Introduction to Generalized Linear Models
Part 1 of 3

Dr Rafael de Andrade Moral
Associate Professor of Statistics, Maynooth University

rafael.deandrademoral@mu.ie
https://rafamoral.github.io

Outline

- The normal model: A recap
- Models for binary data
- Models for binomial data
- Models for multinomial data (generalized logit model)
- Models for count data (Poisson, offset, hnp –end of day 2)
- Extensions: overdispersion models (quasi-Poisson, Negbin, quasi-binomial, betabinomial)
- Extensions: zero-inflated models (ZIP, ZINB, hurdle –end of day 3)

The Normal Model: A Recap

The Normal Model: A Recap

- Y_i is a response variable associated with experimental or observational unit i
- We *assume* it comes from a certain probability distribution with pmf/pdf f and vector of parameters $\boldsymbol{\theta}$
- In general, one of the parameters in $\boldsymbol{\theta}$ is the mean of the distribution
- We also have predictors x_i we are interested in studying
- We may link it to a parameter of interest, typically the mean of the distribution

The Normal Model: A Recap

- For the normal model, we typically write for each observation y_i:

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi} + \varepsilon_i$$

  where $\varepsilon_i \sim \mathsf{N}(0, \sigma^2)$
- Each β coefficient represents the expected mean change in y for a 1-unit increase in its associated predictor
- We can show that, from the equation above, the expected value of Y_i is

$$\mathsf{E}[Y_i] = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}$$

  and the variance is

$$\mathsf{Var}(Y_i) = \sigma^2$$

- Therefore, we are assuming the variance is constant

The Normal Model: A Recap

- The error notation has its advantages...
- However, let's switch things up a bit

Statistical Modelling

What is Statistical Modelling?

Statistical Modelling

What is a Statistical Model?

Statistical Modelling

It's all about Probability!

Statistical Modelling

Building blocks

1. Response variable (Y)
2. Probability distribution
3. Parameters of interest

Statistical Modelling

Building blocks

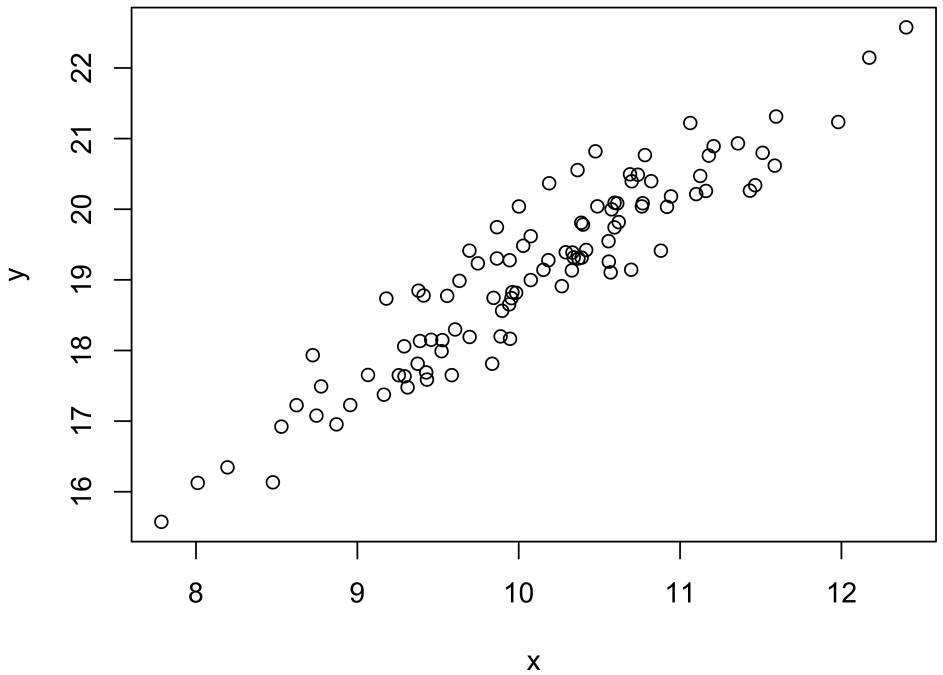
1. Response variable (Y)
2. Probability distribution
3. Parameters of interest $\leftarrow$ covariates / predictors

