Generalized Linear Models (1)

STA 211: The Mathematics of Regression

Yue Jiang

April 4, 2023

The following material was used by Yue Jiang during a live lecture.

Without the accompanying oral comments, the text is incomplete as a record of the presentation.

Review: The exponential family

The **exponential family** of probability distributions are those that can be expressed in a specific form. Suppose X is a random variable with a distribution that depends on (a) parameter(s) $\boldsymbol{\theta}$. A random variable is said to belong to the exponential family if it can be expressed as:

$$f(x|\boldsymbol{\theta}) = h(x) \exp\left(\eta(\boldsymbol{\theta})^T T(x) - \psi(\boldsymbol{\theta})\right),$$

The exponential family

$$f(x|\boldsymbol{\theta}) = h(x)\exp\left(\eta(\boldsymbol{\theta})^T T(x) - \psi(\boldsymbol{\theta})\right),$$

Note in the exponent the $\eta(\boldsymbol{\theta})^T T(x)$ term, which represents the summation $\sum_{i=1}^{k} \eta_i(\boldsymbol{\theta}) T_i(x)$.

In this expression, each $\eta_i(\boldsymbol{\theta})$ and $\psi(\boldsymbol{\theta})$ are real-valued functions of the parameter(s) $\boldsymbol{\theta}$, and each $T_i(x)$ and $h(x)$ are real-valued functions of the data.

If we have just a single parameter θ in the expression above, then we have a member of a **one-parameter exponential family** distribution, expressible as

$$f(x|\theta) = h(x)\exp\left(\eta(\theta) T(x) - \psi(\theta)\right).$$

The exponential family

For simplicity's sake, for now let's consider one-parameter exponential family distributions:

$$f(x|\theta) = h(x) \exp\left(\eta(\theta) T(x) - \psi(\theta)\right).$$

The binomial distribution

Suppose $X \sim Bin(n, p)$ where n is assumed known and we have a single parameter $0 < p < 1$. Then the probability $P(X = x)$ is given by:

$$
\begin{aligned}
f(x|p) &= \binom{n}{x} p^x (1-p)^{n-x} 1_{(x \in \{0,1,\cdots,n\})} \\
&= \binom{n}{x} \left(\frac{p}{1-p}\right)^x (1-p)^n 1_{(x \in \{0,1,\cdots,n\})} \\
&= \binom{n}{x} \exp\left\{ x \log\left(\frac{p}{1-p}\right) + n \log(1-p) \right\} 1_{(x \in \{0,1,\cdots,n\})}
\end{aligned}
$$

The binomial distribution

$$f(x|p) = \binom{n}{x} \exp\left\{x \log\left(\frac{p}{1-p}\right) + n \log(1-p)\right\} 1_{(x \in \{0,1,\cdots,n\})}$$

This is a member of the one-parameter exponential family, with

$$\eta(p) = \log\left(\frac{p}{1-p}\right)$$
$$T(x) = x$$
$$\psi(p) = -n \log(1-p)$$
$$h(x) = \binom{n}{x} 1_{(x \in \{0,1,\cdots,n\})}$$

Canonical form

$$f(x|\theta) = h(x) \exp\left(\eta(\theta)^T T(x) - \psi(\theta)\right),$$

Notice that η and ψ are both functions of θ.

For an invertible function $\eta(\cdot)$, suppose we define the variable $\eta = \eta(\theta)$ such that $\theta = \eta^{-1}(\eta)$ (sorry for using η as both the function and the variable).

Canonical form

We can thus re-write the exponential family in its **canonical form** using the η (the **canonical parameter(s)**):

$$f(x|\theta) = h(x) \exp\left(\eta(\theta)^T T(x) - \psi(\theta)\right)$$
$$f(x|\eta) = h(x) \exp\left(\eta^T T(x) - \psi(\eta^{-1}(\theta))\right)$$

Notice that here, the canonical parameter(s) are directly multiplied with the sufficient statistic(s), and the $\psi(\cdot)$ function is composed with $\eta^{-1}(\cdot)$ as it acts on the (untransformed) parameter θ.

