YSU ASDS, Statistics, Fall 2019

Lecture 14

Michael Poghosyan

05 Oct 2019

Contents

Last Lecture ReCap

- State the LLN and CLT

CLT, Berry-Eseen Inequality

Now, quickly about the convergence rate of CLT:

Theorem(18+, Berry-Esseen): Assume X_k are IID r.v.s with finite $\mathbb{E}(X_1) = \mu$, $Var(X_1) = \sigma^2$ and $\mathbb{E}(|X_1|^3)$. Then, for any $n \in \mathbb{N}$,

$$\sup_{x \in \mathbb{R}} |\mathbb{P}(Z_n \leq x) - \Phi(x)| \leq \frac{\mathbb{E}(|X_1 - \mu|^3)}{\sigma^3 \cdot \sqrt{n}},$$

where

$$Z_n = Standardize(S_n) = Standardize(\overline{X}_n),$$

and $\Phi(x)$ is the CDF of $\mathcal{N}(0, 1)$.

CLT, Roughly

In a non-rigorous way, we can write, for large n (here $\approx$ means approximately distributed as):

$$\frac{S_n - n \cdot \mu}{\sqrt{n} \cdot \sigma} \approx \mathcal{N}(0, 1) \qquad \textit{and} \qquad \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \approx \mathcal{N}(0, 1).$$

CLT, Roughly

In a non-rigorous way, we can write, for large n (here $\approx$ means approximately distributed as):

$$\frac{S_n - n \cdot \mu}{\sqrt{n} \cdot \sigma} \approx \mathcal{N}(0,1) \qquad \textit{and} \qquad \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \approx \mathcal{N}(0,1).$$

or

$$S_n \approx \mathcal{N}(n\mu, n\sigma^2) \qquad \textit{and} \qquad \overline{X}_n \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

CLT, Roughly -2

Let us summarize:

CLT, Roughly -2

Let us summarize:

- If X_k-s are independent, have the Mean $\mathbb{E}(X_k) = \mu$ and $Var(X_K) = \sigma^2$, and **are Normally Distributed**, i.e., $X_k \sim \mathcal{N}(\mu, \sigma^2)$, then

CLT, Roughly -2

Let us summarize:

- If X_k-s are independent, have the Mean $\mathbb{E}(X_k) = \mu$ and $Var(X_K) = \sigma^2$, and **are Normally Distributed**, i.e., $X_k \sim \mathcal{N}(\mu, \sigma^2)$, then

$$S_n \sim \mathcal{N}(n\mu, n\sigma^2) \qquad \textit{and} \qquad \overline{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right);$$

  so we know the **exact Distributions** of S_n and $\overline{X}_n$.

CLT, Roughly -2

Let us summarize:

- If X_k-s are independent, have the Mean $\mathbb{E}(X_k) = \mu$ and $Var(X_K) = \sigma^2$, and **are Normally Distributed**, i.e., $X_k \sim \mathcal{N}(\mu, \sigma^2)$, then

  $$S_n \sim \mathcal{N}(n\mu, n\sigma^2) \qquad \textit{and} \qquad \overline{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right);$$

  so we know the **exact Distributions** of S_n and $\overline{X}_n$.
- If X_k-s are independent, have the Mean $\mathbb{E}(X_k) = \mu$ and $Var(X_K) = \sigma^2$, and **from any Distribution**, then

CLT, Roughly -2

Let us summarize:

- If X_k-s are independent, have the Mean $\mathbb{E}(X_k) = \mu$ and $Var(X_K) = \sigma^2$, and **are Normally Distributed**, i.e., $X_k \sim \mathcal{N}(\mu, \sigma^2)$, then

  $$S_n \sim \mathcal{N}(n\mu, n\sigma^2) \qquad and \qquad \overline{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right);$$

  so we know the **exact Distributions** of S_n and $\overline{X}_n$.

- If X_k-s are independent, have the Mean $\mathbb{E}(X_k) = \mu$ and $Var(X_K) = \sigma^2$, and **from any Distribution**, then

  $$S_n \approx \mathcal{N}(n\mu, n\sigma^2) \qquad and \qquad \overline{X}_n \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right);$$

  and we know the **assymptotic Distributions** (approximate Distributions for large n) of S_n and $\overline{X}_n$.

CLT, Again

Let us "encode" the notion of the convergence in Distributions:

CLT, Again

Let us "encode" the notion of the convergence in Distributions: by CLT, if n is large, then

$$\frac{S_n - n \cdot \mu}{\sqrt{n} \cdot \sigma} \xrightarrow{D} \mathcal{N}(0,1) \qquad \textit{and} \qquad \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} \mathcal{N}(0,1).$$

CLT, Again

Let us "encode" the notion of the convergence in Distributions: by CLT, if n is large, then

$$\frac{S_n - n \cdot \mu}{\sqrt{n} \cdot \sigma} \xrightarrow{D} \mathcal{N}(0,1) \qquad \textit{and} \qquad \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} \mathcal{N}(0,1).$$

This means that, in particular, for large n and any real numbers $a < b$,

$$\mathbb{P}\left(a \leq \frac{S_n - n \cdot \mu}{\sqrt{n} \cdot \sigma} \leq b\right) \overset{Z \sim \mathcal{N}(0,1)}{\approx} \mathbb{P}(a \leq Z \leq b) = \Phi(b) - \Phi(a);$$

and, similarly,

$$\mathbb{P}\left(a \leq \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq b\right) \overset{Z \sim \mathcal{N}(0,1)}{\approx} \mathbb{P}(a \leq Z \leq b) = \Phi(b) - \Phi(a).$$

CLT, Again

Let us "encode" the notion of the convergence in Distributions: by CLT, if n is large, then

$$\frac{S_n - n\cdot\mu}{\sqrt{n}\cdot\sigma} \xrightarrow{D} \mathcal{N}(0,1) \qquad \textit{and} \qquad \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} \mathcal{N}(0,1).$$

This means that, in particular, for large n and any real numbers $a < b$,

$$\mathbb{P}\left(a \le \frac{S_n - n\cdot\mu}{\sqrt{n}\cdot\sigma} \le b\right) \overset{Z\sim\mathcal{N}(0,1)}{\approx} \mathbb{P}(a \le Z \le b) = \Phi(b) - \Phi(a);$$

and, similarly,

$$\mathbb{P}\left(a \le \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \le b\right) \overset{Z\sim\mathcal{N}(0,1)}{\approx} \mathbb{P}(a \le Z \le b) = \Phi(b) - \Phi(a).$$

Here $\Phi(x)$ is the CDF of Standard Normal Distribution, $\mathcal{N}(0,1)$.