Statistical Modelling

- Y_i is a response variable associated with experimental or observational unit i
- We *assume* it comes from a certain probability distribution with pmf/pdf f and vector of parameters $\boldsymbol{\theta}$
- Very often one of the parameters in $\boldsymbol{\theta}$ is the mean of the distribution (or a function of the mean)
  - e.g.[1] for the normal distribution, $\boldsymbol{\theta} = (\mu, \sigma^2)^\top$, where μ is the mean of the distribution
  - e.g.[2] for the Poisson distribution, $\boldsymbol{\theta} = \mu$, where μ is the mean of the distribution
- We also have predictors x_i we are interested in studying
- We may link these predictors to any parameter of interest, but we typically do it for the **mean**

Statistical Modelling

Example



Statistical Modelling

Example

$$\begin{aligned} Y_i &\sim \mathsf{N}(\mu_i, \sigma^2) \\ \mu_i &= \beta_0 + \beta_1 x_i \end{aligned}$$

Statistical Modelling

Example

$$\begin{aligned} Y_i &\sim \mathsf{N}(\mu_i, \sigma^2) \\ \mu_i &= \beta_0 + \beta_1 x_i \end{aligned}$$

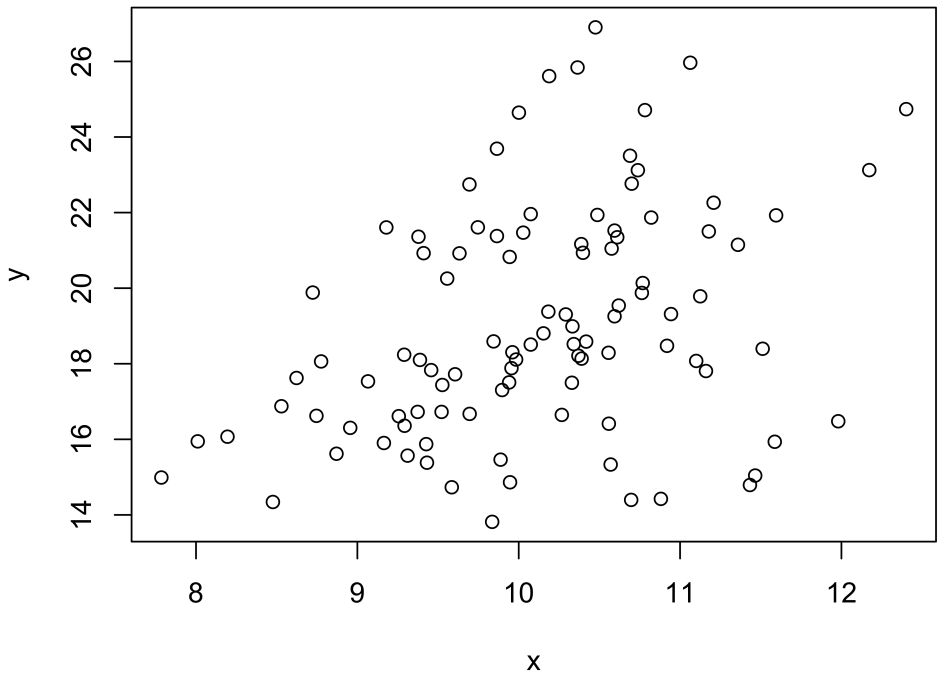
What about

- normality of residuals?
- homogeneity of variances?

This *only* makes sense for the model above!

Statistical Modelling

What if you had



Statistical Modelling

Maybe then you'd assume

$$\begin{aligned} Y_i &\sim \mathsf{N}(\mu_i, \sigma_i^2) \\ \mu_i &= \beta_0 + \beta_1 x_i \\ \log \sigma_i^2 &= \gamma_0 + \gamma_1 x_i \end{aligned}$$

No homogeneity of variances here!

Statistical Modelling

Y can be assumed to have *any* distribution

Statistical Modelling

Y can be assumed to have *any* distribution

Why is the normal distribution used to often then?

