# Which Link Function—Logit, Probit, or Cloglog?

Kevin S. Van Horn

August 14, 2015

## 1 Introduction

A generalized linear model for binary response data has the form

$$\Pr\left(y = 1 \mid x\right) = g^{-1}\left(x'\beta\right)$$

where $y$ is the 0/1 response variable, $x$ is the $n$-vector of predictor variables, $\beta$ is the vector of regression coefficients, and $g$ is the link function. In the Stan[1] modeling language this would be written as

```
y ~ bernoulli(p);
g(p) <- dot_product(x, beta);
```

with $g$ replaced by the name of a link function, and similarly for the BUGS modeling language.

The most common choices for the link function are

- logit:
$$g(p) = \log\left(\frac{p}{1-p}\right);$$
- probit:
$$g^{-1}(\eta) = \Phi(\eta)$$
where $\Phi$ is the cumulative distribution function for the standard normal distribution; and
- complementary log-log (cloglog):
$$g(p) = \log\left(-\log\left(1-p\right)\right).$$

All three of these are strictly increasing, continuous functions with $g(0) = -\infty$ and $g(1) = +\infty$.

In this note we'll discuss when to use each of these link functions.

[1] *Stan Modeling Language Users Guide and Reference Manual*, http://mc-stan.org/

## 2 Probit

The probit link function is appropriate when it makes sense to think of $y$ as obtained by thresholding a normally distributed latent variable $z$:

$$\begin{aligned} z &= x'\beta^* + \varepsilon \\ \varepsilon &\sim \mathrm{Normal}\,(0,\sigma) \\ y &= \begin{cases} 1 & \text{if } z \geq 0 \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Defining $\beta = \beta^*/\sigma$, this yields

$$\begin{aligned} \Pr\,(y = 1 \mid x) &= \Pr\left(x'\beta^* + \varepsilon \geq 0\right) \\ &= \Pr\left(-\varepsilon \leq x'\beta^*\right) \\ &= \Pr\left(\varepsilon \leq x'\beta^*\right) \\ &= \Phi\left(x'\beta\right). \end{aligned}$$

## 3 Logit

Logit is the default link function to use when you have no specific reason to choose one of the others. There is a specific technical sense in which use of logit corresponds to minimal assumptions about the relationship between $y$ and $x$. Suppose that we describe the joint distribution for $x$ and $y$ by giving

- the marginal distribution for $x$, and
- the expected value of $x_i y$ for each predictor variable $x_i$.

Then the maximum-entropy (most spread-out, diffuse, least concentrated) joint distribution for $x$ and $y$ satisfying the above description has a pdf of form

$$p\,(x, y) = \frac{1}{Z} f(x) \exp\left(\sum_{i=1}^{n} \beta_i x_i y\right)$$

for some function $f$, coefficient vector $\beta$ and normalizing constant $Z$. The conditional distribution for $y$ is then

$$\begin{aligned} p\,(y \mid x) &= \frac{p(x, y)}{p(x, 0) + p(x, 1)} \\ &= \frac{\exp\left(\left(x'\beta\right) y\right)}{1 + \exp\left(x'\beta\right)} \end{aligned}$$

and so

$$\begin{aligned}
\Pr(y=1 \mid x) &= \frac{\exp(x'\beta)}{1+\exp(x'\beta)} \\
&= \operatorname{logit}^{-1}\left(x'\beta\right).
\end{aligned}$$

# 4 Cloglog

The complementary log-log link function arises when

$$y = \begin{cases} 1 & \text{if } z > 0 \\ 0 & \text{if } z = 0 \end{cases}$$

where $z$ is a count having a Poisson distribution:

$$\begin{aligned}
z &\sim \operatorname{Poisson}(\lambda) \\
\lambda &= \exp\left(x'\beta\right).
\end{aligned}$$

To see this, let

$$p = \Pr(z > 0 \mid x).$$

Then

$$\begin{aligned}
p &= 1 - \operatorname{Poisson}(0 \mid \lambda) \\
&= 1 - \exp(-\lambda) \\
&= 1 - \exp\left(-\exp\left(x'\beta\right)\right)
\end{aligned}$$

and so

$$\begin{aligned}
\operatorname{cloglog}(p) &= \log(-\log(1-p)) \\
&= \log\left(-\log\left(\exp\left(-\exp\left(x'\beta\right)\right)\right)\right) \\
&= x'\beta.
\end{aligned}$$

# 5 Conclusion

In summary, here is when to use each of the link functions:

- Use probit when you can think of $y$ as obtained by thresholding a normally distributed latent variable.
- Use cloglog when $y$ indicates whether a count is nonzero, and the count can be modeled with a Poisson distribution.
- Use logit if you have no specific reason to choose some other link function.