☼ RNR 613 — Introduction to the Generalized Linear Model

A *Generalized Linear Model* is a probability model where the mean of a response variable (*Y*) is related to a set of explanatory variables (*X's*) by a regression equation; the regression structure is linear in its parameters:

$$\beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p.$$

A function *links* the mean response to the regression structure; this function is the *link function*:

$$g(\mu) = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p.$$

The appropriate link function depends on the distribution of the response variable and allows us to relate explanatory variables to a wide range of response variables.

For ordinary least-squares (OLS) regression, where the response is a continuous variable, the appropriate function is the *identity* link: $g(\mu) = \mu$.

Linear regression

When the response is "normal" (linear regression):

link function → identity

distribution function → normal distribution with constant variance

Binomial data

When the response is binary (values 0 and 1)

link function → logit $\qquad g(p) = \ln\left(\frac{p}{1-p}\right)$

interpretation of μ_i is the probability p of the response Y_i taking on the value one.

distribution function → binomial distribution

A GLM with the binomial distribution and logit link is called logistic regression

Count data

When the response is counts

distribution function → Poisson

link function → log

A GLM with the Poisson distribution and log link is called Poisson regression

Distribution	Default Link Function
binomial	logit
gamma	inverse (power(-1))
inverse Gaussian	inverse squared (power(-2))
multinomial	cumulative logit
negative binomial	log
normal	identity
Poisson	log

Maximum Likelihood

Least squares parameter estimation is replaced by *maximum likelihood* estimation.

If the specified model is correct and sample sizes are large enough, then ML estimators have good properties:

1. They are generally unbiased.
2. They are reasonably precise.
3. Formulas exist for estimating standard deviations of sampling distributions of the estimators (SEs).
4. The sampling distributions are approximately normal.

The aim of maximum likelihood estimation is to find parameter values that make the observed data most "likely."

Parameter estimates are chosen such that estimates have the highest probability of matching the actual observed outcomes.

Once data have been observed they are considered fixed, so there is no "probabilistic" part to them anymore.

The *likelihood* of specific values for model parameters, then, is the probability of the actual outcome, calculated with those parameter values.

In a sense, likelihood works backwards from probability:

> given *B*, we use the conditional probability P(*A* | *B*) to reason about *A*, and, given *A*, we use the likelihood function P(*A* | *B*) to reason about *B*.

To use likelihood methods to estimate parameters, if we define a probability density function $x \rightarrow f(x|\theta)$ where θ is the parameter, then the likelihood function is:

$$L(\theta|x) = f(x|\theta)$$

where *x* is the observed outcome of sampling or an experiment.

In other words, when $f(x|\theta)$ is viewed as a function of *x* with θ fixed, it is a probability density function, and when viewed as a function of θ with *x* fixed, it is a likelihood function.

If the probability of an event X that depends on model parameters *p* is

written as:	P (X | *p*)
then the likelihood as:	L (*p* | X)

that is, the likelihood of the *parameters given the data*.

Probability: Knowing parameters → Predicting outcome
Likelihood: Observing data → Estimating parameters

Example

Toss a coin 100 times and observe 56 heads and 44 tails.

Find the MLE for *p* by finding the value for *p* that makes the observed data most likely.

The observed data are now fixed and are treated as constants and plugged into a binomial probability model :

- n = 100 (total number of tosses)
- h = 56 (total number of heads)

Imagine that *p* was 0.5. Plugging this value into our probability model as follows:

$$L(P = 0.5|data) = \frac{100!}{56!44!}0.5^{56}0.5^{44} = 0.0389$$

But what if p was 0.52 instead?

$$L(P = 0.52|data) = \frac{100!}{56!44!}0.52^{56}0.48^{44} = 0.0581$$

So from this we conclude that *p* is more likely to be 0.52 than 0.5.

The likelihoods for different parameter values to find the MLE of *p*:

p	L
0.48	0.0222
0.50	0.0389
0.52	0.0581
0.54	0.0739
0.56	0.0801
0.58	0.0738
0.60	0.0576
0.62	0.0378

The full range of possible values for *p* gives the likelihood surface.
The best estimate for *p* from any one sample is the proportion of heads observed in that sample.
Similarly, the best estimate for the population mean is the sample mean.

Example

The value of likelihoods may be clearer in a more complex example, here with three parameters.

Donner Party — 40 of 87 people died from famine and exposure on the trip from Springfield, Illinois to California.

If we wish to assess the effects of age and sex on the probability of survival, we can calculate the likelihood under different values of these parameters.

Maximum likelihood estimation chooses as estimates the values that assign the highest probability to the observed outcome.

β_0	β_1(age)	β_2(sex)	log(likelihood)
1.50	–0.050	1.25	–27.7083
		1.80	–28.7931
	–0.80	1.25	–26.1531
		1.80	–25.7272

1.70	–0.050	1.25	–29.0467
		1.80	–30.3692
	–0.80	1.25	–25.7972
		1.80	–25.6904
1.63	–0.078	1.60	–25.6282

For this range of parameter values, the largest likelihood is for

β_0 = 1.63, β_1 = –0.078, β_2 = 1.60.