Lecture 20: Generalized Linear Models

Module 6: part 2

Spring 2025

Logistics

- Continuing Module 6 on Generalized Linear Models

Recap

Logistic Regression

- Consider binary dependent variable which only takes values 0 or 1
- We want to see how certain covariates are associated with dependent variable
- Naively regressing Y onto X doesn't quite work

Logistic Regression

- No longer using the model:

$$Y_i = b_0 + \sum_k b_k x_{i,k} + \varepsilon_i$$

- Consider modeling the expected value of Y_i given $\mathbf{X_i}$

$$E(Y_i \mid \mathbf{X_i}) = b_0 + \sum_k b_k x_{i,k}$$

- Since Y_i is either 0 or 1

$$E(Y_i \mid \mathbf{X_i}) = P(Y_i = 1 \mid \mathbf{X_i}) = \theta(\mathbf{X_i})$$

  where $\theta(\mathbf{X_i})$ is the probability of "success" when the covariates are $\mathbf{X_i}$

- Unfortunately, $b_0 + \sum_k b_k x_{i,k}$ can be arbitrarily large or small, but $\theta(\mathbf{X_i})$ must be between 0 and 1

Logistic Regression

The proposed solution is logistic regression where we assume that

$$\theta(\mathbf{X_i}) = \underbrace{\frac{\exp(b_0 + \sum_k b_k x_{i,k})}{1 + \exp(b_0 + \sum_k b_k x_{i,k})}}_{\text{Sigmoid function}} \tag{1}$$

which is equivalent to

$$\underbrace{\log\left(\frac{\theta(\mathbf{X_i})}{1 - \theta(\mathbf{X_i})}\right)}_{\text{logit function}} = b_0 + \sum_k b_k x_{i,k} \tag{2}$$

Interpretation of logistic regression

- The quantity $\theta/(1-\theta)$ is known as the odds
- Can always map the odds back to the probability

$$P(\text{Success}) = \frac{\text{odds}}{1+\text{odds}}$$

- For a given coefficient b_k, we would say:
- Positive coefficient b_k means that larger values of X_k are associated with larger odds (and probability)
- Negative coefficient b_k means that larger values of X_k are associated with smaller odds (and probability)

Interpretation Example

Suppose two observations have all the same covariate values except differ in X_k by one unit. Then, the odds for the observation with the larger value of X_k would be $\exp(b_k)$ *times the odds for the observation with the smaller value of X_k.*

NFL Field Goals

We use logistic regression to model the log odds of a successful kick as a linear function of

- Distance (yards)
- Wind Speed (mph)
- Raining = 1, Dry = 0

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	6.8185	0.3823	17.84	0.0000
Distance	-0.1174	0.0079	-14.91	0.0000
Wind Speed	-0.0355	0.0128	-2.77	0.0056
Rain	-0.4385	0.2613	-1.68	0.0933

NFL Field Goals

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	6.8185	0.3823	17.84	0.0000
Distance	-0.1174	0.0079	-14.91	0.0000
Wind Speed	-0.0355	0.0128	-2.77	0.0056
Rain	-0.4385	0.2613	-1.68	0.0933

- Considering two attempts with the same rain and wind conditions, the odds of a successful attempt of a kick are $\exp(-.1174) = .889$ of the odds of a kick which is 1 yard longer
- Considering two attempts with the same distance and wind speed, when it is raining, the odds of a successful attempt are $\exp(-.439) = .644$ of the odds when it is not raining

NFL Field Goals

If a kick is from **35 yards**, the wind speed is 10 mph, and it is not raining, then we estimate that

$$\log\left(\frac{\theta(\mathbf{x})}{1-\theta(\mathbf{x})}\right) = 6.819 - .117 \times (35) - .036 \times (10) - .439(0) = 2.364$$

$$\frac{\theta(\mathbf{x})}{1-\theta(\mathbf{x})} = \exp(2.364) = 10.6334$$

$$P(Success) = \theta(\mathbf{x}) = \frac{\exp(2.364)}{1+\exp(2.364)} = .914$$

If a kick is from **36 yards**, the wind speed is 10 mph, and it is not raining, then we estimate that

$$\log\left(\frac{\theta(\mathbf{x})}{1-\theta(\mathbf{x})}\right) = 6.819 - .117 \times (36) - .036 \times (10) - .439(0) = 2.247$$

$$\frac{\theta(\mathbf{x})}{1-\theta(\mathbf{x})} = \exp(2.247) = 9.459 = \exp(2.364) \times .889$$

$$P(Success) = \theta(\mathbf{x}) = \frac{\exp(2.247)}{1+\exp(2.247)} = 0.904$$

NFL Field Goals

If a kick is from 35 yards, the wind speed is 10 mph, and it **is not** raining, then we estimate that