CLT, Again

For example, consider the last approximation:

$$\mathbb{P}\left(a \le \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \le b\right) \overset{Z \sim \mathcal{N}(0,1)}{\approx} \mathbb{P}(a \le Z \le b) = \Phi(b) - \Phi(a).$$

CLT, Again

For example, consider the last approximation:

$$\mathbb{P}\left(a \leq \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq b\right) \overset{Z\sim\mathcal{N}(0,1)}{\approx} \mathbb{P}(a \leq Z \leq b) = \Phi(b) - \Phi(a).$$

This can be written as

$$\mathbb{P}\left(\mu + a \cdot \frac{\sigma}{\sqrt{n}} \leq \overline{X}_n \leq \mu + b \cdot \frac{\sigma}{\sqrt{n}}\right) \approx \mathbb{P}(a \leq Z \leq b) = \Phi(b) - \Phi(a).$$

CLT, Again

For example, consider the last approximation:

$$\mathbb{P}\left(a \leq \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq b\right) \overset{Z\sim\mathcal{N}(0,1)}{\approx} \mathbb{P}(a \leq Z \leq b) = \Phi(b) - \Phi(a).$$

This can be written as

$$\mathbb{P}\left(\mu + a\cdot\frac{\sigma}{\sqrt{n}} \leq \overline{X}_n \leq \mu + b\cdot\frac{\sigma}{\sqrt{n}}\right) \approx \mathbb{P}(a \leq Z \leq b) = \Phi(b) - \Phi(a).$$

Now, let us take as a the quantile of the Standard Normal $z_{\alpha/2}$, $a = z_{\alpha/2}$ and $b = -a = z_{1-\alpha/2}$.

CLT, Again

For example, consider the last approximation:

$$\mathbb{P}\left(a \leq \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq b\right) \overset{Z\sim\mathcal{N}(0,1)}{\approx} \mathbb{P}(a \leq Z \leq b) = \Phi(b) - \Phi(a).$$

This can be written as

$$\mathbb{P}\left(\mu + a \cdot \frac{\sigma}{\sqrt{n}} \leq \overline{X}_n \leq \mu + b \cdot \frac{\sigma}{\sqrt{n}}\right) \approx \mathbb{P}(a \leq Z \leq b) = \Phi(b) - \Phi(a).$$

Now, let us take as a the quantile of the Standard Normal $z_{\alpha/2}$, $a = z_{\alpha/2}$ and $b = -a = z_{1-\alpha/2}$. Then,

$$\mathbb{P}(z_{\alpha/2} \leq Z \leq z_{1-\alpha/2}) =$$

CLT, Again

For example, consider the last approximation:

$$\mathbb{P}\left(a \leq \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq b\right) \overset{Z\sim\mathcal{N}(0,1)}{\approx} \mathbb{P}(a \leq Z \leq b) = \Phi(b) - \Phi(a).$$

This can be written as

$$\mathbb{P}\left(\mu + a \cdot \frac{\sigma}{\sqrt{n}} \leq \overline{X}_n \leq \mu + b \cdot \frac{\sigma}{\sqrt{n}}\right) \approx \mathbb{P}(a \leq Z \leq b) = \Phi(b) - \Phi(a).$$

Now, let us take as a the quantile of the Standard Normal $z_{\alpha/2}$, $a = z_{\alpha/2}$ and $b = -a = z_{1-\alpha/2}$. Then,

$$\mathbb{P}(z_{\alpha/2} \leq Z \leq z_{1-\alpha/2}) = 1 - \alpha.$$

CLT, Again

For example, consider the last approximation:

$$\mathbb{P}\left(a \leq \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq b\right) \overset{Z\sim\mathcal{N}(0,1)}{\approx} \mathbb{P}(a \leq Z \leq b) = \Phi(b) - \Phi(a).$$

This can be written as

$$\mathbb{P}\left(\mu + a \cdot \frac{\sigma}{\sqrt{n}} \leq \overline{X}_n \leq \mu + b \cdot \frac{\sigma}{\sqrt{n}}\right) \approx \mathbb{P}(a \leq Z \leq b) = \Phi(b) - \Phi(a).$$

Now, let us take as a the quantile of the Standard Normal $z_{\alpha/2}$, $a = z_{\alpha/2}$ and $b = -a = z_{1-\alpha/2}$. Then,

$$\mathbb{P}(z_{\alpha/2} \leq Z \leq z_{1-\alpha/2}) = 1 - \alpha.$$

Hence, for large n,

$$\mathbb{P}\left(\mu + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \overline{X}_n \leq \mu + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) \approx 1 - \alpha.$$

CLT, Again

For example, consider the last approximation:

$$\mathbb{P}\left(a \leq \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq b\right) \overset{Z \sim \mathcal{N}(0,1)}{\approx} \mathbb{P}(a \leq Z \leq b) = \Phi(b) - \Phi(a).$$

This can be written as

$$\mathbb{P}\left(\mu + a \cdot \frac{\sigma}{\sqrt{n}} \leq \overline{X}_n \leq \mu + b \cdot \frac{\sigma}{\sqrt{n}}\right) \approx \mathbb{P}(a \leq Z \leq b) = \Phi(b) - \Phi(a).$$

Now, let us take as a the quantile of the Standard Normal $z_{\alpha/2}$, $a = z_{\alpha/2}$ and $b = -a = z_{1-\alpha/2}$. Then,

$$\mathbb{P}(z_{\alpha/2} \leq Z \leq z_{1-\alpha/2}) = 1 - \alpha.$$

Hence, for large n,

$$\mathbb{P}\left(\mu + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \overline{X}_n \leq \mu + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) \approx 1 - \alpha.$$

And this is **for any X_n (IID), with any distribution**.

CLT, Again

For example, consider the last approximation:

$$\mathbb{P}\left(a \leq \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq b\right) \overset{Z \sim \mathcal{N}(0,1)}{\approx} \mathbb{P}(a \leq Z \leq b) = \Phi(b) - \Phi(a).$$

This can be written as

$$\mathbb{P}\left(\mu + a \cdot \frac{\sigma}{\sqrt{n}} \leq \overline{X}_n \leq \mu + b \cdot \frac{\sigma}{\sqrt{n}}\right) \approx \mathbb{P}(a \leq Z \leq b) = \Phi(b) - \Phi(a).$$

Now, let us take as a the quantile of the Standard Normal $z_{\alpha/2}$, $a = z_{\alpha/2}$ and $b = -a = z_{1-\alpha/2}$. Then,

$$\mathbb{P}(z_{\alpha/2} \leq Z \leq z_{1-\alpha/2}) = 1 - \alpha.$$

Hence, for large n,

$$\mathbb{P}\left(\mu + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \overline{X}_n \leq \mu + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) \approx 1 - \alpha.$$

And this is **for any X_n (IID), with any distribution**. This will be our tool to constract Confidence Intervals and design Hypotheses Tests.

Inferential Statistics

Parametric Inference: Point Estimation

Parametric Statistics: General Problem

One of the general Problems of Statistics is the following: we have a Sample, a Dataset $x: \; x_1, ..., x_n$, and our aim is to get an insight from these numbers, to get an information about the Population, about the *process* generating that Dataset.

Parametric Statistics: General Problem

One of the general Problems of Statistics is the following: we have a Sample, a Dataset $x: \ x_1, ..., x_n$, and our aim is to get an insight from these numbers, to get an information about the Population, about the *process* generating that Dataset.

We can, of course, calculate the Sample Mean and the Sample Variance of our Dataset.

Parametric Statistics: General Problem

One of the general Problems of Statistics is the following: we have a Sample, a Dataset $x: \ x_1, ..., x_n$, and our aim is to get an insight from these numbers, to get an information about the Population, about the *process* generating that Dataset.

We can, of course, calculate the Sample Mean and the Sample Variance of our Dataset. Or, we can plot the Histogram or KDE.

Parametric Statistics: General Problem

One of the general Problems of Statistics is the following: we have a Sample, a Dataset $x: \ x_1, ..., x_n$, and our aim is to get an insight from these numbers, to get an information about the Population, about the *process* generating that Dataset.

We can, of course, calculate the Sample Mean and the Sample Variance of our Dataset. Or, we can plot the Histogram or KDE. But will this give an info about the Population or the process generating the Dataset?

Parametric Statistics: General Problem

One of the general Problems of Statistics is the following: we have a Sample, a Dataset $x: \ x_1, ..., x_n$, and our aim is to get an insight from these numbers, to get an information about the Population, about the *process* generating that Dataset.

We can, of course, calculate the Sample Mean and the Sample Variance of our Dataset. Or, we can plot the Histogram or KDE. But will this give an info about the Population or the process generating the Dataset? Well, no, in general.

Parametric Statistics: General Problem

One of the general Problems of Statistics is the following: we have a Sample, a Dataset $x: \ x_1, ..., x_n$, and our aim is to get an insight from these numbers, to get an information about the Population, about the *process* generating that Dataset.

We can, of course, calculate the Sample Mean and the Sample Variance of our Dataset. Or, we can plot the Histogram or KDE. But will this give an info about the Population or the process generating the Dataset? Well, no, in general.

Say, if we are interested in the (undergrad) Statistics course average grade all over the world, then will the average of grades of our section give complete info about the quantity we are interested in?

Parametric Statistics: General Problem

One of the general Problems of Statistics is the following: we have a Sample, a Dataset $x: \ x_1, ..., x_n$, and our aim is to get an insight from these numbers, to get an information about the Population, about the *process* generating that Dataset.

We can, of course, calculate the Sample Mean and the Sample Variance of our Dataset. Or, we can plot the Histogram or KDE. But will this give an info about the Population or the process generating the Dataset? Well, no, in general.

Say, if we are interested in the (undergrad) Statistics course average grade all over the world, then will the average of grades of our section give complete info about the quantity we are interested in? Of course, no.

Parametric Statistics: General Problem

One of the general Problems of Statistics is the following: we have a Sample, a Dataset $x: \ x_1, ..., x_n$, and our aim is to get an insight from these numbers, to get an information about the Population, about the *process* generating that Dataset.

We can, of course, calculate the Sample Mean and the Sample Variance of our Dataset. Or, we can plot the Histogram or KDE. But will this give an info about the Population or the process generating the Dataset? Well, no, in general.

Say, if we are interested in the (undergrad) Statistics course average grade all over the world, then will the average of grades of our section give complete info about the quantity we are interested in? Of course, no. And, of course, our section average grade will be much higher than the total average.

Parametric Statistics: General Problem

One of the general Problems of Statistics is the following: we have a Sample, a Dataset $x: \; x_1, ..., x_n$, and our aim is to get an insight from these numbers, to get an information about the Population, about the *process* generating that Dataset.

We can, of course, calculate the Sample Mean and the Sample Variance of our Dataset. Or, we can plot the Histogram or KDE. But will this give an info about the Population or the process generating the Dataset? Well, no, in general.

Say, if we are interested in the (undergrad) Statistics course average grade all over the world, then will the average of grades of our section give complete info about the quantity we are interested in? Of course, no. And, of course, our section average grade will be much higher than the total average. And, of course, not because our exam test will be easier than other's tests $\ddot\smile$.

