Chapter 5

Generalized Linear Models (GLMs): A Unifying Theory

Learning Objectives

- Determine if a probability distribution can be expressed in one-parameter exponential family form.
- Identify canonical links for distributions of one parameter exponential family form.

5.1 One parameter exponential families

Thus far, we have expanded our repertoire of models from OLS to include Poisson regression. But in the early 1970s Nelder and Wedderburn identified a broader class of models that generalizes the multiple linear regression we considered in the introductory chapter and are referred to as generalized linear models (GLMs). GLMs have similar forms for the likelihoods, MLEs, and variance. This makes it easier to find model estimates and their corresponding uncertainty. To determine whether a model is a GLM, we consider the following properties. When a probability formula can be written in the form below

\[ f(y; \theta) = \exp(a(y)b(\theta) + c(\theta) + d(y)) \]

and if the support (the set of possible values) does not depend upon \( \theta \), it is said to have a one-parameter exponential family form. We demonstrate that the Poisson distribution is a member of the one parameter exponential family by writing its probability mass function (pmf) in the form of Equation 5.1 and assessing its support.

5.1.1 One Parameter Exponential Family: Possion

Recall we begin with

\[ P(Y = y) = \frac{e^{-\lambda} \lambda^y}{y!} \text{ where } y = 0, 1, 2, \ldots \infty \]

A useful identity when attempting to establish exponential form is:

\[ \lambda^y = e^{\log(\lambda)} \]

Determining whether the Poisson model is a member of the one-parameter exponential family is a matter of writing the Poisson pmf in the form of Equation 5.1 and checking that the support does not depend upon \( \lambda \). First, consider the condition concerning the support of the distribution. The set of possible values for any Poisson random variable is \( y = 0, 1, 2, \ldots \infty \) which does not depend on \( \lambda \). The support condition is met. Now we see if we can rewrite the pmf in one-parameter exponential family form.

\[ P(Y = y) = e^{-\lambda} e^{\log(\lambda)} e^{-\log(y!)} \]
The exponent for Equation 5.1 must be the product of two factors, one solely a function of \(y\), \(a(y)\) and another, \(b(\lambda)\) a function of \(\lambda\) only. We must check the other two terms in the exponent to confirm that the Poisson distribution is in fact from the one-parameter exponential family. The middle term in the exponent must be a function of \(\lambda\) only, no \(y\)'s should appear. The last term has only \(y\)'s and no \(\lambda\). Since this appears to be the case here we can identify the different functions in this form.

\[
a(y) = y \tag{5.3}
\]
\[
b(\lambda) = \log(\lambda) \tag{5.4}
\]
\[
c(\lambda) = -\lambda
\]
\[
d(y) = -\log(y!) \tag{5.5}
\]

These functions have useful interpretations in statistical theory. We won’t be going in to this in detail, but we will note that function \(b(\lambda)\) or more generally \(b(\theta)\) will be particularly helpful in GLMs. The function \(b(\theta)\) is referred to as the \textbf{canonical link}. The canonical link is often a good choice to model as a linear function of the explanatory variables. In fact, there is a distinct advantage modeling the canonical link as opposed to other functions of \(\theta\), but it is also worth noting that other choices are possible and at times preferred depending upon the context of the application.

We’ll find that other distributions are members of the one parameter exponential family by writing their pdf or pmf in this manner and verifying the support condition. So for example, we’ll see that the binomial distribution meets these conditions so it is a member of the one parameter exponential family. The normal distribution is a special case. We note that we have two parameters, a mean and standard deviation. If it is possible, however, to assume that the standard deviation is known, we can show that a normal random variable is from a one parameter exponential family.

\subsection{One parameter exponential family: Normal}

Here we determine whether a normal distribution is a one parameter exponential family member. First we will need to assume that \(\sigma\) is known. Next, possible values for a normal random variable range from \(-\infty\) to \(\infty\), so we have no problems with the support. Finally, we’ll need to write the probability density function (pdf) in the one parameter exponential family form. We start with the familiar form:

\[
f(y) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}
\]

Even writing \(1/\sigma \sqrt{2\pi}\) as \(e^{-\log(\sigma) - \log(2\pi)/2}\) we still do not have the pdf written in one parameter exponential family form. We will first need to expand the exponent so that we have

\[
f(y) = e^{-\log(\sigma) - \log(2\pi)/2} e^{-\frac{y^2 - 2\mu y + \mu^2}{2\sigma^2}}
\]

Assuming \(\sigma\) is known, we have

\[
f(y) \propto e^{-2\mu y + \mu^2 + y^2}
\]

From this result, we can see that the canonical link for a normal response is \(\mu\) which is consistent with what we’ve been doing with OLS. The simple linear regression model has the form:

\[
\mu_{Y|X} = \hat{\beta}_0 + \hat{\beta}_1 X.
\]
There are multiple benefits of identifying a response as being from a one parameter exponential family. Wedderburn’s observation has unearthed a unifying theory for regression modeling. By identifying the one parameter exponential family form, we can identify the function of the parameter (the canonical link) that can be modeled by a linear combination of the predictors. We can also use the one parameter exponential family form to determine the expected value, \( E(Y|X) \) and standard deviation of \( Y|X \). With statistical theory you can show that

\[
\mu = \frac{c'(\theta)}{b'(\theta)} \quad \text{and} \quad \sigma^2 = \frac{b''(\theta)c'\theta - c''\theta b'\theta}{[b'(\theta)]^3}
\]

where differentiation is with respect to \( \theta \).

