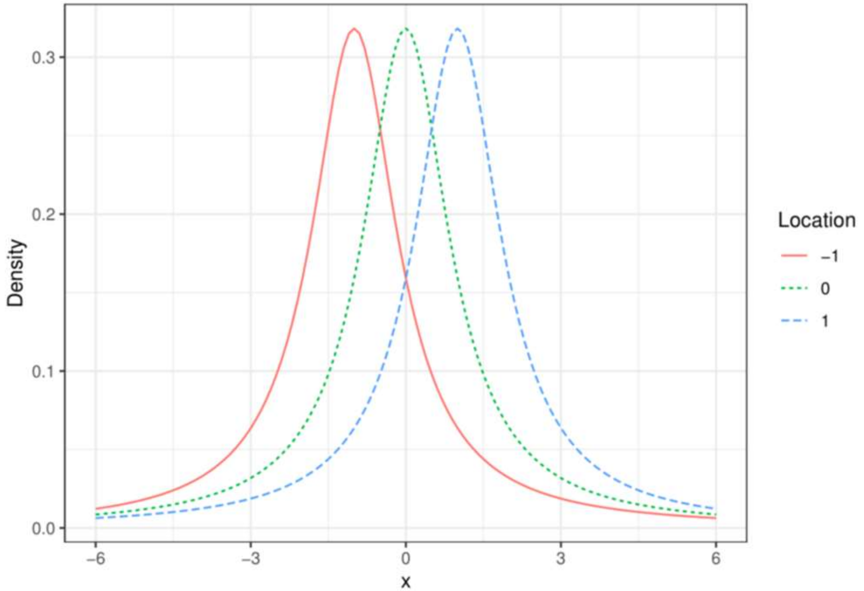
这个函数给出了随机变量X的值小于或等于x的概率

$$F_\theta(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_\theta(z)dz = \frac{1}{\pi}\arctan(x-\theta) + \frac{1}{2}.$$

The population median of X is $F_\theta^{-1}(0.5) = \inf\left\{x \in \mathbb{R} : F_\theta(x) \geq \frac{1}{2}\right\} = \theta.$

柯西分布的中位数可以通过其累积分布函数的逆来定义

Suppose that we have $X_1, \cdots, X_n$ are i.i.d. Cauchy random variables. We can try the sample median $\hat{\theta} = \text{Median}(X_1, \cdots, X_n)$ as an estimator for θ.



Example 3

Example 3. Suppose $X_1, \cdots, X_n$ are i.i.d. Cauchy random variables with parameters θ.

Attempt: We estimate θ with the sample median $\text{Median}(X_1, \cdots, X_n)$.

Simulation 2. Compare $\text{Median}(X_1, \cdots, X_n)$ for different sample sizes (n from 1 to 10000).

```
set.seed(0)
num_trials_per_sample_size <- 10
max_sample_size <- 10000
theta <- 1 # True parameter theta

# 1. create data frame consisting of all pairs of sample_size and trial
df <- crossing(trial=seq(num_trials_per_sample_size),
               sample_size=seq(to=sqrt(max_sample_size) , by=0.1)**2)
# 2. for each pair, simulate a sequence of Cauchy random variables
df <- mutate(df, simulation=pmap(.l=list(trial,sample_size),
                                 .f=~rcauchy(.y, location=theta) ))
# 3. for each sequence, compute its sample median
sim_by_n_df <- mutate(df, sample_median=map_dbl(.x=simulation, .f=median))
```