The binomial distribution (again)

Suppose $X \sim Bin(n, p)$ where n is assumed known and we have a single parameter $0 < p < 1$. Then the probability $P(X = x)$ is in the exponential family with:

$$f(x|p) = \binom{n}{x} \exp\left\{x \log\left(\frac{p}{1-p}\right) + n \log(1-p)\right\} 1_{(x \in \{0,1,\cdots,n\})}$$

Taking $\eta = \log\left(\frac{p}{1-p}\right)$, then we have $1 - p = \frac{1}{1+e^{\eta}}$, and so in canonical form, the binomial distribution is expressed as

$$f(x|p) = \binom{n}{x} \exp\left\{\eta x - n \log(1 + e^{\eta})\right\} 1_{(x \in \{0,1,\cdots,n\})}$$

Canonical form

$$f(x|p) = \binom{n}{x} \exp\left\{\eta x - n\log(1+e^{\eta})\right\} 1_{(x \in \{0,1,\cdots,n\})}$$

In canonical form, we have

$$\eta = \log\left(\frac{p}{1-p}\right)$$
$$T(x) = x$$
$$A(\eta) = n\log(1+e^{\eta})$$
$$h(x) = \binom{n}{x} 1_{(x \in \{0,1,\cdots,n\})}$$

Generalized linear models

A **generalized linear model** has three components:

1. An outcome Y that follows a distribution from the exponential family*
2. The linear predictor $\mathbf{X}\beta$
3. A link function g that links the conditional expectation of Y with the linear predictor:

$$E(Y|\mathbf{X}) = g^{-1}(\mathbf{X}\beta)$$

Dispersion parameters

We often incorporate a **dispersion parameter** when thinking of the exponential family distributions in GLMs:

$$f(x|\theta) = h(x) \exp\left(\eta(\theta) T(x) - \psi(\theta)\right).$$

The dispersion parameter often gets at notions of "variance" - we now essentially have a two-parameter exponential family (with θ "corresponding" to some notion of mean, and ϕ for variance)

$$f(x|\theta, \phi) = h(x, \phi) \exp\left(\frac{\eta(\theta) T(x) - \psi(\theta)}{c(\phi)}\right).$$

Generalized linear models

Generalized linear models:

$$E(Y|\mathbf{X}) = g^{-1}(\mathbf{X}\beta)$$

Keep in mind that for exponential distributions in **canonical form**, the parameter relates to the sufficient statistic through the identity function. Furthermore, the mean can be found by differentiating the log-partition with respect to the canonical parameter.

The link function

Generalized linear models:

$$E(Y|\mathbf{X}) = g^{-1}(\mathbf{X}\beta)$$

The **link function** relates the linear predictor to the conditional expectation of the response (the "mean" - how convenient!).

There are any number of link functions that might be used. One important link function in particular is the **canonical link function** which directly relates the canonical parameter to the linear predictor.

Binary regression

Let's formulate binary regression as a generalized linear model:

1. The outcome Y follows a Bernoulli distribution (this is in the exponential family - just a binomial with fixed $n = 1$)
2. We'll assume the functional form of the predictors is a linear combination
3. We'll use a link function g to specifically relate the conditional mean of Y with the predictors

$$E(Y|\mathbf{X}) = g^{-1}(\mathbf{X}\beta)$$

- What might be some candidates for $g(\cdot)$ (it must be invertible; we generally like smooth, monotonic functions)?
- What properties or interpretations might different choices of $g(\cdot)$ provide us?

Canonical form

$$f(y|p) = \exp\left\{\eta y - n\log(1+e^{\eta})\right\} 1_{(y\in\{0,1\})}$$

In canonical form, we have

$$\begin{aligned}\eta &= \log\left(\frac{p}{1-p}\right)\\ T(y) &= y\\ A(\eta) &= n\log(1+e^{\eta})\\ h(y) &= 1_{(y\in\{0,1\})}\end{aligned}$$

Note that $p = E(Y)$ for a Bernoulli distributed Y.