A Brief History of GLMs

- Multiple linear regression: a normal model with the identity link (Legendre, Gauss, Galton, 19th century)
- Analysis of variance (ANOVA): a normal model with the identity link (Fisher, 1918)
- Analysis of dilution assays: a binomial model with the complementary log-log link (Fisher, 1922)
- The exponential family class of distributions (Fisher, 1934)
- Probit analysis: a binomial distribution with the probit link (Bliss, 1935)
- Logistic regression: a binomial distribution with the logit link (Berkson, 1944; Dyke and Patterson, 1952)
- Item analysis: a Bernoulli distribution with the logit link (Rasch, 1960)
- Log-linear models: a Poisson distribution with the log link (Birch, 1963)
- Regression for survival data: an exponential distribution with the inverse or log links (Feigl and Zelen, 1965; Zippin and Armitage, 1966; Gasser, 1967)
- Inverse polynomials: a gamma distribution with the inverse link (Nelder, 1966)

A Brief History of GLMs

J. R. Statist. Soc. A,
(1972), **135**, *Part* 3, p. 370

Generalized Linear Models

By J. A. NELDER and R. W. M. WEDDERBURN

Rothamsted Experimental Station, Harpenden, Herts

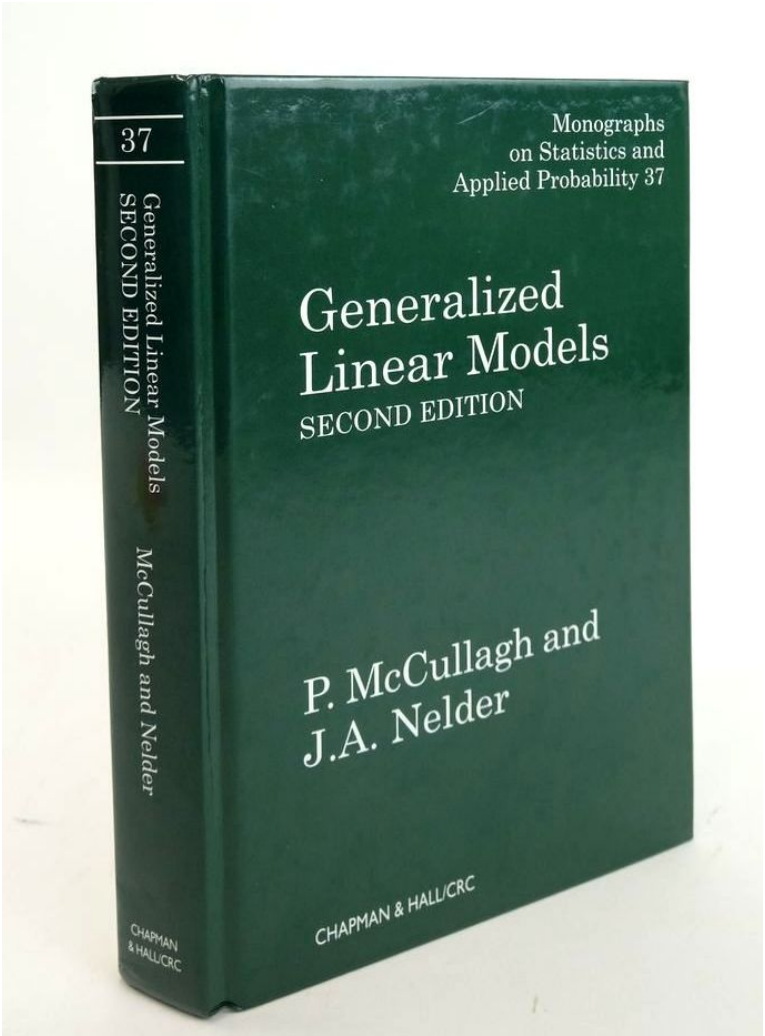
SUMMARY

The technique of iterative weighted linear regression can be used to obtain maximum likelihood estimates of the parameters with observations distributed according to some exponential family and systematic effects that can be made linear by a suitable transformation. A generalization of the analysis of variance is given for these models using log-likelihoods. These generalized linear models are illustrated by examples relating to four distributions; the Normal, Binomial (probit analysis, etc.), Poisson (contingency tables) and gamma (variance components).

The implications of the approach in designing statistics courses are discussed.