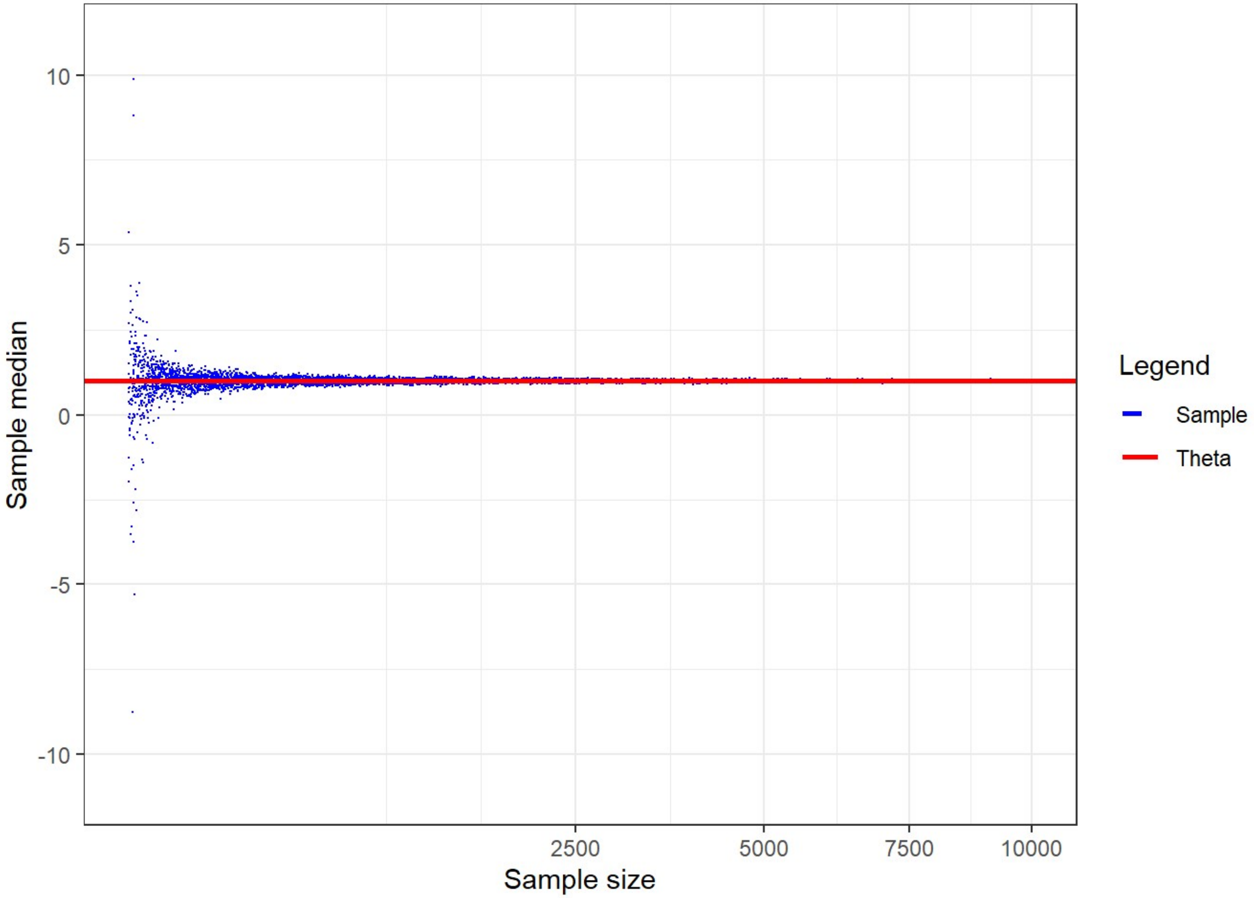
Example 3

Example 3. Suppose $X_1, \cdots, X_n$ are i.i.d. Cauchy random variables with parameters θ.

Attempt: We estimate θ with the sample median Median$(X_1, \cdots, X_n)$.

Simulation 2. Compare Median$(X_1, \cdots, X_n)$ for different sample sizes (n from 1 to 10000).

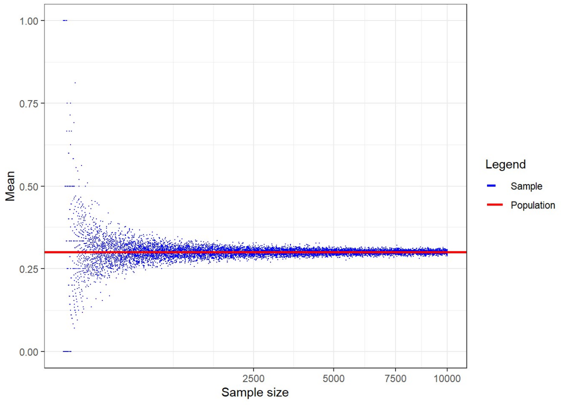
Conclusion: the sample median approaches to θ for large n.



