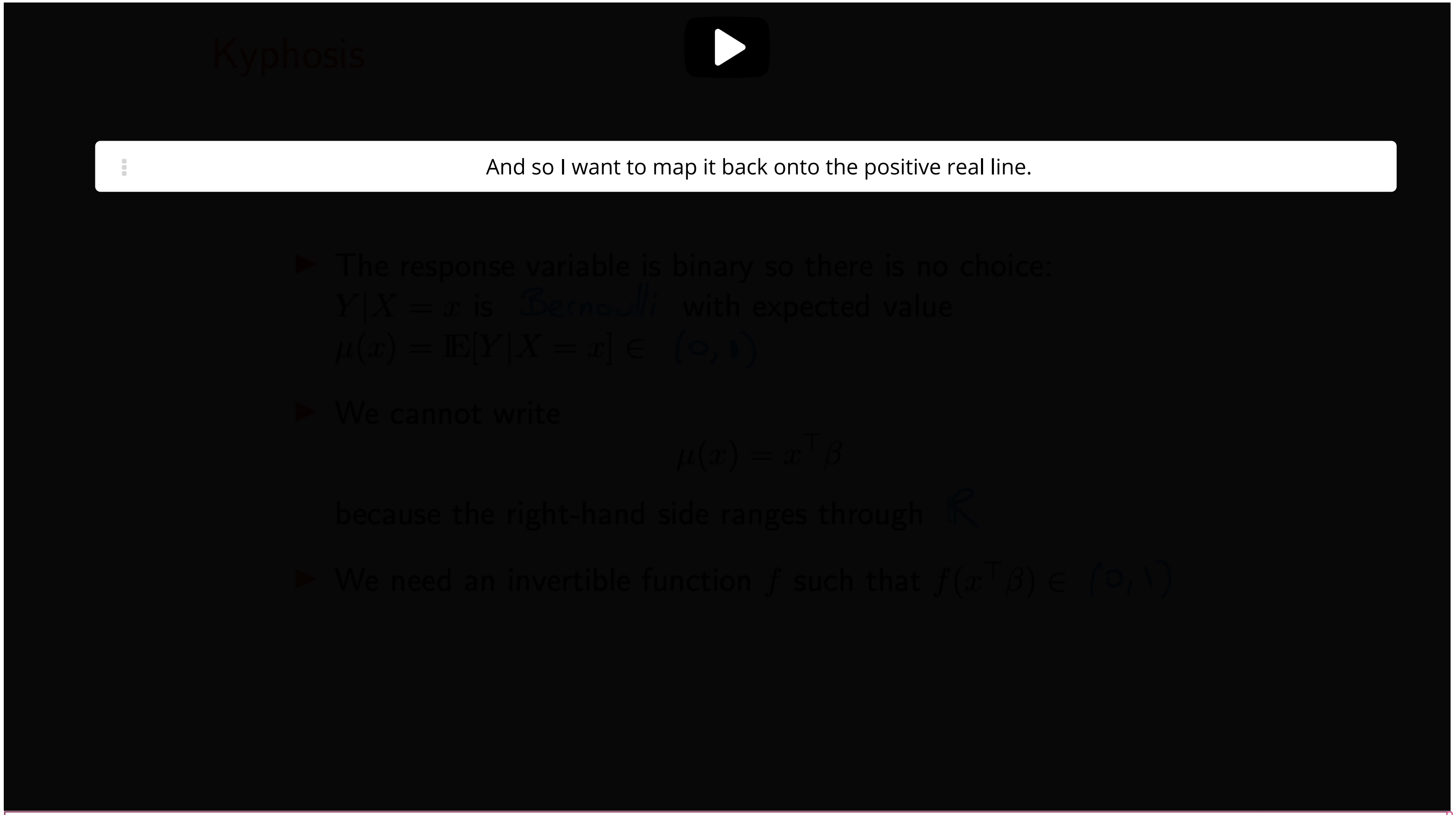
4. Motivation

Motivation: Kyphosis



7:49 / 7:49 1.50x

Video
Download video file

Transcripts
Download SubRip (.srt) file
Download Text (.txt) file

Computing the Regression Function for a Known Joint Distribution

1/1 point (graded)

Consider the pair of random variables (X, Y) where we first choose X uniformly at random from $[0, 1]$, then Y is chosen uniformly at random from $[0, X]$. What is the regression function $\mu(x)$?

(Recall that $\mu(x)$ is defined to be $\mathbb{E}[Y \mid X = x]$, so use lower case x for the variable.)

