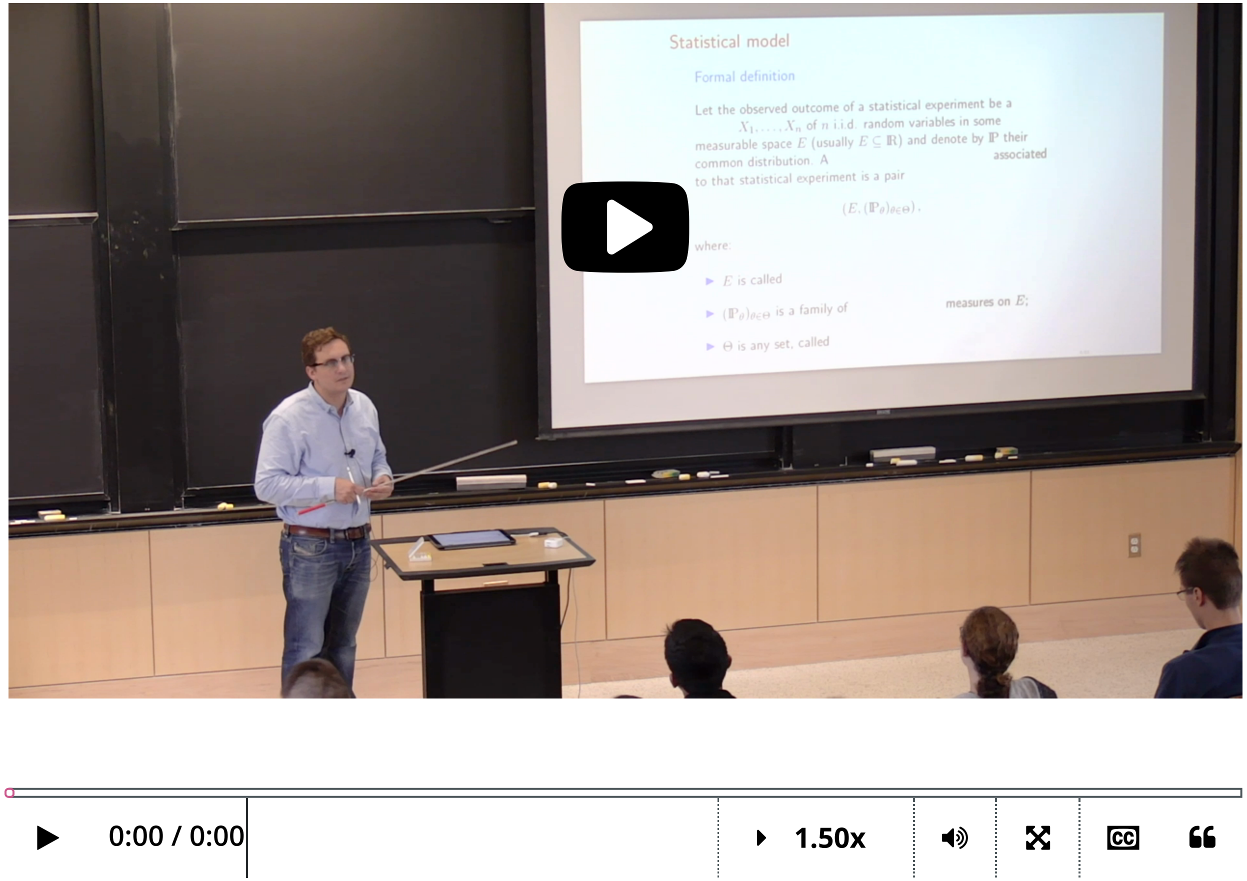
5. Statistical model

Statistical model: definition

Start of transcript. Skip to the end.



Statistical model

Formal definition

Let the observed outcome of a statistical experiment be a $X_1, \dots, X_n$ of n i.i.d. random variables in some measurable space E (usually $E \subseteq \mathbb{R}$) and denote by $\mathbb{P}$ their common distribution. A associated to that statistical experiment is a pair

$$(E, (\mathbb{P}_\theta)_{\theta \in \Theta}),$$

where:

- E is called
- $(\mathbb{P}_\theta)_{\theta \in \Theta}$ is a family of measures on E;
- Θ is any set, called

So let's talk about a statistical model.