Keywords: ANALYSIS OF VARIANCE; CONTINGENCY TABLES; EXPONENTIAL FAMILIES; INVERSE POLYNOMIALS; LINEAR MODELS; MAXIMUM LIKELIHOOD; QUANTAL RESPONSE; REGRESSION; VARIANCE COMPONENTS; WEIGHTED LEAST SQUARES

A Brief History of GLMs



The Generalized Linear Model

- The generalized linear model can be defined using three components:

1. The random component: *a distribution belonging to the exponential family*
2. The systematic component: *a linear predictor*
3. The link function: *a function that links the mean to the linear predictor*

The Generalized Linear Model

- The generalized linear model can be defined using three components:

1. The random component: *a distribution belonging to the exponential family*
2. The systematic component: *a linear predictor*
3. The link function: *a function that links the mean to the linear predictor*

- More specifically, we have *independent* random variables $Y_i, i = 1, \ldots, n$
- The linear predictor can be written as $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$ where $\mathbf{X}$ is the $n \times p$ design (or model) matrix and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^\top$ is the vector of model coefficients
- The link function $g(\cdot)$ relates the mean μ_i to η_i, i.e. $g(\mu_i) = \eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}$, where $\mathbf{x}_i$ is the $i-$th row of $\mathbf{X}$

The Simple Linear Regression Model as a GLM

1. Random component: $Y_i \sim \mathsf{N}(\mu_i, \sigma^2)$
2. Systematic component: $\eta_i = \beta_0 + \beta_1 x_{1i}$
3. Link function: $g(\mu_i) = \mu_i$, the *identity* link

The Simple Linear Regression Model as a GLM

1. Random component: $Y_i \sim \mathsf{N}(\mu_i, \sigma^2)$
2. Systematic component: $\eta_i = \beta_0 + \beta_1 x_{1i}$
3. Link function: $g(\mu_i) = \mu_i$, the *identity* link

- More simply put:

$$\begin{aligned} Y_i &\sim \mathsf{N}(\mu_i, \sigma^2) \\ \mu_i &= \beta_0 + \beta_1 x_{1i} \end{aligned}$$

The Simple Linear Regression Model as a GLM

1. Random component: $Y_i \sim \mathsf{N}(\mu_i, \sigma^2)$
2. Systematic component: $\eta_i = \beta_0 + \beta_1 x_{1i}$
3. Link function: $g(\mu_i) = \mu_i$, the *identity* link

- More simply put:

$$\begin{aligned} Y_i &\sim \mathsf{N}(\mu_i, \sigma^2) \\ \mu_i &= \beta_0 + \beta_1 x_{1i} \end{aligned}$$

- Example: weight vs. height

Models for Binary Data

Models for Binary Data

- Example: `affairs` dataset

Models for Binary Data

- We have that our response variable Y_i is *binary*, i.e.

$$Y_i = \begin{cases} 1, & \text{if success} \\ 0, & \text{if failure} \end{cases}$$

Models for Binary Data

- We have that our response variable Y_i is *binary*, i.e.

$$Y_i = \begin{cases} 1, & \text{if success} \\ 0, & \text{if failure} \end{cases}$$

- We can then write

$$\begin{aligned} \mathsf{P}(\text{success}) &= \mathsf{P}(Y_i = 1) = \pi_i \\ \mathsf{P}(\text{failure}) &= \mathsf{P}(Y_i = 0) = 1 - \pi_i \end{aligned}$$

Models for Binary Data

- We have that our response variable Y_i is *binary*, i.e.

$$Y_i = \begin{cases} 1, & \text{if success} \\ 0, & \text{if failure} \end{cases}$$

- We can then write

$$\begin{aligned} \mathsf{P}(\text{success}) &= \mathsf{P}(Y_i = 1) = \pi_i \\ \mathsf{P}(\text{failure}) &= \mathsf{P}(Y_i = 0) = 1 - \pi_i \end{aligned}$$

- Y_i has a *Bernoulli* distribution, i.e.,

$$Y_i \sim \text{Bernoulli}(\pi_i)$$

The Bernoulli GLM

- Going back to our GLM components, we would like to model the success probabilities π_i as a function of predictors
- Can we simply write $\pi_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}$?
- Since the β coefficients are unbounded (i.e. can go from minus infinity to infinity), this would result in unbounded π_i values
- However, π_i are probabilities, and therefore, bounded in the $(0, 1)$ interval
- Therefore, we need a *link function* that maps the $(0, 1)$ interval do the real line
- One such link function is called the *logit link*