$\mu(x) =$ x/2 ✔ **Answer:** x/2

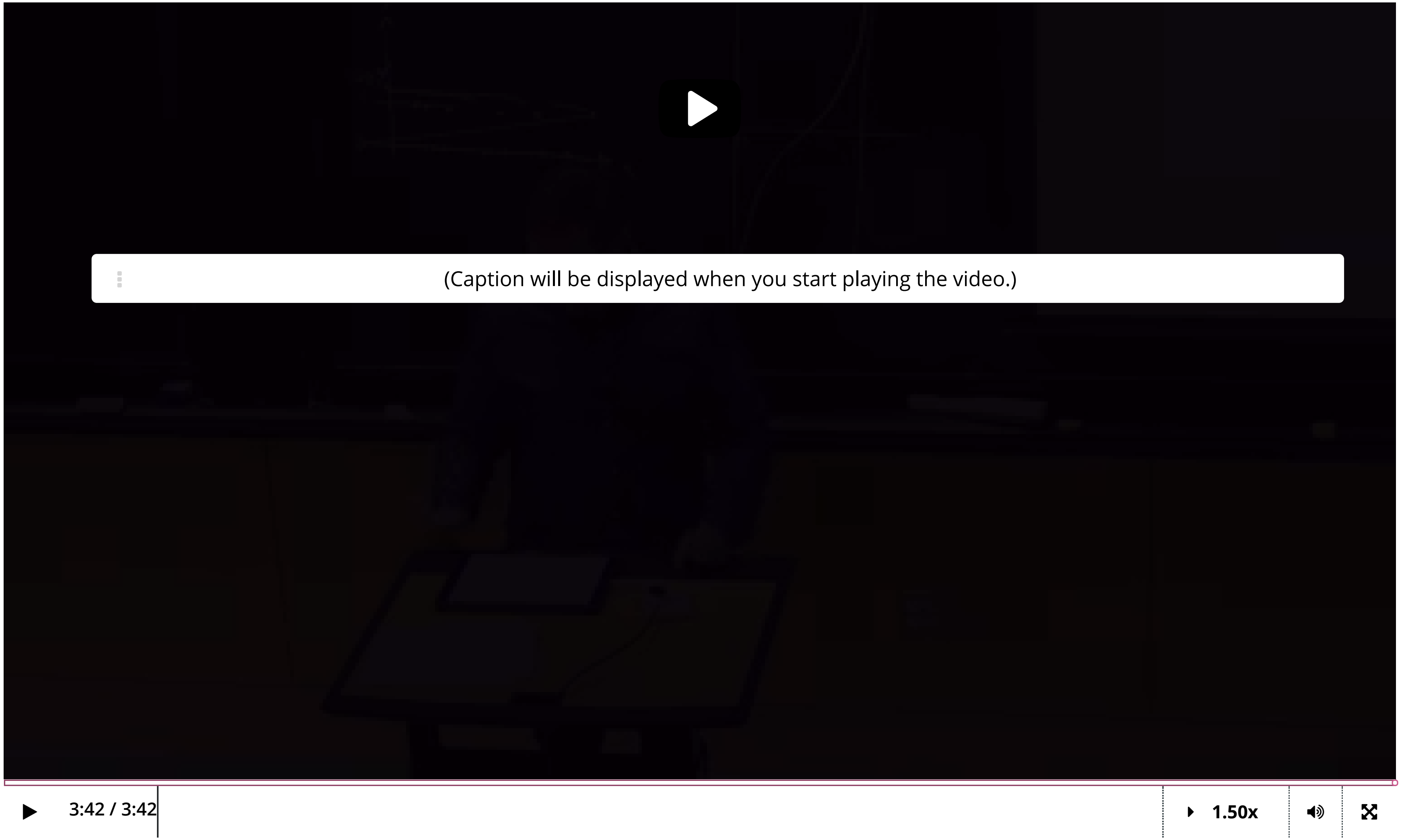
$\frac{x}{2}$

STANDARD NOTATION

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Generalizing the Linear Model; the Link Function



Video

[Download video file]

An Example: Flaky Airline Passengers

1/1 point (graded)

An airline company wishes to build a predictor for **whether or not** a passenger will show up on time for a flight (a yes/no answer), based on predictive features such as: (1) how many times the passenger missed their flight in the past, (2) the time that the ticket was purchased and (3) the predicted amount of traffic for that day.

Let $\mathbf{X}$ be the vector of predictive features and let Y be the desired feature we wish to predict for future passengers.

Which of the following sets best describes the **range** of the regression function $\mu(\mathbf{x})$?