Here I said, you know, I want to replace

my PDF by a particular statistical model, which was

in this case a Poisson model.

A model just means something which

is like slightly simpler than what reality actually

is, but hopefully captures most of it.

Well, that would be a good model.

And so I'm going to have a statistical

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

A Basic Statistical Model: Sample space

1/1 point (graded)

You have a coin that either lands heads, which you denote by 1, or tails, which you denote by 0. Let X be a random variable representing this coin flip, with an (unknown) distribution. You run a **statistical experiment** consisting of n iid tosses of the coin and record your data set as $X_1, X_2, X_3, \ldots X_n$.

(It makes sense to assume the coin tosses $X_1, \ldots, X_n$ as identically distributed, since we always toss the same coin; and as independent, since these tosses do not affect each other.)

We now construct a **statistical model** $\left(E, \{P_\theta\}_{\theta\in\Theta}\right)$ associated with this experiment, where

$$\begin{aligned} E &\quad \text{is a sample space for} X, \text{i.e. a set that contains all possible outcomes of } X, \\ \{\mathbb{P}_\theta\}_{\theta\in\Theta} &\quad \text{is a family of probability distributions on } E, \\ \Theta &\quad \text{is a parameter set, i.e. a set consisting of some possible values of } \theta. \end{aligned}$$

What is the **smallest sample space** for X? We can use this as the sample space E in our statistical model.
(Below, $[0,1]$ denotes the closed interval between 0 and 1. In contrast, $\{0,1\}$ denotes the set with two elements, 0 and 1.)

● $\{0,1\}$ ✓

○ $[0,1]$

○ $\mathbb{R}$

○ $\mathbb{R}^2$

Solution:

Here the coin is either heads (denoted by 1) or tails (denoted by 0), so $\{0,1\}$ is the smallest sample space of X. The remaining choices are valid, but not the smallest, sample spaces of X.

Submit You have used 1 of 2 attempts

ℹ Answers are displayed within the problem

A Basic Statistical Model: Family of distributions and Parameter set

2/2 points (graded)

Continuing from the previous problem, which of the following is the smallest family of probability distributions that the distribution of X belongs to? We can use this family as $\{\mathbb{P}_\theta\}_{\theta\in\Theta}$ in our statistical model.

● Bernoulli ✓

○ Poisson

○ Binomial

The distribution of X is a member of the family with some unknown parameter θ. According to the information given about the experiment, which of the following represents the set of all possible values of the parameter θ? We can use this set as the parameter set Θ in our statistical model.

○ $\{0,1\}$

○ $\{0,1/2,1\}$

● $[0,1]$ ✓

○ $\mathbb{R}$

Solution:

1. Since the (smallest) sample space of X is $\{0,1\}$, X follows a Bernoulli distribution.

2. The first and second choices, $\{0,1\}$ and $\{0,1/2,1\}$, place too many restrictions on the distribution of X. Also, be sure to not confuse the space where the parameter θ lives with the sample space, where the random variable X lives! The fourth choice, $\mathbb{R}$, allows for values of θ that do not make sense according to modeling X as $\mathrm{Ber}\,(\theta)$. For example, there is no such thing as $\mathrm{Ber}\,(-1/2)$.
   We are not given any assumptions on the distribution of the coin, so we need to allow θ to take all possible values that make sense according to our modeling assumption. Since θ represents the probability that $X=1$, we must have $0 \leq \theta \leq 1$. Hence, the third choice, $[0,1]$, is correct.

Using this problem and the previous one, we can construct the statistical model $(\{0,1\},\{\mathrm{Ber}\,(\theta)\}_{\theta\in[0,1]})$ for the distribution of the RV X representing the outcome of the coin flip.

Submit

You have used 1 of 3 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 2 Foundation of Inference:Lecture 3: Parametric Statistical Models / 5. Statistical model

Add a Post

Show all posts ▼ by recent activity ▼

? Set of outcomes of random variable X. What is it, the Domain or the Image of the function X?

As the definition of the random variable is the function I would like to clarify the definition of Sample Space. The set of the outcomes, is it the Domai... 3