The Bernoulli GLM

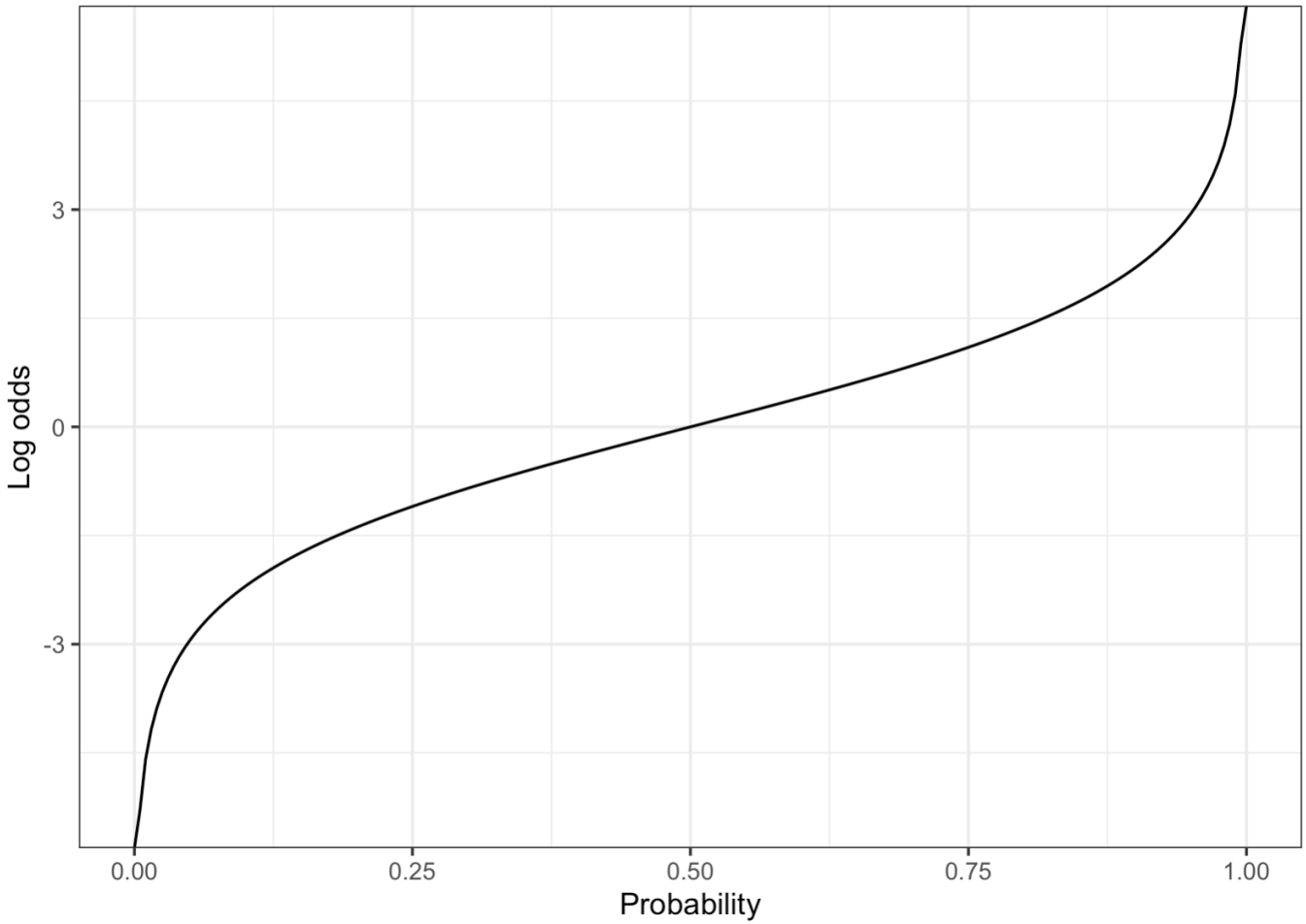
$$\begin{aligned} Y_i &= \sim \text{Bernoulli}(\pi_i) \\ \log\left(\frac{\pi_i}{1-\pi_i}\right) &= \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi} \end{aligned}$$

- The function $\log\left(\frac{\pi_i}{1-\pi_i}\right)$ can also be written as $\text{logit}(\pi_i)$
- "logit" stems from the words **log**istic un**it**, since it based on the cumulative distribution function of the logistic distribution
- It is simply the natural logarithm of the *odds*

The Bernoulli GLM: Important Concepts

- **Natural logarithm**: it is the logarithm base e, where $e = 2.71\ldots$ is called Euler's constant
- **Odds**: is the ratio between the probability of success and failure
  - e.g.^1 if the probability of rain is 90%, or 0.9, the odds of raining are $0.9/0.1$, i.e., 9 to 1
  - e.g.^2 if the odds of a horse winning a race are 3 to 1, it means that the probability the horse will win is 0.75, or 75%
  - odds are different than probabilities!
- **Odds ratio**: it is the ratio between two odds
  - e.g. the odds of a thoroughbred horse winning the race are 6 to 1, and the odds of a quarter horse winning the race are 3 to 1; the odds ratio is $6/3 = 2$, i.e. thoroughbred horses have twice the odds of winning the race than quarter horses