$$\log\left(\frac{\theta(\mathbf{x})}{1-\theta(\mathbf{x})}\right) = 6.819 - .117 \times (35) - .036 \times (10) - .439(0) = 2.364$$

$$\frac{\theta(\mathbf{x})}{1-\theta(\mathbf{x})} = \exp(2.364) = 10.6334$$

$$P(Success) = \theta(\mathbf{x}) = \frac{\exp(2.364)}{1+\exp(2.364)} = .914$$

If a kick is from 35 yards, the wind speed is 10 mph, and it **is** raining, then we estimate that

$$\log\left(\frac{\theta(\mathbf{x})}{1-\theta(\mathbf{x})}\right) = 6.819 - .117 \times (35) - .036 \times (10) - .439(1) = 1.935$$

$$\frac{\theta(\mathbf{x})}{1-\theta(\mathbf{x})} = \exp(1.935) = 6.855 = \exp(2.364) \times .644$$

$$P(Success) = \theta(\mathbf{x}) = \frac{\exp(1.935)}{1+\exp(1.935)} = .873$$

Generalized Linear Models

Generalized Linear Model

Logistic regression is a specific example of a **Generalized Linear Model** (GLM)

In general GLM's have 3 pieces

- **Distribution (family)**: What is the distribution of the dependent variable Y?

- **Link function**: The conditional mean satisfies for a link function g:

$$g(E(Y_i \mid \mathbf{X_i})) = a_i.$$

  This can be written:

$$g(\theta(\mathbf{X_i})) = a_i.$$

- **Linear model of covariates**: The "input" a_i is a linear function of some covariates

$$a_i = b_0 + \sum_k b_k x_{ik}$$

Logistic Regression

- **Distribution (family)**: Y_i follows a Bernoulli distribution
  - Y_i can be either 0 or 1
  - Mean: $E(Y \mid \mathbf{X_i}) = \theta(\mathbf{X_i})$ where θ is also the probability of success
  - Sometimes will only write θ instead of $\theta(\mathbf{X_i})$ for notational convenience
  - Variance: $\text{var}(Y) = \theta(1 - \theta)$
- **Link function**: We use the logit function:

$$\log\left(\frac{\theta}{1-\theta}\right) = a_i$$

- **Linear model of covariates**: The value $a_i = b_0 + \sum_k b_k x_{ik}$

Poisson distribution

Distribution (family): Y_i follows a Poisson distribution and can be any non-negative whole number: $0, 1, 2, \ldots$

- Poisson distribution can be used to model count data
- Examples:
  - Number of customers who enter a shop in a given hour
  - Number of birds which pass a sensor
  - Poisson can be thought of as a Binomial where number of trials is very very large, and probability of success is very very small
  - Each millisecond is a trial; A success is a bird passing the sensor in that given millisecond
- Poisson distribution has 1 parameter:
  - Mean parameter: $E(Y_i) = \theta$ (sometimes notation used is λ)
  - Variance: $\text{var}(Y_i) = \theta$ (so mean and variance are the same!)

Poisson distribution

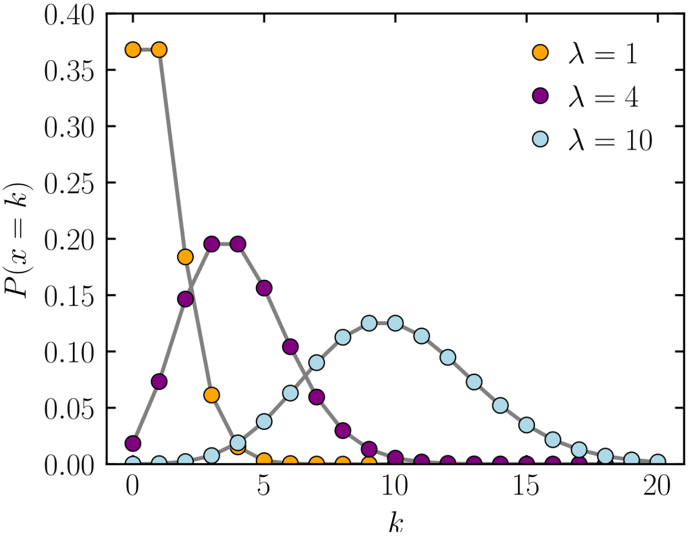

Figure: Poisson distribution (via wikipedia)