Parametric Statistics: General Problem

One of the general Problems of Statistics is the following: we have a Sample, a Dataset $x: \ x_1, ..., x_n$, and our aim is to get an insight from these numbers, to get an information about the Population, about the *process* generating that Dataset.

We can, of course, calculate the Sample Mean and the Sample Variance of our Dataset. Or, we can plot the Histogram or KDE. But will this give an info about the Population or the process generating the Dataset? Well, no, in general.

Say, if we are interested in the (undergrad) Statistics course average grade all over the world, then will the average of grades of our section give complete info about the quantity we are interested in? Of course, no. And, of course, our section average grade will be much higher than the total average. And, of course, not because our exam test will be easier than other's tests $\ddot\smile$. But because you had one of the best instructors in the world.

Parametric Statistics: General Problem

One of the general Problems of Statistics is the following: we have a Sample, a Dataset $x: \ x_1, ..., x_n$, and our aim is to get an insight from these numbers, to get an information about the Population, about the *process* generating that Dataset.

We can, of course, calculate the Sample Mean and the Sample Variance of our Dataset. Or, we can plot the Histogram or KDE. But will this give an info about the Population or the process generating the Dataset? Well, no, in general.

Say, if we are interested in the (undergrad) Statistics course average grade all over the world, then will the average of grades of our section give complete info about the quantity we are interested in? Of course, no. And, of course, our section average grade will be much higher than the total average. And, of course, not because our exam test will be easier than other's tests $\ddot\smile$. But because you had one of the best instructors in the world. For Basic Math, I mean.

Parametric Statistics: General Problem

One of the general Problems of Statistics is the following: we have a Sample, a Dataset $x: \ x_1, ..., x_n$, and our aim is to get an insight from these numbers, to get an information about the Population, about the *process* generating that Dataset.

We can, of course, calculate the Sample Mean and the Sample Variance of our Dataset. Or, we can plot the Histogram or KDE. But will this give an info about the Population or the process generating the Dataset? Well, no, in general.

Say, if we are interested in the (undergrad) Statistics course average grade all over the world, then will the average of grades of our section give complete info about the quantity we are interested in? Of course, no. And, of course, our section average grade will be much higher than the total average. And, of course, not because our exam test will be easier than other's tests $\ddot\smile$. But because you had one of the best instructors in the world. For Basic Math, I mean. No kidding! $\ddot\smile$.

Parametric Statistics: Modeling

Again, assume we have a Dataset $x: \ x_1, ..., x_n$.

Parametric Statistics: Modeling

Again, assume we have a Dataset $x: \ x_1, ..., x_n$. For a Statistical analysis, statisticians assume that this Dataset is a realization of a finite sequence of r.v.s $X_1, X_2, ..., X_n$.

Parametric Statistics: Modeling

Again, assume we have a Dataset $x: \ x_1, ..., x_n$. For a Statistical analysis, statisticians assume that this Dataset is a realization of a finite sequence of r.v.s $X_1, X_2, ..., X_n$.

Why is this? Because statisticians think about the Dataset x as being **one of the possible realizations** in the universe of all possible scenarios.

Parametric Statistics: Modeling

Again, assume we have a Dataset $x: \ x_1, ..., x_n$. For a Statistical analysis, statisticians assume that this Dataset is a realization of a finite sequence of r.v.s $X_1, X_2, ..., X_n$.

Why is this? Because statisticians think about the Dataset x as being **one of the possible realizations** in the universe of all possible scenarios. Say, if we consider the obervations in the `cars` Dataset:

```
str(cars)
```

```
## 'data.frame':    50 obs. of  2 variables:
##  $ speed: num  4 4 7 7 8 9 10 10 10 11 ...
##  $ dist : num  2 10 4 22 16 10 18 26 34 17 ...
```

Parametric Statistics: Modeling

Again, assume we have a Dataset $x: \ x_1, ..., x_n$. For a Statistical analysis, statisticians assume that this Dataset is a realization of a finite sequence of r.v.s $X_1, X_2, ..., X_n$.

Why is this? Because statisticians think about the Dataset x as being **one of the possible realizations** in the universe of all possible scenarios. Say, if we consider the obervations in the `cars` Dataset:

```
str(cars)
```

```
## 'data.frame':    50 obs. of  2 variables:
##  $ speed: num  4 4 7 7 8 9 10 10 10 11 ...
##  $ dist : num  2 10 4 22 16 10 18 26 34 17 ...
```

then we think about this as a Sample, and we realize that if we would make another Sampling, we'd get other numbers. Even with the same cars!

Parametric Statistics: Modeling

So we will think about our Dataset $x_1, ..., x_n$ as being one possible realization (possible values) of the r.v. Sample $X_1, X_2, ..., X_n$.

Parametric Statistics: Modeling

So we will think about our Dataset $x_1, ..., x_n$ as being one possible realization (possible values) of the r.v. Sample $X_1, X_2, ..., X_n$. Then we can think about the sequence $X_1, ..., X_n$ as the **process generating samples** $x_1, ..., x_n$.

Parametric Statistics: Modeling

So we will think about our Dataset $x_1, ..., x_n$ as being one possible realization (possible values) of the r.v. Sample $X_1, X_2, ..., X_n$. Then we can think about the sequence $X_1, ..., X_n$ as the **process generating samples** $x_1, ..., x_n$. And we can think about X_1 (or, generic, X) as the process generating numbers in our Dataset.

Example: If we consider the weights (in Kg) of 10 persons:

$$69.5, 77.1, ..., 109,$$

then we make the following model: let X_1 be the weight of the first person (say, the first person we will meet when performing the experiment), X_2 be the weight of the second person,. . . , X_{10} be the weight of the 10-th person.

Parametric Statistics: Modeling

So we will think about our Dataset $x_1, ..., x_n$ as being one possible realization (possible values) of the r.v. Sample $X_1, X_2, ..., X_n$. Then we can think about the sequence $X_1, ..., X_n$ as the **process generating samples** $x_1, ..., x_n$. And we can think about X_1 (or, generic, X) as the process generating numbers in our Dataset.