The Bernoulli GLM: The Logit Link

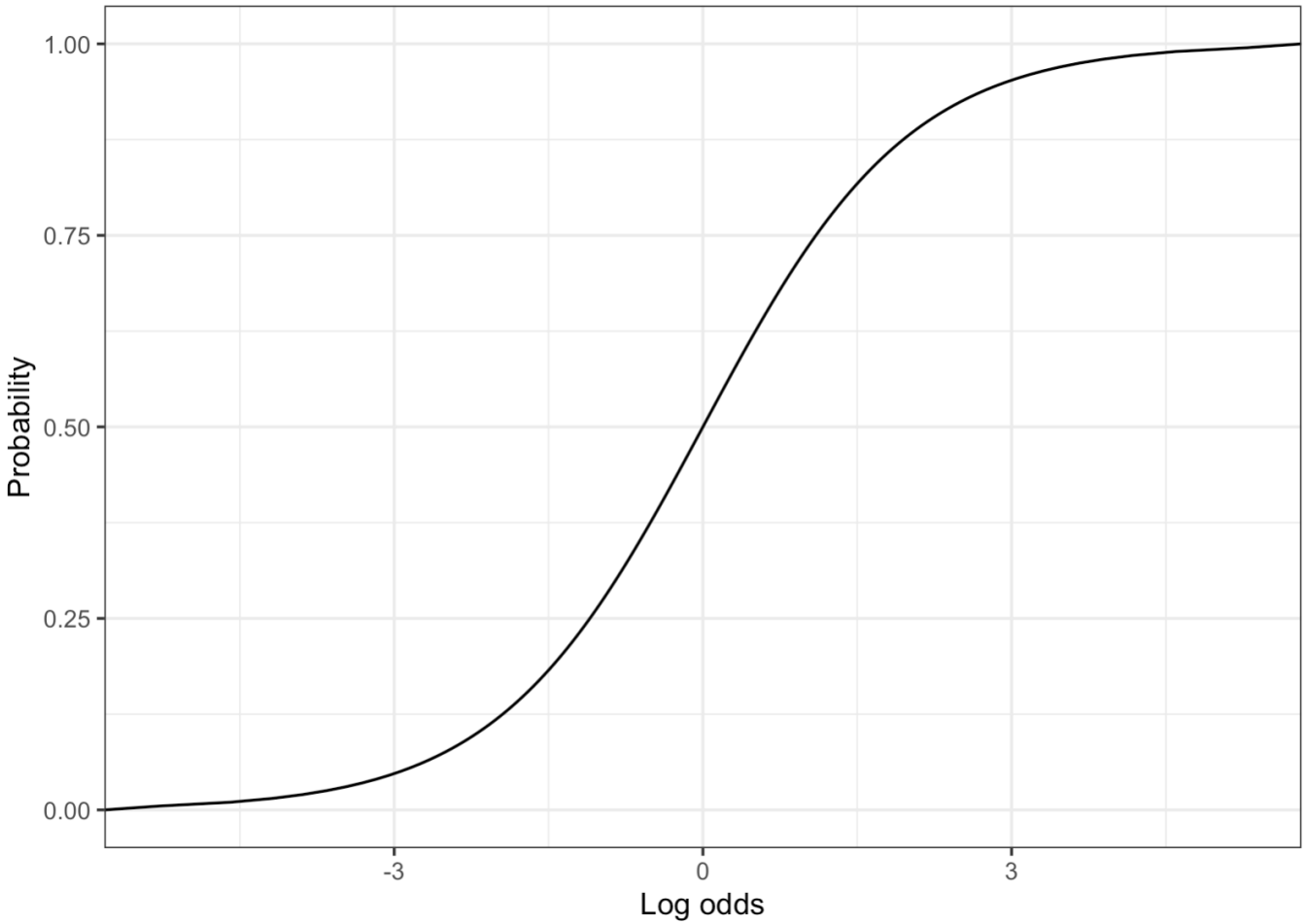


The Bernoulli GLM: The Inverse Logit Function

- We can go from probabilities to logits, but we can also go back
- This is especially important when interpreting Bernoulli model coefficients and calculating estimated probabilities
- We can easily show that if $\eta_i = \log\left(\frac{\pi_i}{1-\pi_i}\right)$, then