Canonical link function

For an inverse logit link function

$$g^{-1}(\mathbf{X}\beta) = \frac{\exp(\mathbf{X}\beta)}{1 + \exp(\mathbf{X}\beta)}$$

We have

$$E(Y|\mathbf{X}) = g^{-1}(\mathbf{X}\beta)$$
$$E(Y|\mathbf{X}) = \frac{\exp(\mathbf{X}\beta)}{1 + \exp(\mathbf{X}\beta)}$$
$$\log\left(\frac{E(Y|\mathbf{X})}{1 - E(Y|\mathbf{X})}\right) = \mathbf{X}\beta$$

which is exactly the logistic regression model. Note the form of the link function, which is the canonical parameter of the Bernoulli distribution.

Logistic regression

1. The outcome Y follows a Bernoulli distribution (this is in the exponential family - just a binomial with fixed $n = 1$)
2. We'll assume the functional form of the predictors is a linear combination
3. We use the canonical link function $logit(\cdot)$ to link the conditional mean of Y with the predictors

$$g(E(Y|\mathbf{X}\beta)) = \mathbf{X}\beta$$
$$\log\left(\frac{E(Y|\mathbf{X})}{1 - E(Y|\mathbf{X})}\right) = \mathbf{X}\beta$$

Probit regression

1. The outcome Y follows a Bernoulli distribution (this is in the exponential family - just a binomial with fixed $n = 1$)
2. We'll assume the functional form of the predictors is a linear combination
3. We can use the non-canonical link function Φ^{-1} (inverse of the normal cdf) to link the conditional mean of Y with the predictors

$$g(E(Y|\mathbf{X}\beta)) = \mathbf{X}\beta$$
$$\Phi^{-1}(E(Y|\mathbf{X}\beta)) = \mathbf{X}\beta$$

A linear probability model

1. The outcome Y follows a Bernoulli distribution (this is in the exponential family - just a binomial with fixed $n = 1$)
2. We'll assume the functional form of the predictors is a linear combination
3. We can use the non-canonical identity link function to link the conditional mean of Y with the predictors

$$g(E(Y|\mathbf{X}\beta)) = \mathbf{X}\beta$$
$$E(Y|\mathbf{X}\beta) = \mathbf{X}\beta$$

Why canonical link functions?

Honestly, often just computational convenience.

For instance, the sufficient statistic under canonical link is $\mathbf{X}^T Y$. We also get nice properties inherited from exponential families under canonical link (such as easily finding MLEs). Various computational algorithms also coincide (e.g., Newton-Raphson and Fisher Scoring are equivalent - we'll talk about this next time).

But there are also reasons for using non-canonical links (even some computational)! For instance, you might want the interpretation associated with a non-canonical link, or have a specific computational reason (e.g., Bayesian logistic regression and Gibbs samplers with probit link on normally-distributed priors).

Homework (page 1 of 2)

1. Consider the linear regression model under the normality assumption (and constant variance). Is this a GLM? If so, identify the three components needed and specifically identify whether the link function is canonical. If not, explain why not.
2. Suppose we're trying to model the number of cancer cases per month (Y) in a city, conditionally on various demographic and exposure factors. Consider the log-linear regression model $\log(E(Y|\mathbf{X})) = \mathbf{X}\beta$, where Y takes on a Poisson distribution with parameter λ. Is this a GLM? If so, identify the three components needed (including specifics regarding the exponential family) and specifically identify whether the link function is canonical. If not, explain why not.

Homework (page 2 of 2)

3 Suppose we're trying to model the waiting time until the next bus arrives (Y), conditionally on weather conditions and traffic. Consider the log-linear regression model $\log(E(Y|\mathbf{X})) = \mathbf{X}\beta$, where Y takes on an Exponential distribution with parameter λ. Is this a GLM? If so, identify the three components needed (including specifics regarding the exponential family) and specifically identify whether the link function is canonical. If not, explain why not.