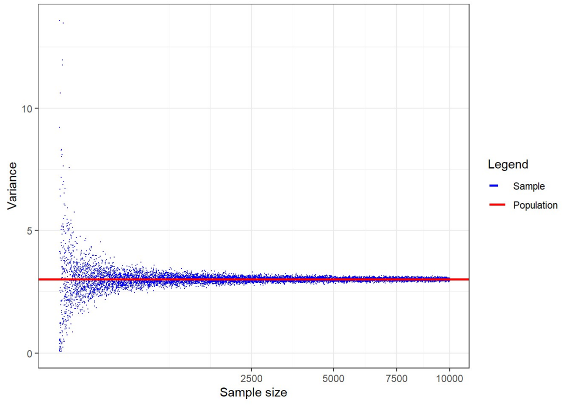
3. Consistent estimators

We are interested in statistical estimators $\hat{\theta}$ that tend towards the true parameter θ as $n \to \infty$.

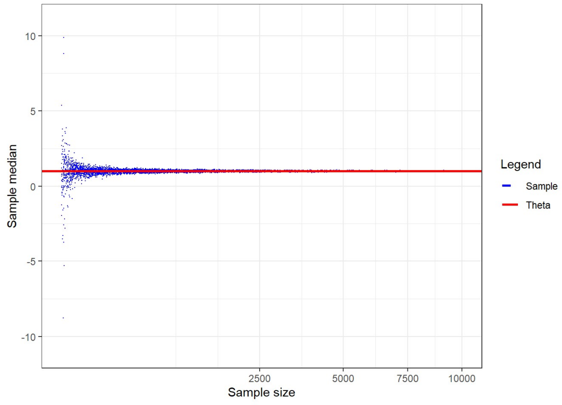
Example 1



Example 2



Example 3



These estimators are referred to as "consistent" estimators

Consistent estimators

Consistent estimators

A sample statistic $\hat{\theta} = g(X_1, \cdots, X_n)$ of a population parameter θ is consistent if for all $\epsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}\{|g(X_1, \cdots, X_n) - \theta| > \epsilon\} = 0.$$

So a consistent estimator converges to the population parameter in probability, as n goes to infinity.

Consequence of the weak law of large numbers

Recall the **weak law of large numbers**: let $X : \Omega \to \mathbb{R}$ be a random variable with a well-defined expectation $\mu = \mathbb{E}(X)$. Let $X_1, \cdots, X_n : \Omega \to \mathbb{R}$ be a sequence of independent copies of X. Then for all $\epsilon > 0$, $\lim_{n\to\infty} \mathbb{P}\big(\big|\frac{1}{n}\sum X_i - \mu\big| \geq \epsilon\big) = 0$.

Suppose that $X_1, \cdots, X_n$ are i.i.d. random variables with mean μ and variance σ^2, as a consequence of the weak law of large numbers, we have

1. The sample mean $\bar{X} := \frac{1}{n}\sum_{i=1}^n X_i$ is a consistent estimator of μ (e.g., Example 1)

2. The sample variance $\frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ is a consistent estimator of σ^2 (e.g., Example 2).

Moreover, since the standard deviation σ is the square root of the variance, we have

3. The sample deviation $\sqrt{\frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2}$ is a consistent estimator of σ.

4. Statistical bias

The bias of an estimator

The bias of an estimator $\hat{\theta} = g(X_1, \cdots, X_n)$ of a population parameter θ is

$$\mathrm{Bias}\left(\hat{\theta}\right) := \mathbb{E}\left(\hat{\theta}\right) - \theta.$$

The estimator is said to be unbiased if $\mathrm{Bias}\left(\hat{\theta}\right) = 0$.

