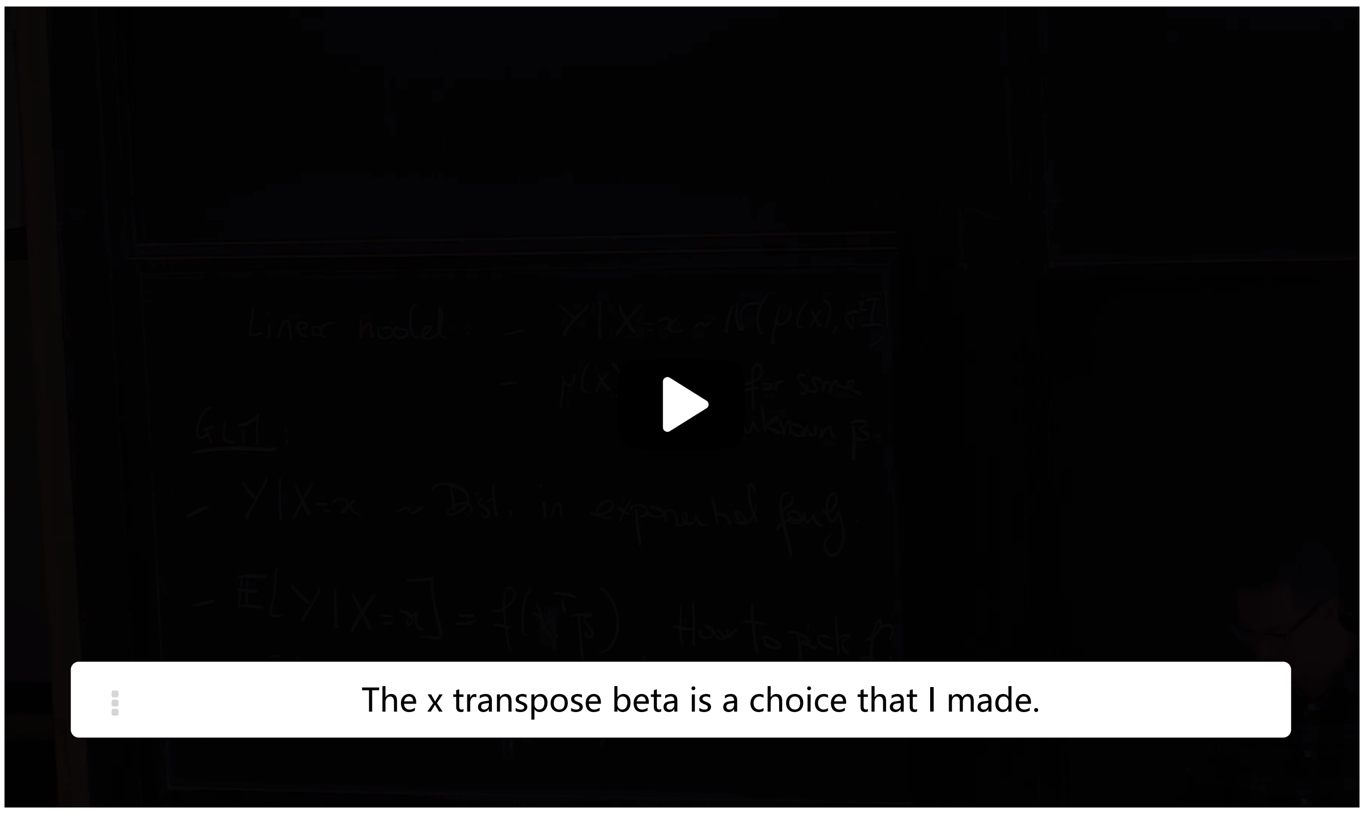
2. Recap of Generalized Linear Model Definitions and the Link Function

Recap: Generalized Linear Model and Link Function



7:05 / 7:05 1.0x

at all possible function that map our p where x lives to r,

I'm going to take a look at very specific functions

where there's a given link function f.

And then I'm just mapping--

I'm just collapsing x into one dimension

by looking at x transpose beta.

So there's still some linearity happening there.

The x transpose beta is a choice that I made.

End of transcript. Skip to the start.

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

Link Function?

1/1 point (graded)

Which one of the following are **valid** link functions? Recall that a link function g is required to be **monotone increasing** and **differentiable**. (Choose all that apply.)

Note: The link function, in general, can be monotone increasing or monotone decreasing. In this class, we have chosen as convention to require it to be monotone increasing.

- [x] $g(\mu) = \mu, \mu \in \mathbb{R}$ ✔
- [x] $g(\mu) = -\frac{1}{\mu}, \mu > 0$ ✔
- [ ] $g(\mu) = \mu^2, \mu \in \mathbb{R}$

- [x] $\ln\left(\frac{\mu^3}{1-\mu^3}\right), 1 > \mu > 0$ ✔

- [x] $-\ln\left[-\ln\left(\frac{\mu}{n}\right)\right], 0 < \mu < n$ and $n > 0$ known ✔

✔

Solution:

Choices 1, 2, 4, and 5 are valid link functions. One can verify that these functions are differentiable and monotone increasing in their domain.

Submit You have used 2 of 2 attempts

Answers are displayed within the problem

Concept Check: Linear Model and Generalized Linear Model

0 points possible (ungraded)
Which one of the following data modeling scenarios require one to **strictly use a generalized linear model over a Gaussian linear model**? (Choose all that apply.)

Note: While it is true that one can use a Gaussian linear model to fit any data (without paying attention to whether it is appropriate or not), in this problem we should use a GLM when it is more appropriate under a given scenario.

- [x] We observe data $Y_i \in \{0, 1, \ldots, n_i\}$ as a function of integers $n_i > 0$ and we wish to model the proportions Y_i/n_i. ✔

- [x] We observe $Y_i \in \mathbb{R}$ that we know are non-linearly related to the explanatory variables $\mathbf{X}_i$. ✔

- [x] The dependent variable $Y > 0$ has a discrete distribution whose expectation we wish to relate to the explanatory variable $\mathbf{X}$. ✔

✔

Solution:

All of the scenarios require us to use generalized linear models. We examine the scenarios in order:

- The first choice requires us to model proportions that lie between 0 and 1. A generalized linear model is clearly a better fit when compared to a linear model.

- The second choice suggests that we should apply a generalized linear model because we know the ground truth that the dependent variable is non-linearly related to the explanatory variables.

- In the third choice, the dependent variable Y has a discrete distribution and it is stated that $Y > 0$. If we are to fit the data using a model, a generalized linear model is better than a linear model because of multiple reasons. For one, the restriction $Y > 0$ can be satisfied if we try to explain Y for an unobserved data sample $\mathbf{X}$ via the regression function $\mu(\mathbf{X})$ and the link function $g(\cdot)$: $\mu(\mathbf{X}) = g^{-1}(\mathbf{X}^T\boldsymbol{\beta})$. Secondly, a Gaussian linear model assumes that $Y|\mathbf{X} = \mathbf{x}$ is normally distributed with some mean, which is clearly not the case here because $Y > 0$ and Y is discrete.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem