Link Functions: Where do they come from?

Danny

3/30/2020

As a student presented with the concept of link functions I was always left wanting a bit more. The standard way a social science student usually learns about link functions is through the introduction of logistic regression. In this note, I will attempt to provide an earlier version of myself what I always wanted - a step-by-step break down of how the logit link function is derived.

Of note, we'll be dealing with link functions in specific scenarios. First, we'll be dealing with the canonical link function, only. For instance, the logit is the cannonical link to the binomial whereas the probit is not. Roughly speaking, this means that the logit link can be directly derived from the binomial whereas the probit link cannot. Ironically, early on, probit regression was more widely used than logistic regression.

We'll be using a lot of algebra - there will be a hint of calculus. If this isn't your cup of tea, perhaps just try to read the step. But, I try to explain all the steps below in a way that I hope is clear, including the "rules" I'm using through each step - so give it a shot! Usually mathematical derivations skip steps which makes them actually look less daunting. I tried to show every single step to make it obvious. I think a little effort doing algebra can serve most of us students in social science more than we realize.

Preliminaries:

I'll refer to these rules in the post below.

1. When I use log I mean the natural log, or ln . This is the logarithm with base e such that $log_e(1) = 0$ because $e^0 = 1$.
2. $log(a) + log(b) = log(a * b)$.
3. $log(a) - log(b) = log(\frac{a}{b})$.
4. $log(a)^x = x log(a)$.

Exponential Family of Probability Functions

A generalized linear model's link function takes the "structural" or linear component of the model and links it to the outcome, the expectation of y given x (sometimes, this is written, y-hat) but you can also think of it as μ. To expand, even when you have a standard linear regression of the form,

$$y_i = \alpha + \beta_1 X_{1i} + \epsilon_i$$

y has some conditional distribution, an error distribution. In other words, we need to have some way to go from the structural portion of the model,

$$\alpha + \beta_1 X_1$$

, which is just a straight line, and link it to its outcome y and have some probability density function for describing or modeling the data, since we are not working with deterministic data, afterall. In this case, we have an ϵ term that "adds" noise and we assume that it's normally distibuted for each value of X. So Y, for a given value of x, is modeled with a normal distribution. In other words,

$$Y = N(\mu = \alpha + B_1 * X_1, \sigma^2)$$

or in matrix notation:

$$Y = N(\mu = XB, \sigma^2 I)$$

where I is the identity matrix.

In this case, the Gaussian/normal, the link function is the idenity - or, said another way, the identify link takes $E[Y|X]$ and outputs the same thing - it doesn't change it - $g(E[Y|X]) = E[Y|X]$.

When the outcome can only take a value of 1 or 0, the normal distribution is not a grear distribution to use, really - it best models data that are continuous (among many other properties). So, the identity link won't do. We need to see if we can use some other function to transform the expectation such that we can still use a linear model to model the data.

Welcome to link functions for the generalized linear model.

Family of Exponential Distributions

This is where the bulk of the hard work will be. It turns out that probability density functions (pdf) like the normal distribution or discrete probability distributions (probability mass functions, or, pmf) like the binomial are of a family of distributions called the exponential family of distributions. In the case of the binomial, it is in the exponential family only when there are a fixed number of trials.

What's special about this is that members of this family can be written in the general form,

$$f_X(x|\theta) = h(x)exp[\eta(\theta) * T(x) - A(\theta)] \tag{A}$$

For whatever reason, I like an equivalent form of this distribution a bit better:

$$f(x, \theta, \phi) = exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right] \tag{B}$$

Here, y is your outcome (or number of successes in the case of a binomial), θ is a parameter, $b(\theta)$ is a function of θ, $a(\phi)$ is a function of parameter ϕ, and `c()` is a function or some set of functions of data and ϕ.

The logit link and logistic regression

The binomial distribution looks like:

$$f(y) = \binom{n}{y} p^y (1-p)^{n-y} \tag{C}$$

This looks nothing like `equation B`. Remember from `equation A,` above, that a `GLM` gives us the expectation of y given x. In linear regression, the expectation of y given an x value is simply, $\hat{y} = \alpha + B1_X 1$

We now have a dichotomous outcome which simply can't be normally distributed given each value of `X` (aka, it can't have normally distributed errors). This will make more sense if we can link the structural portion of the model to the parameter of the binomial, which is, in the case of 1 trial, simply a bernoulli distribution with expecation, p, or, generally, with n trials, $n * p$ where p is the probability of success for a trial. In this way, we're saying our outcome is the result of bernoulli trials. Regardless, we have to transform to find the link function.