Poisson distribution

- **Distribution (family)**: Y_i follows a Poisson distribution and can be any non-negative whole number: $0, 1, 2, \ldots$
- **Link function**: is the log function

$$\log\left(E(Y_i \mid \mathbf{X_i})\right) = b_0 + \sum_k b_k x_{ik}$$

Poisson Regression

$$\log\left(E(Y_i \mid \mathbf{X_i})\right) = b_0 + \sum_k b_k x_{i,k}$$

Suppose two observations have all the same covariate values except differ in x_k by one unit with $x_{1,k} = x_{2,k} + 1$

$$\log\left(E(Y_1 \mid \mathbf{X_1})\right) - \log\left(E(Y_2 \mid \mathbf{X_2})\right) = b_k(x_{1,k} - x_{2,k}) = b_k$$

Since,

$$\log\left(E(Y_1 \mid \mathbf{X_1})\right) - \log\left(E(Y_2 \mid \mathbf{X_2})\right) = \log\left(\frac{E(Y_1 \mid \mathbf{X_1})}{E(Y_2 \mid \mathbf{X_2})}\right)$$

then, we also have

$$\frac{E(Y_1 \mid \mathbf{X_1})}{E(Y_2 \mid \mathbf{X_2})} = \exp(b_k)$$

Poisson Regression

$$\frac{E(Y_1 \mid \mathbf{X_1})}{E(Y_2 \mid \mathbf{X_2})} = \exp(b_k)$$

Suppose two observations have all the same covariate values except differ in x_k by one unit with $x_{1,k} = x_{2,k} + 1$, then the expected mean for observations with covariates $\mathbf{x_1}$ is $\exp(b_k)$ times the expected mean for observations with covariates $\mathbf{x_2}$

GLMs

GLMs are an entire semester worth of material, which we can't cover, but there's a lot more they can do:

- You can use GLMs to model various other distributions
- Categorical data: data which falls into more than 2 categories
  - Type of house: bungalow, cottage, ranch style, etc
- Ordinal data: categorical data which has a natural ordering
  - Likert scale: Very poor, poor, fair, good, very good
- Exponential/gamma data: continuous data that is only positive and skewed right
  - How long do I have to wait to be seated at a busy restaurant?
  - How long will a machine at a factory last before it needs to be repaired?

GLM vs transformation

At first, GLMs may seem similar to using a transformation of the dependent variable

When we use some function g to transform the dependent variable Y, we are fitting a model which assumed:

$$E\left(g(Y_i)\right) = b_0 + \sum_k b_k x_{ik}$$

When we are using a glm with link function g, we are fitting a model which assumes:

$$g\left(E(Y_i)\right) = b_0 + \sum_k b_k x_{ik}$$

and in general, if $g(a)$ is not a linear function of a (i.e., $g(a) = b_0 + b_1 a$), then

$$E\left(g(Y_i)\right) \neq g\left(E(Y_i)\right)$$

GLM vs transformation

Suppose Y_i is a Bernoulli random variable with probability of success $\theta = .5$ and suppose $g(a) = a^2$

Because Y_i is either 0 or 1, then $Y_i^2 = Y_i$ so

$$E\left(g(Y_i)\right) = E(Y_i^2) = E(Y_i) = \theta$$

However,

$$g\left((E(Y_i)\right) = g\left(\theta\right) = \theta^2$$

$$\theta \neq \theta^2$$

GLM vs transformation

Suppose Y_i is a Poisson random variable with mean $\theta = 2$ and suppose $g(a) = \log(a)$

Transforming Y_i may not make sense since Y_i can be 0, and $\log(0) = -\infty$

$$E\left(g(Y_i)\right) = E(Y_i^2) = E(Y_i) = \theta$$

However, since $\theta > 0$, then

$$g\left((E(Y_i)\right) = g\left(\theta\right) = \log(\theta)$$

still makes sense

GLM vs transformation

GLM's also typically specify how the conditional variance depends on the conditional mean

- Before, we assumed homoscedasticity so that the conditional variance does not depend on the covariates or the conditional mean (fitted value $\hat{y}_i$)
- Now, we assume that the conditional variance depends on the conditional mean $E(Y \mid \mathbf{X})$ in a specific way
  - Bernoulli: $\text{var}(Y_i \mid \mathbf{X_i}) = \theta(\mathbf{X_i})(1 - \theta(\mathbf{X_i}))$
  - Poisson: $\text{var}(Y_i \mid \mathbf{X_i}) = \theta(\mathbf{X_i})$
  - If the true variance is smaller/larger than what we specify, we would say that the data is underdispersed/overdispersed.

Recap

- GLMs are a framework for regression with broader types of data
- Requires specifying a distribution of the outcome and a link function which connects a parameter of the distribution to a linear function of the covariates
- Details on how to fit these models on Monday