$$\pi_i = \frac{e^{\eta_i}}{1+e^{\eta_i}} = \frac{1}{1+e^{-\eta_i}}$$

The Bernoulli GLM: The Inverse Logit Function



The Bernoulli GLM: Interpretation

- Bernoulli regression coefficients are interpreted in the *log-odds scale*
- "For every extra year of marriage, the log-odds of having an affair increases by 0.0588 on average"
  - this is not very intuitive
- However, take the difference between two log-odds

$$\eta_1 - \eta_2 = \log\left(\frac{\pi_1}{1-\pi_1}\right) - \log\left(\frac{\pi_2}{1-\pi_2}\right) = \log\left(\frac{\pi_1}{1-\pi_1} \Big/ \frac{\pi_2}{1-\pi_2}\right)$$

- This is the log of the *odds ratio*

The Bernoulli GLM: Interpretation (cont'd)

- For an increase in one year of marriage, we have

$$\begin{aligned} \eta_1 &= \beta_0 + \beta_1(\text{years} + 1) \\ \eta_2 &= \beta_0 + \beta_1 \text{years} \\ \eta_1 - \eta_2 &= \beta_0 + \beta_1(\text{years} + 1) - (\beta_0 + \beta_1 \text{years}) = \beta_1 \end{aligned}$$

- Therefore, β_1 represents the log of the odds ratio
- If we exponentiate, we get the odds ratio:

$$e^{\beta_1} = \frac{\pi_1}{1-\pi_1} \Big/ \frac{\pi_2}{1-\pi_2}$$

- $e^{0.0588} = 1.06$: "For every extra year of marriage, the odds of having an affair increases by 6% on average"

Prediction

- After estimating the coefficients of a model, we can obtain predicted probabilities by using the inverse logit function:

$$\hat{\pi}_i = \frac{1}{1 + e^{-\left(\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_p x_{pi}\right)}}$$

Deviance

- Once we have the estimates of the parameters, we can calculate *goodness of fit*
- Nelder and Wedderburn introduced the concept of *deviance* for GLMs
- The deviance of a model is defined as

$$D = -2 \log L(\hat{\boldsymbol{\beta}}|\mathbf{y})$$

  where $L(\cdot)$ is called the *likelihood function*, and $\hat{\boldsymbol{\beta}}$ are the *maximum likelihood estimates* (MLE)
- This is a counterpart to R^2 for GLMs

Deviance: Model Testing

- Take a model with p_1 predictors (model $\mathcal{M}_1$) and a model with $p_2 < p_1$ predictors (model $\mathcal{M}_2$)
- Model $\mathcal{M}_1$ is called the *full model*, while model $\mathcal{M}_2$ is called the *current* or *reduced model*
- If model $\mathcal{M}_2$ is *nested* within model $\mathcal{M}_1$, we can use the difference between deviances as a test statistic to compare both models:

$$\Delta_D = D_2 - D_1 = -2 \log \frac{L(\hat{\boldsymbol{\beta}}_2|\mathbf{y})}{L(\hat{\boldsymbol{\beta}}_1|\mathbf{y})}$$

  where $\hat{\beta}_1$ and $\hat{\beta}_2$ are the MLEs of the full and reduced models, respectively
- Under the null hypothesis that both models are equivalent in terms of fit, Δ_D has an asymptotic χ^2 distribution with $p_1 - p_2$ d.f.
- In practice, $\Delta_D \approx p_1 - p_2$ indicates little support to the alternative hypothesis that the full model fits the data better

Models for Binomial Data

The Binomial GLM

- Example: *Diaphorina citri* mortality data

The Binomial GLM

- If Y_i represents the total number of successes out of m_i independent Bernoulli trials, all with the same probability of success π_i, then Y_i has a binomial distribution:

$$Y_i \sim \text{Binomial}(m_i, \pi_i)$$

- Important assumptions:
  - *independent* Bernoulli trials
  - *same probability of success* for each Bernoulli trial
- π_i is intrinsically related to the mean of the binomial distribution:

$$\mu_i = m_i \pi_i$$

- We are interested in modelling the probability of success π_i as a function of predictors
- Typically m_i is known *a priori*

The Binomial GLM

- We can also use the *logit* link function:

$$\begin{aligned} Y_i &\sim \text{Binomial}(m_i, \pi_i) \\ \log\left(\frac{\pi_i}{1-\pi_i}\right) &= \log\left(\frac{\mu_i}{m_i-\mu_i}\right) = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi} \end{aligned}$$

- The logit link allows for the same interpretation as for the Bernoulli model
- e^{β} translates into the odds ratio
- Computing $\hat{\pi}_i$ is the same as for the Bernoulli model
- We can also easily compute $\hat{\mu}_i = m_i \hat{\pi}_i$
- More on inference and goodness-of-fit when we discuss overdispersion!