Transforming the Binomial into Exponential Form

Steps 1-4

1. The first step is taking the log of both sides of `equation C`, because, from **preliminary point 2**, this will take the multiplication and turn it into addition which we need to do to get closer to `B`.
2. Using **prelimiary point 4**, we factor and rearrange, and place the exponents in front of the log fuctions (for multiplication).
3. To get rid of the log on the left side, we exponentiate with e both sides (using $exp(a)$ to mean e^a).
4. Re-write so f(y) back to normal. Proceed to step 5. The algebra is below.

$$log(f(y)) = log\left[\binom{n}{y}\right] + log(p)^y + log(1-p)^{n-y} \tag{1}$$

$$log(f(y)) = log\left[\binom{n}{y}\right] + ylog(p) + (n-y)log(1-p). \tag{2}$$

$$exp(log(f(y))) = exp\left[log\left(\binom{n}{y}\right) + ylog(p) + (n-y)log(1-p)\right] \tag{3}$$

$$f(y) = exp\left[log(\binom{n}{y}) + ylog(p) + (n-y)log(1-p)\right] \tag{4}$$

Steps 5 - 7

To save on typing, we'll only write the term inside the `exp[ ]` term now.

5. We'll expand out $(n-y)(log(1-p))$ in step 5.

6. Rewrite with step 5 expansion inserted back in. To make things simpler, we'll group our outcome, y, our parameter of interest p, and terms with the full data in it (n).

7. From **preliminary 3**, since first two terms have the same log base and multiplicative constant we can regroup such that $ylog(p) - ylog(1-p)$ becomes $y\left[log(\frac{p}{1-p})\right]$.

$$nlog(1-p) - ylog(1-p) \tag{5}$$

$$= ylog(p) - ylog(1-p) + nlog(1-p) + log\left(\binom{n}{y}\right) \tag{6}$$

$$= y\left[log(\frac{p}{1-p})\right] + nlog(1-p) + log\binom{n}{y} \tag{7}$$

3

Steps 8 - 10

This is starting to look just like we want, and if you've used logistic regression, there should be some familiar terms.

Remember that our exponential form from equation B:

$$f(x, \theta, \phi) = exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right] \tag{B}$$

It appears the $y\theta$ portion matches up with:

$$y\left[log(\frac{p}{1-p})\right]$$

$$\theta = \left[log(\frac{p}{1-p})\right] \tag{8}$$

From equation 2, we see that the second term is also a function of theta, so we need to rewrite $nlog(1-p)$ in terms of θ. So, the easiest way to do this is to solve for θ in terms of p.

Exponentiate both sides of 8 then simply solve for θ by: A. multiplying both sides by $1 - p$, expanding/multiplying out the left side, adding $exp(\theta) * p$ to both sides

B. factoring out p, the common term from the right side,

C. Divide both sides by $(1 + exp(\theta))$

$$1 - p = \frac{1}{1 + exp(\theta)} = (1 + exp(\theta))^{-1}$$

$$exp(\theta) = \frac{p}{1 - p} \tag{9a}$$

$$exp(\theta)(1 - p) = p \tag{1}$$

$$exp(\theta) - exp(\theta) * (p) = p \tag{2}$$

$$exp(\theta) = p + exp(\theta) * p \tag{3}$$

Factor:

$$exp(\theta) = p(1 + exp(\theta)) \tag{9b}$$

$$\frac{exp(\theta)}{1 + exp(\theta)} = p \tag{9c}$$

$$1 = \frac{1 + exp(\theta)}{1 + exp(\theta)} \tag{4}$$

$$1 - p = \frac{1}{1 + exp(\theta)} \tag{9d}$$

4

After that fair amount of algebra, let's rewrite

$$1 - p$$

as

$$(1 + exp(\theta))^{-1}$$

because a negative exponent denotes a fraction. This step is really important because of **prelimiary 4**. Thus, inputting $nlog(1 + exp(\theta))^{-1}$ back in -

We can rewrite, now:

10a. Input $nlog(1 + exp(\theta))^{-1}$ in place of $nlog(1 - p)$.