Example: If we consider the weights (in Kg) of 10 persons:

$$69.5, 77.1, ..., 109,$$

then we make the following model: let X_1 be the weight of the first person (say, the first person we will meet when performing the experiment), X_2 be the weight of the second person,. . . , X_{10} be the weight of the 10-th person. Before making the observations, $X_1, ..., X_{10}$ are random.

Parametric Statistics: Modeling

So we will think about our Dataset $x_1, ..., x_n$ as being one possible realization (possible values) of the r.v. Sample $X_1, X_2, ..., X_n$. Then we can think about the sequence $X_1, ..., X_n$ as the **process generating samples** $x_1, ..., x_n$. And we can think about X_1 (or, generic, X) as the process generating numbers in our Dataset.

Example: If we consider the weights (in Kg) of 10 persons:

$$69.5, 77.1, ..., 109,$$

then we make the following model: let X_1 be the weight of the first person (say, the first person we will meet when performing the experiment), X_2 be the weight of the second person,. . . , X_{10} be the weight of the 10-th person. Before making the observations, $X_1, ..., X_{10}$ are random. And, in generic, let X be the r.v. showing the weight of a random person.

Parametric Statistics: Modeling

So we will think about our Dataset $x_1, ..., x_n$ as being one possible realization (possible values) of the r.v. Sample $X_1, X_2, ..., X_n$. Then we can think about the sequence $X_1, ..., X_n$ as the **process generating samples** $x_1, ..., x_n$. And we can think about X_1 (or, generic, X) as the process generating numbers in our Dataset.

Example: If we consider the weights (in Kg) of 10 persons:

$$69.5, 77.1, ..., 109,$$

then we make the following model: let X_1 be the weight of the first person (say, the first person we will meet when performing the experiment), X_2 be the weight of the second person,. . . , X_{10} be the weight of the 10-th person. Before making the observations, $X_1, ..., X_{10}$ are random. And, in generic, let X be the r.v. showing the weight of a random person.

Our Dataset of weights is just one of the possible realizations of $X_1, ..., X_{10}$.

Example

Example: Let me make a simulation: say, I want to have a model for the height of a 21 year male person. To that end I will use a Sample of size 6. Instead of randomly asking 6 persons, I will use computer to get that Sample.

Example

Example: Let me make a simulation: say, I want to have a model for the height of a 21 year male person. To that end I will use a Sample of size 6. Instead of randomly asking 6 persons, I will use computer to get that Sample. Now I start from the other end - assume I *already know* the solution: I know beforehand that the heights are Normally Distributed with the mean $155cm$ and variance $30cm^2$ (well, in fact, this is what we want to obtain).

Example

Example: Let me make a simulation: say, I want to have a model for the height of a 21 year male person. To that end I will use a Sample of size 6. Instead of randomly asking 6 persons, I will use computer to get that Sample. Now I start from the other end - assume I *already know* the solution: I know beforehand that the heights are Normally Distributed with the mean $155cm$ and variance $30cm^2$ (well, in fact, this is what we want to obtain). Then I can run

```
rnorm(6, mean = 155, sd = sqrt(30))
```

```
## [1] 155.6089 156.0454 158.5412 149.2057 157.9417 151.663
```

This is my Sample.

Example

Example: Let me make a simulation: say, I want to have a model for the height of a 21 year male person. To that end I will use a Sample of size 6. Instead of randomly asking 6 persons, I will use computer to get that Sample. Now I start from the other end - assume I *already know* the solution: I know beforehand that the heights are Normally Distributed with the mean $155cm$ and variance $30cm^2$ (well, in fact, this is what we want to obtain). Then I can run

```
rnorm(6, mean = 155, sd = sqrt(30))
```

```
## [1] 155.6089 156.0454 158.5412 149.2057 157.9417 151.663
```

This is my Sample. If I will run the code again (in some sense, ask another 7 random persons), I will get, say,

```
rnorm(6, mean = 155, sd = sqrt(30))
```

```
## [1] 163.3288 156.5705 147.0360 153.0763 158.5617 150.610
```

Example

Example: Let me make a simulation: say, I want to have a model for the height of a 21 year male person. To that end I will use a Sample of size 6. Instead of randomly asking 6 persons, I will use computer to get that Sample. Now I start from the other end - assume I *already know* the solution: I know beforehand that the heights are Normally Distributed with the mean $155cm$ and variance $30cm^2$ (well, in fact, this is what we want to obtain). Then I can run

```
rnorm(6, mean = 155, sd = sqrt(30))
```

```
## [1] 155.6089 156.0454 158.5412 149.2057 157.9417 151.66
```

This is my Sample. If I will run the code again (in some sense, ask another 7 random persons), I will get, say,

```
rnorm(6, mean = 155, sd = sqrt(30))
```

```
## [1] 163.3288 156.5705 147.0360 153.0763 158.5617 150.61
```

And so on.

Example, Cont'd

Now let's start from the initial point of Parametric Statistics: assume that the heights are Normally Distributed, but I do not know the parameters - the Mean and Variance.

Example, Cont'd

Now let's start from the initial point of Parametric Statistics: assume that the heights are Normally Distributed, but I do not know the parameters - the Mean and Variance. But I just have one of the above Samples as an observation of heights.

Example, Cont'd

Now let's start from the initial point of Parametric Statistics: assume that the heights are Normally Distributed, but I do not know the parameters - the Mean and Variance. But I just have one of the above Samples as an observation of heights. And our task is not to get an information about that specific observation, but *the total process* generating heights, i.e., information about the Distribution of heights.

Example, Cont'd

Now let's start from the initial point of Parametric Statistics: assume that the heights are Normally Distributed, but I do not know the parameters - the Mean and Variance. But I just have one of the above Samples as an observation of heights. And our task is not to get an information about that specific observation, but *the total process* generating heights, i.e., information about the Distribution of heights.

So, again, having a Dataset $x_1, ..., x_n$, statisticians work with a r.v.s $X_1, X_2, ..., X_n$ to work not only with a particular Sample, but with **all possible samples** from the Distribution (Process) behind the phenomenon.

