AUA CS108, Statistics, Fall 2020

Lecture 25

Michael Poghosyan

23 Oct 2020

Contents

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: 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:

```r
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] 152.7771 153.6793 158.3788 160.5891 155.1615 156.69
```

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] 152.7771 153.6793 158.3788 160.5891 155.1615 156.69
```

This is my Sample. If I will run the code again (in some sense, ask another 6 random persons), I will get, say,

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

```
## [1] 143.3883 160.9973 157.3698 157.2665 150.7009 157.84
```

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] 152.7771 153.6793 158.3788 160.5891 155.1615 156.69
```

This is my Sample. If I will run the code again (in some sense, ask another 6 random persons), I will get, say,

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

```
## [1] 143.3883 160.9973 157.3698 157.2665 150.7009 157.84
```

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

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.

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.**

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 $x_1, ..., x_n$.