10b. Input $-nlog(1 + exp(\theta))$ because $log(a^4) = 4log(a)$

$$f(x, \theta, \phi) = y(\theta) + nlog(1 + exp(\theta))^{-1} + log\binom{n}{y} \tag{10a}$$

$$f(x, \theta, \phi) = y(\theta) - nlog(1 + exp(\theta)) + log\binom{n}{y} \tag{10b}$$

Believe it or not, this is now in exponential form, where:

$$y(\theta) = ylog(\frac{p}{1 - p})$$

$$b(\theta) = nlog(1 + exp(\theta))$$

$$c(y, \phi) = log\binom{n}{y}$$

$$a(\phi) = 1$$

$$f(x, \theta, \phi) = exp([y(\theta) - nlog(1 + exp(\theta))] + log\binom{n}{y})$$

Finding Expectations - a little bit of calculus (optional):

It turns out, that the expectation of a distribution in the exponential form is the first derivative of $b(\theta)$ and the variance is $a(\phi)$ times the second derivative of $b(\theta)$ 1. $\mu = b'(\theta)$ 2. $\sigma^2 = b''(\theta) * a(\phi)$

Expectation/mean

$$= b'(\theta) = n\left[log(1 + exp(\theta))\right] \tag{5}$$

$$= n\left[exp(\theta) * \frac{1}{1 + exp(\theta)}\right] \tag{6}$$

$$= n\left[\frac{exp(\theta)}{1 + exp(\theta)}\right] \tag{7}$$

$$= n * p = np \tag{8}$$

This is great news. The first derivative worked out and we found the mean. This is what we need to pass through the link function, eventually.

Variance (optional)

Given the definition above, this is how I solved. I use the chain and product rule instead of the quotient rule. . . sue me.

$$a(\phi)b''(\theta) = 1 * n \left[\frac{exp(\theta)}{1 + exp(\theta)} \right] \tag{9}$$

$$= n \left[exp(\theta) * (1 + exp(\theta)^{-1}) \right] + \left[exp(\theta) * -1 * exp(\theta) * (exp(1 + \theta)^{-2}) \right] \tag{10}$$

$$= n \left[\frac{exp(\theta)}{1 + exp(theta)} - \frac{exp(2\theta)}{(exp(1 + \theta)^2)} \right] \tag{11}$$

$$= n \frac{exp(\theta)}{1 + exp(\theta)} (1 - exp(\theta)) \tag{12}$$

$$= npq \tag{13}$$

If you know the mean of the binomial, you'll realize this isn't super important to go through all this work, but it's good to see.

Linking μ and θ

So the key aspect of a link function is how we go from μ to θ. That is,

$$f(\mu) = \theta$$

is effectively the link function.

Finding the link function

If you skipped the calculus portion, all we need to know now is μ in terms of θ

$$\mu = n * p = n * \left(\frac{exp(\theta)}{1 + \exp(\theta)} \right)$$

We need to solve for θ. We'll start by multiplying both sides of the equation by $1 + exp(\theta)$ and then follow the same process as `step 9` but replace `p` with $n * exp(\theta)$

$$\mu(1 + exp(\theta)) = n * exp(\theta) \tag{14}$$

$$\mu + (\mu * \exp(\theta)) = n * exp(\theta) \tag{15}$$

$$\mu = n * exp(\theta) - \mu * exp(\theta) \tag{16}$$

$$\mu = exp(\theta)(n - \mu) \tag{17}$$

$$\frac{\mu}{n - \mu} = exp(\theta) \tag{18}$$

$$\log(\frac{\mu}{n - \mu}) = \theta \tag{19}$$

And that's your link function! But note, $\mu = n * p$ so we can rewrite it as:

$$\theta = log(\frac{n * p}{n - (n * p)})$$

Which, either factoring out n/n, or by substituting 1 for n as is the case for bernoulli trials like in scenarios we're used to with logistic regression:

$$\theta = log(\frac{1 * p}{1 - (1 * p))})$$

$$\theta = log(\frac{p}{1 - p})$$

And, now this looks awfully like logistic regression.