Parametric Statistics: Modeling

And we give a

Definition: We will call the collection of IID r.v.s $X_1, X_2, ..., X_n$ a **Random Sample** of size n.

Parametric Statistics: Modeling

And we give a

Definition: We will call the collection of IID r.v.s $X_1, X_2, ..., X_n$ a **Random Sample** of size n.

In the rest of our story, we will work with Random Samples.

Parametric Statistics: Modeling

And we give a

Definition: We will call the collection of IID r.v.s $X_1, X_2, ..., X_n$ a **Random Sample** of size n.

In the rest of our story, we will work with Random Samples.

Now, if we have a Random Sample $X_1, ..., X_n$, then, because they are IID, we will have that all X_k-s are coming from the same Distribution:

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

Parametric Statistics: Modeling

And we give a

Definition: We will call the collection of IID r.v.s $X_1, X_2, ..., X_n$ a **Random Sample** of size n.

In the rest of our story, we will work with Random Samples.

Now, if we have a Random Sample $X_1, ..., X_n$, then, because they are IID, we will have that all X_k-s are coming from the same Distribution:

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

Our general Problem will be: get an information about $\mathcal{F}$, estimate/recover $\mathcal{F}$.

Parametric Statistics: Modeling

And we give a

Definition: We will call the collection of IID r.v.s $X_1, X_2, ..., X_n$ a **Random Sample** of size n.

In the rest of our story, we will work with Random Samples.

Now, if we have a Random Sample $X_1, ..., X_n$, then, because they are IID, we will have that all X_k-s are coming from the same Distribution:

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

Our general Problem will be: get an information about $\mathcal{F}$, estimate/recover $\mathcal{F}$.

This problem is very general, and hence, not so much can be said about $\mathcal{F}$.

Parametric Statistics: Modeling

And we give a

Definition: We will call the collection of IID r.v.s $X_1, X_2, ..., X_n$ a **Random Sample** of size n.

In the rest of our story, we will work with Random Samples.

Now, if we have a Random Sample $X_1, ..., X_n$, then, because they are IID, we will have that all X_k-s are coming from the same Distribution:

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

Our general Problem will be: get an information about $\mathcal{F}$, estimate/recover $\mathcal{F}$.

This problem is very general, and hence, not so much can be said about $\mathcal{F}$. So, we need to impose some conditions about $\mathcal{F}$ to be able to get some more information about it.

Parametric Statistics: Modeling

In Parametric Statistics, we assume that we have a Random Sample

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

and $\mathcal{F}$ is a member of the Parametric Familiy of Distributions:

$$\mathcal{F} = \mathcal{F}_\theta, \qquad \theta \in \Theta \subset \mathbb{R}^m.$$

Parametric Statistics: Modeling

In Parametric Statistics, we assume that we have a Random Sample

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

and $\mathcal{F}$ is a member of the Parametric Familiy of Distributions:

$$\mathcal{F} = \mathcal{F}_\theta, \qquad \theta \in \Theta \subset \mathbb{R}^m.$$

Here

- θ is our Parameter, it is **fixed, but unknown**, it is, in general, m-dimensional;

Parametric Statistics: Modeling

In Parametric Statistics, we assume that we have a Random Sample

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

and $\mathcal{F}$ is a member of the Parametric Familiy of Distributions:

$$\mathcal{F} = \mathcal{F}_\theta, \qquad \theta \in \Theta \subset \mathbb{R}^m.$$

Here

- θ is our Parameter, it is **fixed, but unknown**, it is, in general, m-dimensional;
- $\Theta \subset \mathbb{R}^m$ is our Parameter Set, the set of all possible values of our Parameter θ;

Parametric Statistics: Modeling

In Parametric Statistics, we assume that we have a Random Sample

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

and $\mathcal{F}$ is a member of the Parametric Familiy of Distributions:

$$\mathcal{F} = \mathcal{F}_\theta, \qquad \theta \in \Theta \subset \mathbb{R}^m.$$

Here

- θ is our Parameter, it is **fixed, but unknown**, it is, in general, m-dimensional;
- $\Theta \subset \mathbb{R}^m$ is our Parameter Set, the set of all possible values of our Parameter θ;
- $\mathcal{F}_\theta$ is our Parametric Distribution.

Parametric Statistics: Modeling

We will consider one of the main Problems of the Parametric Statistics: **Using the observations from our Random Sample, estimate the value of the Parameter θ.**

Parametric Statistics: Modeling

We will consider one of the main Problems of the Parametric Statistics: **Using the observations from our Random Sample, estimate the value of the Parameter θ.**

Note: Sometimes we are required to estimate not the parameter itself, but some function of the parameter. Say, we want to estimate $\dfrac{\theta^2+1}{\log\theta+2}$.

Parametric Statistics: Modeling

We will consider one of the main Problems of the Parametric Statistics: **Using the observations from our Random Sample, estimate the value of the Parameter θ.**

Note: Sometimes we are required to estimate not the parameter itself, but some function of the parameter. Say, we want to estimate $\dfrac{\theta^2+1}{\log\theta+2}$. Or, if our model is $\mathcal{F}_\theta$ and $X \sim \mathcal{F}_\theta$, then we want to estimate $\mathbb{E}(X)$ or $Var(X)$ - these are functions of θ.

Parametric Statistics: Modeling

We will consider one of the main Problems of the Parametric Statistics: **Using the observations from our Random Sample, estimate the value of the Parameter θ.**

Note: Sometimes we are required to estimate not the parameter itself, but some function of the parameter. Say, we want to estimate $\dfrac{\theta^2+1}{\log\theta+2}$. Or, if our model is $\mathcal{F}_\theta$ and $X \sim \mathcal{F}_\theta$, then we want to estimate $\mathbb{E}(X)$ or $Var(X)$ - these are functions of θ.

The set of Distributions

$$\{\mathcal{F}_\theta : \theta \in \Theta\}$$

is called a **Statistical Model**.

