**Section 2.3 – Generalized linear models**

Logistic regression models fall within a family of models known as generalized linear models. Each generalized linear model (GLM) has three different components:

1. Random: This specifies the distribution for Y. For the logistic regression model, Y has a Bernoulli distribution.
2. Systematic: This specifies a linear combination of the regression parameters with the explanatory variables, and it is often referred to as the linear predictor. For the logistic regression model, we have β0 + β1x1 + + βpxp.
3. Link: This specifies how the expected value of the random component E(Y) is linked to the systematic component. For the logistic regression model, we have  where E(Y) = π and the logit transformation is the link function.

Comments:

* Note that “linear” in generalized linear models comes from the regression parameters simply being coefficients for the explanatory variables in the model. Nonlinear models involve more complex functional forms such as xβ.
* The GLM abbreviation is where the glm() function gets its name.

Other generalized linear models include:

* Normal linear regression model:

E(Y) = β0 + β1x1 + … + βpxp

Random: Y has a normal distribution

Systematic: β0 + β1x1 + + βpxp

Link: Identity

* Poisson regression model from Chapter 4:

log[E(Y)] = β0 + β1x1 + … + βpxp

Random: Y has a Poisson distribution

Systematic: β0 + β1x1 + + βpxp

Link: log

We will focus in this section on generalized linear models when the random component has Y with a Bernoulli distribution. In addition to the logistic regression model, two models sometimes used for binary regression are

* Probit regression model:

probit(π) = β0 + β1x1 + βpxp

where probit(π) represents the inverse transformation of the standard normal CDF. In other words, suppose Z is a standard normal random variable. Find the quantile such that the probability to the left of it is π. For example, P(Z < 1.96) is 0.975:

> qnorm(0.975)

[1] 1.959964

> pnorm(1.96)

[1] 0.9750021

Equivalently, Z0.975 = 1.96. This model can be re-expressed as

π = Φ(β0 + β1x1 + βpxp)

where Φ() is the standard normal CDF. In other words, if β0 + β1x1 + βpxp = 1.96, then π is 0.975.

The term “probit” is a shortened version of “probability unit” (Hubert, 1992). Chapter 8 of Salsburg (2001) gives an interesting account of how this model was developed by Chester Bliss.

* Complementary log-log model:

log[-log(1 – π)] = β0 + β1x1 + βpxp

Solving π, we obtain

π = 1 – exp[-exp(β0 + β1x1 + βpxp)]

Thus, the model's name comes from using 1 – π (which is like using P(Ac) = 1-P(A) where Ac is the complement of an event A) and a double log transformation.

Why are these link functions used? The *inverse* link function needs to be chosen to guarantee 0 < π < 1. For example, the inverse link function for logistic regression is the  transformation:



Inverse link functions used in practice are based on cumulative distribution functions (CDFs) for random variables. The CDF of a logistic probability distribution is used for logistic regression, which results in the name for the model.

CDF review:

Let X be a continuous random variable with probability density function f(x). An observed value of X is denoted by x. The CDF of X is F(x) = P(Xx) = . Note that u is substituted into the probability density function to avoid confusion with the upper limit of integration. If X is a discrete random variable, the CDF of X is F(x) = P(Xx) =  =  where the sum is over all values of X ≤ x.

An informal definition is the CDF “cumulates” probabilities as a function of x.

Example: Logistic probability distribution (Logistic.R)

A random variable X with a logistic probability distribution has a probability density function of



for -∞ < x < ∞ and parameters -∞ < μ < ∞ and σ > 0. The CDF is



Below is a plot of the PDF and CDF for μ = -2 and σ = 2.



> mu <- -2

> sigma <- 2

> #Examples for f(-2) and F(-2)

> dlogis(x = -2, location = mu, scale = sigma)

[1] 0.125

> plogis(q = -2, location = mu, scale = sigma)

[1] 0.5

> dev.new(width = 10, height = 6, pointsize = 12)

> par(mfrow = c(1,2))

> curve(expr = dlogis(x = x, location = mu, scale = sigma),

 ylab = "f(x)", xlab = "x", xlim = c(-15, 15), main =

 "PDF", col = "black",, n = 1000)

> abline(h = 0)

> curve(expr = plogis(q = x, location = mu, scale = sigma),

 ylab = "F(x)", xlab = "x", xlim = c(-15, 15), main =

 "CDF", col = "black", n = 1000)

We see that the distribution is symmetric and centered at μ. Larger values of σ lead to a larger variance. Overall, E(X) = μ and Var(X) = σ2π2/3. In comparison to a normal distribution with mean μ and variance σ2, the distributions are centered at the same location, but the logistic distribution has a larger variance.

We can equivalently work with  for a logistic distribution, similar to what is often done for a normal distribution. This leads to the CDF



Solving for z results in the inverse of this function:



This has the same mathematical form as the link function used with logistic regression. Thus, the link function is the inverse CDF for a logistic distribution.

Note that the previous CDF plot is actually the same plot as in our initial example of a logistic regression model!

Model comparisons

|  |  |  |  |
| --- | --- | --- | --- |
|  | Logistic | Probit | Complementary log-log |
| Link | Inverse CDF of standard logistic distribution | Inverse CDF of standard normal distribution | Inverse CDF of Gumbel (extreme value) distribution |
| Inverse link | CDF of standard logistic distribution | CDF of standard normal distribution | CDF of standard Gumbel (extreme value) distribution |
| glm() code | link = logit | link = probit | link = cloglog |
| Symmetry | Rotationally symmetric about π = 0.5 | Rotationally symmetric about π = 0.5 | Not rotationally symmetric |
| OR for g(π) = β0 + β1x | Not dependent on x | Dependent on x | Dependent on x |
| Use?  | Everywhere! | A few places | Very little |

where g() represents the appropriate transformation for the corresponding model.

The yellow highlighted cells of the table are why we will not discuss the probit and complementary log-log models much further.

Comments:

* Probit regression models generally provide a very similar fit as the logistic regression model. The one place where the fit can be a little different is close to π = 0 or 1, because the logistic distribution has fatter tails (more probability in the outer portions of the distribution) than the normal distribution.
* Below is the fit of the three binary regression models for the placekicking example when distance is the only explanatory variable. The code is in Placekick.R.



The estimated models are

* 
* 
* 

You may think there are large differences among these models due to their regression parameter estimates. However, examining the estimates alone does not take into account the functional form of the models. We can see from the plots that the models are VERY similar except at larger distances for the complementary log-log model when compared to the logistic and probit models.

* A logistic regression model does NOT assume Y has a logistic distribution! Simply, the model uses the CDF’s mathematical function form for its inverse link function.