Given an independent and identically distributed sample $X_1, \cdots, X_n$, we have

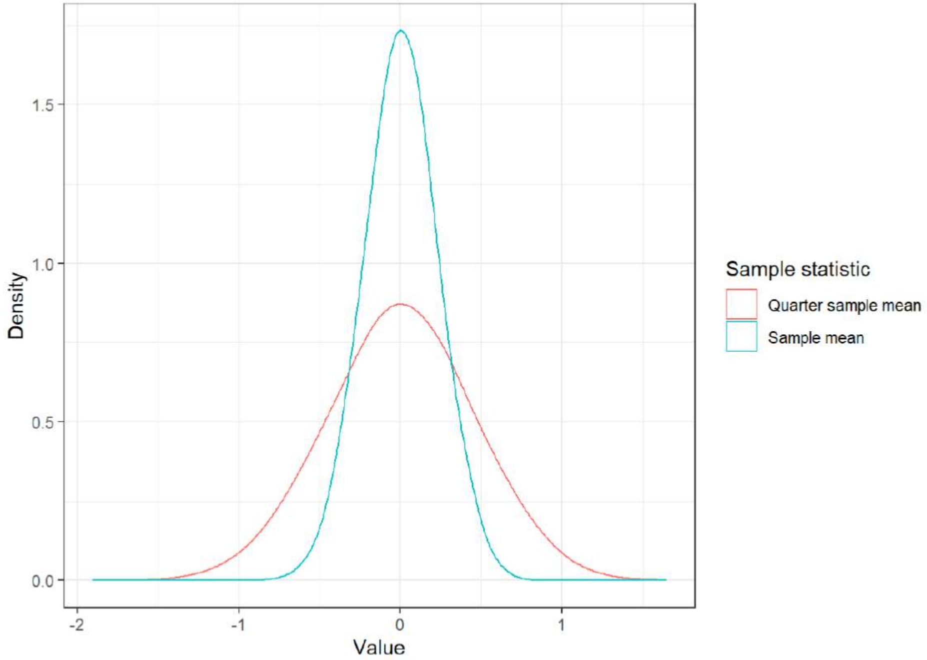
$$\mathrm{Bias}(\bar{X}) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \mu = 0$$

$$\mathrm{Bias}(s^2) = \mathbb{E}\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right) - \sigma^2 = 0.$$

Statistical bias

The sample mean is an unbiased estimator for the population mean.

The sample mean computed from just the first quarter of the data is also unbiased.



尽管一个估计量是无偏的，它的有效性（如何紧密集中于真值周围）可能会因为样本大小或选取的样本而变化

There are many unbiased estimators.

However, the bias is not sufficient for measuring the accuracy of an estimator. In fact, many such estimators are of very high variance!

5. Variance of an estimator

Variance of an estimator

The variance of an estimator $\hat{\theta} = g(X_1, \cdots, X_n)$ of a population parameter θ is

$$\mathrm{Var}\left(\hat{\theta}\right) := \mathbb{E}\left\{\left(\hat{\theta} - \mathbb{E}\left(\hat{\theta}\right)\right)^2\right\}.$$