Verifying these results for the Poisson response:

\[
\mu = -\frac{1}{\lambda} \quad \text{and} \quad \sigma^2 = \frac{1}{\lambda^2} \quad \frac{1}{(1/\lambda)^3}
\]

Verifying this result for the normal distribution is left as an exercise.

**Useful Facts for Writing Exponential Family Forms**

\[ a = e^{log(a)} \]

\[ a^x = e^{xlog(a)} \]

\[ log(ab) = log(a) + log(b) \]

\[ log\frac{a}{b} = log(a) - log(b) \]

### 5.2 Generalized Linear Modeling

GLM theory suggests that the canonical link can be modeled as a linear combination of the explanatory variable(s). This approach unifies a number of modeling results used throughout the text. One example is that likelihoods will be used to compare models in the same way for parameter exponential family members.

We have now generalized our modeling to handle non-normal responses. For example, in addition to normally distributed responses, we are able to handle Poisson responses, binomial responses, and more. Writing a pmf or pdf for a response in one parameter exponential family form reveals the canonical link which can be modeled as a linear function of the predictors. This linear function of the predictors is the last piece of the puzzle for performing generalized linear modeling. But, in fact, it is really nothing new. We have used linear combinations when modeling normally distributed data.
Table 5.1: One parameter exponential family form and canonical links.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>One-parameter Exponential Family Form</th>
<th>Canonical Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binary Binomial</td>
<td>$P(Y = y) = e^{y \log \lambda - y}$</td>
<td>logit(p)</td>
</tr>
<tr>
<td>Poisson</td>
<td>$f(y) = e^{-y \mu + \mu^2 + y^2}$</td>
<td>log(\lambda)</td>
</tr>
<tr>
<td>Normal</td>
<td></td>
<td>$\mu$</td>
</tr>
<tr>
<td>Exponential</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gamma</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geometric</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Three Components of a GLM

1. Distribution of Y (e.g., Poisson)
2. Link Function (a function of the parameter, e.g., $\log(\lambda)$ for Poisson)
3. Linear Predictor (choice of predictors, e.g., $\beta_1, \beta_2$)

In the chapter on Poisson modeling, we provided heuristic rationale for using the log() function as our link. That is, counts would be non-negative but a linear function inevitably goes negative. By taking the logarithm of our parameter $\lambda$ we could use a linear predictor and not worry that it can take on negative values. Now we have theoretical justification for this choice, the log is the canonical link for poisson data. In the next chapter we encounter yet another type of response, a binary response, which calls for a different link function. Our work here suggests that we will model a logit(p) using a linear predictor.

Completing Table ?? is left as an exercise.

Note that generalized linear models (GLMs) differs from General Linear Models.$^1$

5.3 Exercises

1. For each distribution,
   - Write the pdf in one parameter exponential form, if possible.
   - Describe an example of a setting where this random variable might be used.
   - Identify the canonical link function, and
   - Compute $\mu = -\frac{\psi'(\theta)}{\psi''(\theta)}$ and $\sigma^2 = \frac{\psi''(\theta)\psi'(\theta) - \psi''(\theta)\psi'(\theta)}{[\psi'(\theta)]^3}$

(a) Binary
   Y = 1 for a success, 0 for a failure

   
   $p(y) = p^y(1 - p)^{1-y}$

---

$^1$ The general linear model is a statistical linear model with multivariate vectors as responses. For example, each subject in a study may have their height, weight, and shoe size recorded and modeled as a function of age and sex. The response is a vector, $Y = (height, weight, shoe size)$, for each study participant. Age and sex are explanatory variables in the model. The residual is usually assumed to follow a multivariate normal distribution. If the residual is not a multivariate normal distribution, generalized linear models may be used to relax assumptions about $Y$ and $U$. 

(b) Binomial
   \( Y = \text{number of successes in } n \text{ independent, identical trials} \)
   \[
p(y) = \binom{n}{y} p^y (1-p)^{1-y}
   \]

(c) Poisson
   \( Y = \text{number of events occurring in a given time (or space) when the average event rate is } \lambda \text{ per time (or space)} \)
   \[
P(Y = y) = \frac{e^{-\lambda} \lambda^y}{y!}
   \]

(d) Normal
   \[
f(y) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(y-\mu)^2}{2\sigma^2}}
   \]

(e) Exponential
   \( Y = \text{time spent waiting for the first event for a process with an average rate of waiting } 1/\lambda \text{ per event} \)
   \[
f(y) = \lambda e^{-\lambda y}
   \]

(f) Gamma
   \( Y = \text{time spent waiting for the } r^{th} \text{ event for a process with an average rate of waiting } 1/\lambda \text{ per event} \)
   \[
f(y; \lambda) = \frac{\lambda^r y^{(r-1)} e^{-\lambda y}}{\Gamma(r)}
   \]

(g) Geometric
   \( Y = \text{number of trials up to and including the first success} \)
   \[
p(y) = (1-p)^{(y-1)} p
   \]

(h) Negative Binomial (for fixed \( r \))
   \( Y = \text{number of trials up to and including the } r^{th} \text{ success} \)
   \[
p(y) = \binom{y-1}{r-1} (1-p)^{(y-r)} p^r
   \]

(i) Pareto
   \[
f(y) = \frac{\theta k^\theta}{y^{(\theta+1)}}
   \]

\[y \geq k; \theta \geq 1\]

2. Complete Table 5.1 containing your results of the preceding exercises.

5.4 References cited