Example

Example: Assume we have a coin, and we are tossing it n times, and let $x_1, x_2, ..., x_n$ be the result of that n tosses: $x_k = 1$, of the k-the toss resulted in Heads, and $x_k = 0$ otherwise.

Example

Example: Assume we have a coin, and we are tossing it n times, and let $x_1, x_2, ..., x_n$ be the result of that n tosses: $x_k = 1$, of the k-the toss resulted in Heads, and $x_k = 0$ otherwise. Then we can model this experiment as a realization of a Random Sample $X_1, X_2, ..., X_n$, where

$$X_1, X_2, ..., X_n \sim$$

Example

Example: Assume we have a coin, and we are tossing it n times, and let $x_1, x_2, ..., x_n$ be the result of that n tosses: $x_k = 1$, of the k-the toss resulted in Heads, and $x_k = 0$ otherwise. Then we can model this experiment as a realization of a Random Sample $X_1, X_2, ..., X_n$, where

$$X_1, X_2, ..., X_n \sim Bernoulli(p), \qquad p \in [0, 1],$$

where p is the Probability of Heads for our coin.

Example

Example: Assume we have a coin, and we are tossing it n times, and let $x_1, x_2, ..., x_n$ be the result of that n tosses: $x_k = 1$, of the k-the toss resulted in Heads, and $x_k = 0$ otherwise. Then we can model this experiment as a realization of a Random Sample $X_1, X_2, ..., X_n$, where

$$X_1, X_2, ..., X_n \sim Bernoulli(p), \qquad p \in [0, 1],$$

where p is the Probability of Heads for our coin. In this case,

- our Parameter is $p = \theta$;

Example

Example: Assume we have a coin, and we are tossing it n times, and let $x_1, x_2, ..., x_n$ be the result of that n tosses: $x_k = 1$, of the k-the toss resulted in Heads, and $x_k = 0$ otherwise. Then we can model this experiment as a realization of a Random Sample $X_1, X_2, ..., X_n$, where

$$X_1, X_2, ..., X_n \sim Bernoulli(p), \qquad p \in [0, 1],$$

where p is the Probability of Heads for our coin. In this case,

- our Parameter is $p = \theta$;
- the set $[0, 1] = \Theta \subset \mathbb{R}$ is our Parameters set;

Example

Example: Assume we have a coin, and we are tossing it n times, and let $x_1, x_2, ..., x_n$ be the result of that n tosses: $x_k = 1$, of the k-the toss resulted in Heads, and $x_k = 0$ otherwise. Then we can model this experiment as a realization of a Random Sample $X_1, X_2, ..., X_n$, where

$$X_1, X_2, ..., X_n \sim Bernoulli(p), \qquad p \in [0, 1],$$

where p is the Probability of Heads for our coin. In this case,

- our Parameter is $p = \theta$;
- the set $[0, 1] = \Theta \subset \mathbb{R}$ is our Parameters set;
- the Parametric family of Distributions is the family $Bernoulli(p)$, $p \in [0, 1]$.

Example

Example: Assume we have a coin, and we are tossing it n times, and let $x_1, x_2, ..., x_n$ be the result of that n tosses: $x_k = 1$, of the k-the toss resulted in Heads, and $x_k = 0$ otherwise. Then we can model this experiment as a realization of a Random Sample $X_1, X_2, ..., X_n$, where

$$X_1, X_2, ..., X_n \sim Bernoulli(p), \qquad p \in [0, 1],$$

where p is the Probability of Heads for our coin. In this case,

- our Parameter is $p = \theta$;
- the set $[0, 1] = \Theta \subset \mathbb{R}$ is our Parameters set;
- the Parametric family of Distributions is the family $Bernoulli(p)$, $p \in [0, 1]$.

And a generic $X \sim Bernoulli(p)$ will be the result of our coin toss.

Example

Example: Assume we have a coin, and we are tossing it n times, and let $x_1, x_2, ..., x_n$ be the result of that n tosses: $x_k = 1$, of the k-the toss resulted in Heads, and $x_k = 0$ otherwise. Then we can model this experiment as a realization of a Random Sample $X_1, X_2, ..., X_n$, where

$$X_1, X_2, ..., X_n \sim Bernoulli(p), \qquad p \in [0, 1],$$

where p is the Probability of Heads for our coin. In this case,

- our Parameter is $p = \theta$;
- the set $[0, 1] = \Theta \subset \mathbb{R}$ is our Parameters set;
- the Parametric family of Distributions is the family $Bernoulli(p)$, $p \in [0, 1]$.

And a generic $X \sim Bernoulli(p)$ will be the result of our coin toss.

In this problem, p is **fixed, but unknown**.

Example

Example: Assume we have a coin, and we are tossing it n times, and let $x_1, x_2, ..., x_n$ be the result of that n tosses: $x_k = 1$, of the k-the toss resulted in Heads, and $x_k = 0$ otherwise. Then we can model this experiment as a realization of a Random Sample $X_1, X_2, ..., X_n$, where

$$X_1, X_2, ..., X_n \sim Bernoulli(p), \qquad p \in [0, 1],$$

where p is the Probability of Heads for our coin. In this case,

- our Parameter is $p = \theta$;
- the set $[0, 1] = \Theta \subset \mathbb{R}$ is our Parameters set;
- the Parametric family of Distributions is the family $Bernoulli(p)$, $p \in [0, 1]$.

And a generic $X \sim Bernoulli(p)$ will be the result of our coin toss.

In this problem, p is **fixed, but unknown**. And our aim will be to estimate p, using our observations $x_1, ..., x_n$.

Example

Example: Assume $x_1, x_2, ..., x_n$ is the daily number of car accidents for day $1, 2, ..., n$ for some city. Then we can model this by using a Random Sample

$$X_1, X_2, ..., X_n \sim$$

Example

Example: Assume $x_1, x_2, ..., x_n$ is the daily number of car accidents for day $1, 2, ..., n$ for some city. Then we can model this by using a Random Sample

$$X_1, X_2, ..., X_n \sim Pois(\lambda),$$

where $\lambda \in (0, +\infty)$.