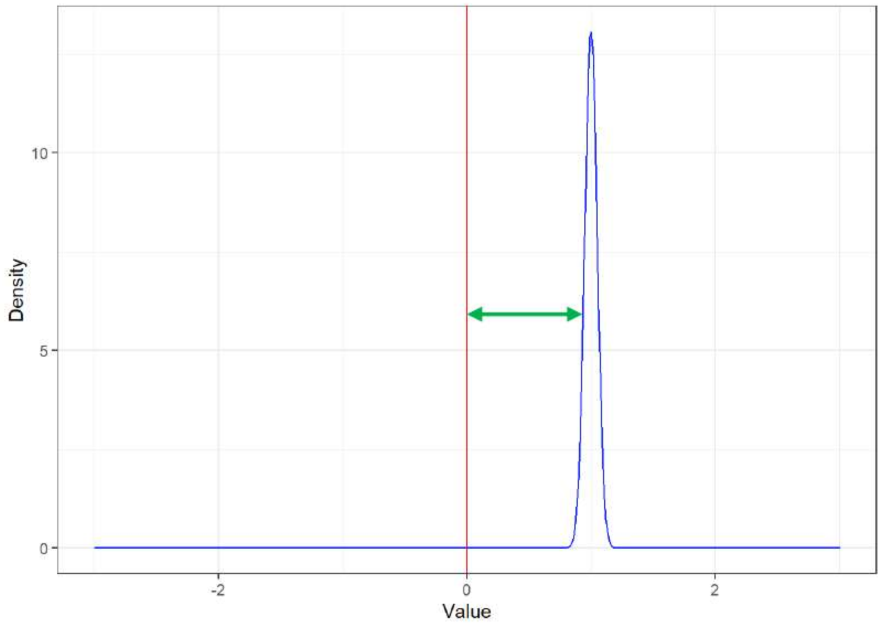
An estimator with a small variance is not necessarily a good estimator either.

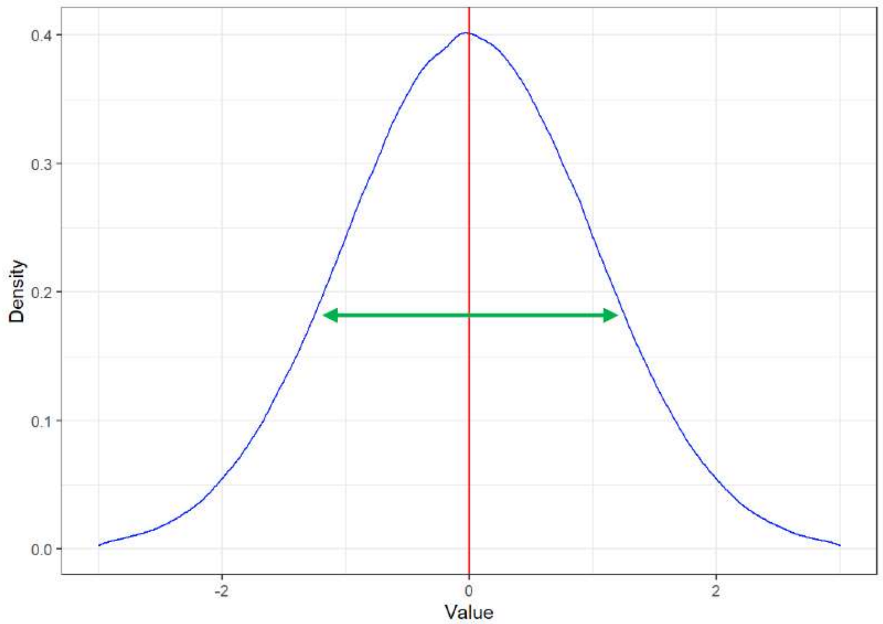
Consider an extreme case where $\hat{\theta} = 0$ for any sample. So $\hat{\theta}$ does not reflect useful information about the population.

Bias and variance of an estimator

$$\text{Bias}(\hat{\theta}) := \mathbb{E}(\hat{\theta}) - \theta.$$

$$\text{Var}(\hat{\theta}) := \mathbb{E}\left[\left\{\hat{\theta} - \mathbb{E}(\hat{\theta})\right\}^2\right]$$





6. Mean squared error

The mean square error of an estimator

The mean square error of an estimator $\hat{\theta} = g(X_1, \cdots, X_n)$ of a population parameter θ is

$$\mathrm{MSE}\left(\hat{\theta}\right) := \mathbb{E}\left\{\left(\hat{\theta} - \theta\right)^2\right\}.$$

The MSE tells the difference between the estimate $\hat{\theta}$ and the parameter θ in the sense of mean square distance.