- $\mathbb{R}$, the set of all reals.
- $\mathbb{R}^+$, the set of positive reals.
- $(0,1)$, the unit interval.
- $\{0,1\}$, a set with two possible values.

Solution:

The best choice is $(0,1)$, the unit interval. Since the model calls for Y being a yes/no indicator, the distribution of Y given X ought to be a $\{0,1\}$ -valued random variable – for example, distributed like Bernoulli(p) where p might depend on X in some way. The range of μ, then, ought to be $(0,1)$ since it is defined as an expectation $\mathbb{E}[Y|X=x]$ – which is p if we do indeed decide to model it as a Bernoulli RV.

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Flaky Airline Passengers, Continued

1/1 point (graded)

Remark: In this problem, we consider reasons for which a more generalized version of regression – as opposed to simple linear regression – might be more appropriate.

In the same setting as the previous problem (and in the context of the discussion of the solution), which of the following are true statements about μ and the pair $(\mathbf{X}, Y)$? Choose all that apply.

- [x] The range of values of Y is bounded.
- [x] The range of values of μ is strictly positive.
- [x] Based on the range of values of Y, it is harder to assume that the noise is Gaussian.
- [ ] Mathematically, linear regression is impossible to compute for Yes/No responses.
- [ ] Mathematically, linear regression is impossible to compute for integer-valued features, (e.g. $X_1 =$ the number of missed flights).

✔

Solution:

- "**Y takes values in a bounded interval**" is True, since $Y \in \{0, 1\}$, which may or may not be Bernoulli given x. Such random variables are sometimes referred to as **categorical** random variables (takes values ranging between multiple choices).

- **"The range of values of μ is strictly positive"** is True, since μ takes values in the interval $(0, 1)$.

- **"Based on the range of values of Y, it is harder to assume that the noise is Gaussian"** is True, since Gaussian random variables are generally unbounded; i.e. taking values on $\mathbb{R}$ which can be arbitrarily large. If the predictor is to be a **probability** in the interval $(0, 1)$, it is harder to justify using a normal distribution to model the noise (except perhaps in the rare case where the variance is extremely tiny, which we shouldn't take for granted).

- **"Mathematically, linear regression is impossible for Yes/No responses"** is False. We can **always** perform linear regression on datasets $\{(X_i, Y_i)\}_{i=1}^{N}$, by simply calculating a best-fit line. Such an approach operates on the assumption that the provided observations are approximately linear, and does not care about misspecification. The only question is whether or not such a technique is actually appropriate.

- **"Mathematically, linear regression is impossible for integer-valued features, (e.g. $X_1 =$ the number of missed flights)"** is False for the same reason.

*Note: The fact that we mathematically **cannot** perform linear regression (or any other statistical technique, for that matter) is not the true reason why we consider generalized linear models.*

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

In this unit, we focus on **generalized linear models** , which is a much more powerful and expressive family of models. As it turns out, this comes at a cost: finding the Maximum Likelihood Estimator becomes more difficult (in general). We relax the assumption that μ is linear. Instead, we assume that $g \circ \mu$ is linear, for some function g:

$$g(\mu(\mathbf{x})) = \mathbf{x}^T \beta.$$

The function g is assumed to be known, and is referred to as the **link function** .

We have done this with the following strategy in mind: **Through an appropriate choice of the link function, which depends on the model, we will hope to be able to compute an estimator $\hat{\beta}$, usually the MLE**.

Domain and Range of the Link Function

1/1 point (graded)
Let (X, Y) be some pair of random variables. Assume that the regression function $\mu(x)$ takes values in the range $[-3, 5]$. In order to fit into the generalized linear model context, which of the following gives the **most accurate description** of the domain and range of any candidate link function g?

- ○ $g : [-3, 5] \to [-3, 5]$
- ● $g : [-3, 5] \to \mathbb{R}$

$g : [-3, 5] \to \mathbb{R}^+$

$g : \mathbb{R} \to [-3, 5]$

$g : \mathbb{R}^+ \to [-3, 5]$

Solution:

We want link functions g so that $g(\mu(x)) = x^T\beta$ for some β. The right hand side is a linear function of x, which takes values in $\mathbb{R}$, while μ takes values in $[-3, 5]$. Therefore, the correct choice is g mapping $[-3, 5]$ to all of $\mathbb{R}$.

Submit

You have used 1 of 3 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 7 Generalized Linear Models:Lecture 21: Introduction to Generalized Linear Models; Exponential Families / 4. Motivation

Add a Post

Show all posts ▼ by recent activity ▼

There are no posts in this topic yet.