Example

Example: Assume $x_1, x_2, ..., x_n$ is the daily number of car accidents for day $1, 2, ..., n$ for some city. Then we can model this by using a Random Sample

$$X_1, X_2, ..., X_n \sim Pois(\lambda),$$

where $\lambda \in (0, +\infty)$.

- Here our Parameter is $\lambda = \theta$; it is 1D;

Example

Example: Assume $x_1, x_2, ..., x_n$ is the daily number of car accidents for day $1, 2, ..., n$ for some city. Then we can model this by using a Random Sample

$$X_1, X_2, ..., X_n \sim Pois(\lambda),$$

where $\lambda \in (0, +\infty)$.

- Here our Parameter is $\lambda = \theta$; it is 1D;
- The set of Parameters is $(0, +\infty) = \Theta \subset \mathbb{R}$;

Example

Example: Assume $x_1, x_2, ..., x_n$ is the daily number of car accidents for day $1, 2, ..., n$ for some city. Then we can model this by using a Random Sample

$$X_1, X_2, ..., X_n \sim Pois(\lambda),$$

where $\lambda \in (0, +\infty)$.

- Here our Parameter is $\lambda = \theta$; it is 1D;
- The set of Parameters is $(0, +\infty) = \Theta \subset \mathbb{R}$;
- The Parametric Family of Distributions is $Pois(\lambda)$.

Example

Example: Assume $x_1, x_2, ..., x_n$ is the daily number of car accidents for day $1, 2, ..., n$ for some city. Then we can model this by using a Random Sample

$$X_1, X_2, ..., X_n \sim Pois(\lambda),$$

where $\lambda \in (0, +\infty)$.

- Here our Parameter is $\lambda = \theta$; it is 1D;
- The set of Parameters is $(0, +\infty) = \Theta \subset \mathbb{R}$;
- The Parametric Family of Distributions is $Pois(\lambda)$.

A generic $X \sim Pois(\lambda)$ can be interpreted as the number of accidents for a generic day.

Example

Example: Assume $x_1, x_2, ..., x_n$ is the daily number of car accidents for day $1, 2, ..., n$ for some city. Then we can model this by using a Random Sample

$$X_1, X_2, ..., X_n \sim Pois(\lambda),$$

where $\lambda \in (0, +\infty)$.

- Here our Parameter is $\lambda = \theta$; it is 1D;
- The set of Parameters is $(0, +\infty) = \Theta \subset \mathbb{R}$;
- The Parametric Family of Distributions is $Pois(\lambda)$.

A generic $X \sim Pois(\lambda)$ can be interpreted as the number of accidents for a generic day.

And our problem here will be to estimate our unknown λ, using the realizations $x_1, x_2, ..., x_n$.

Example

Example: We consider a Sample of heights of some persons: $x_1, x_2, ..., x_n$. We assume this is a realization of a Random Sample

$$X_1, X_2, ..., X_n \sim$$

Example

Example: We consider a Sample of heights of some persons: $x_1, x_2, ..., x_n$. We assume this is a realization of a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2), \qquad \mu \in \mathbb{R}, \sigma^2 \geq 0.$$

Example

Example: We consider a Sample of heights of some persons: $x_1, x_2, ..., x_n$. We assume this is a realization of a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2), \qquad \mu \in \mathbb{R}, \sigma^2 \geq 0.$$

Here

- Our Parameter is $(\mu, \sigma^2) = \theta$ (or $(\mu, \sigma) = \theta$), which is 2D;

Example

Example: We consider a Sample of heights of some persons: $x_1, x_2, ..., x_n$. We assume this is a realization of a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2), \qquad \mu \in \mathbb{R}, \sigma^2 \geq 0.$$

Here

- Our Parameter is $(\mu, \sigma^2) = \theta$ (or $(\mu, \sigma) = \theta$), which is 2D;
- The Parameter Set is $\mathbb{R} \times [0, +\infty) = \Theta \subset \mathbb{R}^2$;

Example

Example: We consider a Sample of heights of some persons: $x_1, x_2, ..., x_n$. We assume this is a realization of a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2), \qquad \mu \in \mathbb{R}, \sigma^2 \geq 0.$$

Here

- Our Parameter is $(\mu, \sigma^2) = \theta$ (or $(\mu, \sigma) = \theta$), which is 2D;
- The Parameter Set is $\mathbb{R} \times [0, +\infty) = \Theta \subset \mathbb{R}^2$;
- The Parametric Family of Distributions is $\mathcal{N}(\mu, \sigma^2)$.

Example

Example: We consider a Sample of heights of some persons: $x_1, x_2, ..., x_n$. We assume this is a realization of a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2), \qquad \mu \in \mathbb{R}, \sigma^2 \geq 0.$$

Here

- Our Parameter is $(\mu, \sigma^2) = \theta$ (or $(\mu, \sigma) = \theta$), which is 2D;
- The Parameter Set is $\mathbb{R} \times [0, +\infty) = \Theta \subset \mathbb{R}^2$;
- The Parametric Family of Distributions is $\mathcal{N}(\mu, \sigma^2)$.

We can think about $X \sim \mathcal{N}(\mu, \sigma^2)$ as the height of a generic, random person.

Example

Example: We consider a Sample of heights of some persons: $x_1, x_2, ..., x_n$. We assume this is a realization of a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2), \qquad \mu \in \mathbb{R}, \sigma^2 \geq 0.$$

Here

- Our Parameter is $(\mu, \sigma^2) = \theta$ (or $(\mu, \sigma) = \theta$), which is 2D;
- The Parameter Set is $\mathbb{R} \times [0, +\infty) = \Theta \subset \mathbb{R}^2$;
- The Parametric Family of Distributions is $\mathcal{N}(\mu, \sigma^2)$.

We can think about $X \sim \mathcal{N}(\mu, \sigma^2)$ as the height of a generic, random person.

Our Problem here is, using the observation $x_1, x_2, ..., x_n$, to estimate μ and σ^2.