The Bias-variance decomposition

Theorem (Bias-variance decomposition)

Suppose that $\hat{\theta}$ is an estimator of a parameter θ. Then

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

Proof.

$$\begin{aligned}
\mathrm{MSE}\left(\hat{\theta}\right) &:= \mathbb{E}\left(\left(\hat{\theta} - \theta\right)^2\right) \\
&= \mathbb{E}\left\{\left[\left(\hat{\theta} - \mathbb{E}(\hat{\theta})\right) + \left(\mathbb{E}(\hat{\theta}) - \theta\right)\right]^2\right\} \\
&= \mathbb{E}\left\{\left[\hat{\theta} - \mathbb{E}(\hat{\theta})\right]^2\right\} + \mathbb{E}\left\{\left[\mathbb{E}(\hat{\theta}) - \theta\right]^2\right\} + 2\mathbb{E}\left\{\left[\left(\hat{\theta} - \mathbb{E}(\hat{\theta})\right) \cdot \left(\mathbb{E}(\hat{\theta}) - \theta\right)\right]\right\} \\
&= \mathbb{E}\left\{\left[\hat{\theta} - \mathbb{E}(\hat{\theta})\right]^2\right\} + \left(\mathbb{E}(\hat{\theta}) - \theta\right)^2 \\
&= \mathrm{Var}(\hat{\theta}) + \mathrm{Bias}(\hat{\theta})^2.
\end{aligned}$$

7. Minimum variance unbiased estimator

Recall that an estimator is said to be unbiased if $\text{Bias}\left(\hat{\theta}\right) = 0$.

Minimum variance unbiased estimator

An estimator $\hat{\theta}$ is said to be a minimum variance unbiased estimator (MVUE) if

1. $\hat{\theta}$ is unbiased, i.e., $\mathbb{E}(\hat{\theta}) = \theta$.
2. $\hat{\theta}$ has minimum variance, i.e., $\text{Var}(\hat{\theta}) \leq \text{Var}(\tilde{\theta})$ for any unbiased estimator $\tilde{\theta}$.

Remark: A MVUE also has minimal mean square error over all unbiased estimators, because of the bias-variance decomposition

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

which holds for all unbiased estimators.

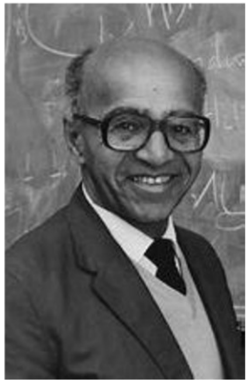
Minimum variance unbiased estimator

Recall that a **minimum variance unbiased estimator** (MVUE) has minimum variance over all unbiased estimators.

Example 1. Suppose that $X_1, \cdots, X_n \sim \mathcal{B}(q)$ are i.i.d.

Then $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is a MVUE of q.

Note: This is a consequence of the Rao-Blackwell theorem.



Minimum variance unbiased estimator

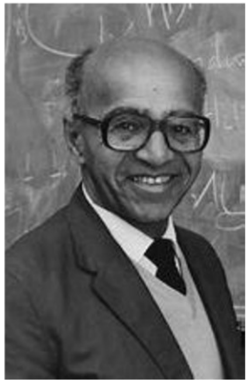
Recall that a **minimum variance unbiased estimator** (MVUE) has minimum variance over all unbiased estimators.

Example 2. Suppose $X_1, \cdots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ are i.i.d.

Then $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is a MVUE of $\mu = \mathbb{E}(X_i)$.

$s^2 = \frac{1}{n-1}\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$ is a MVUE of $\sigma^2 = \mathbb{E}\big((X_i - \mu)^2\big)$.

Note: This is also a consequence of the Rao-Blackwell theorem.



What have we covered?

We considered sample statistics as estimators of parameters of interest.

We introduced the concept of statistical consistency.

We also considered the idea of statistical bias and the bias-variance decomposition.

We have discussed the concept of a minimum variance unbiased estimator.

Thanks for listening!

Dr. Rihuan Ke

rihuan.ke@bristol.ac.uk

Statistical Computing and Empirical Methods
Unit EMATM0061, MSc Data Science