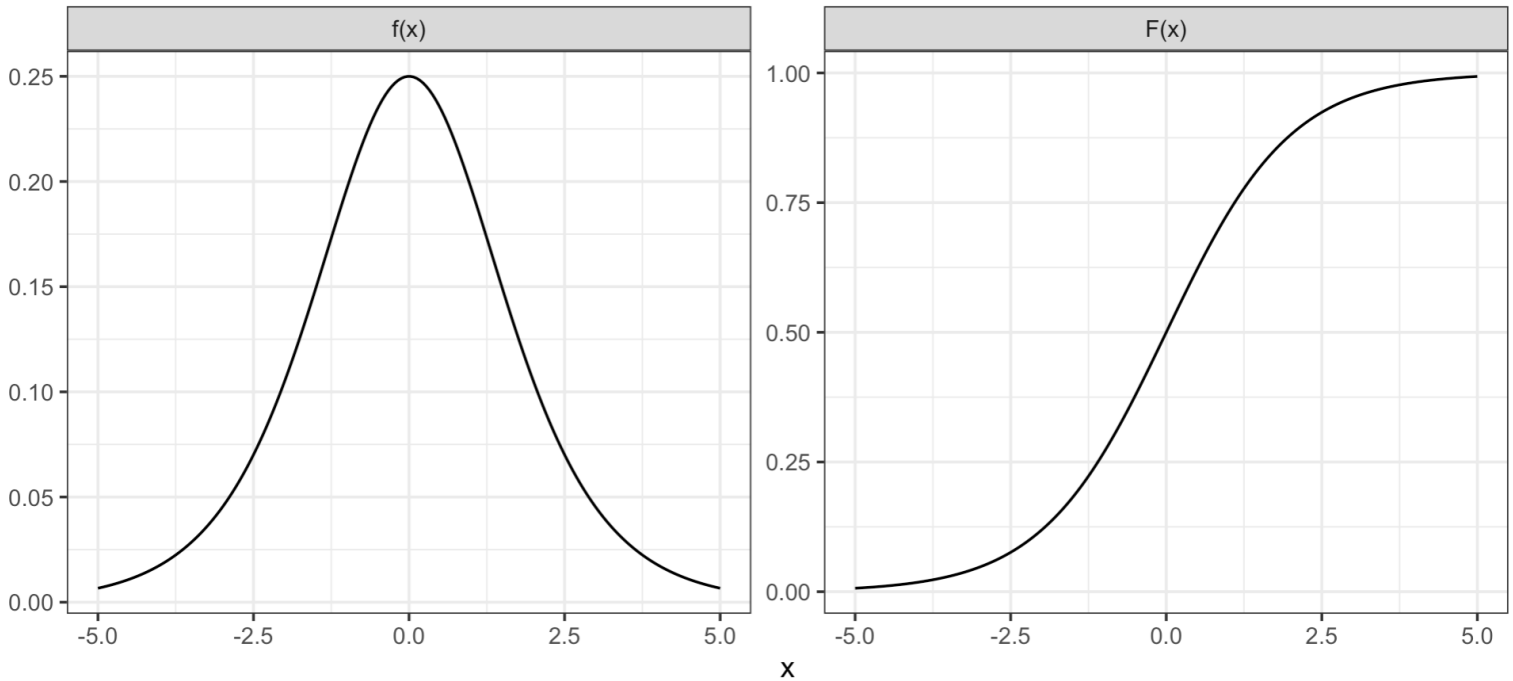
The Binomial GLM: Alternative Link Functions

- There are many alternatives to the logit link in the literature implemented in R
- We will cover the *probit* and *complementary log-log* links in this course
- But first, we go back to the logit link and introduce the *logistic distribution*

The Binomial GLM: Alternative Link Functions

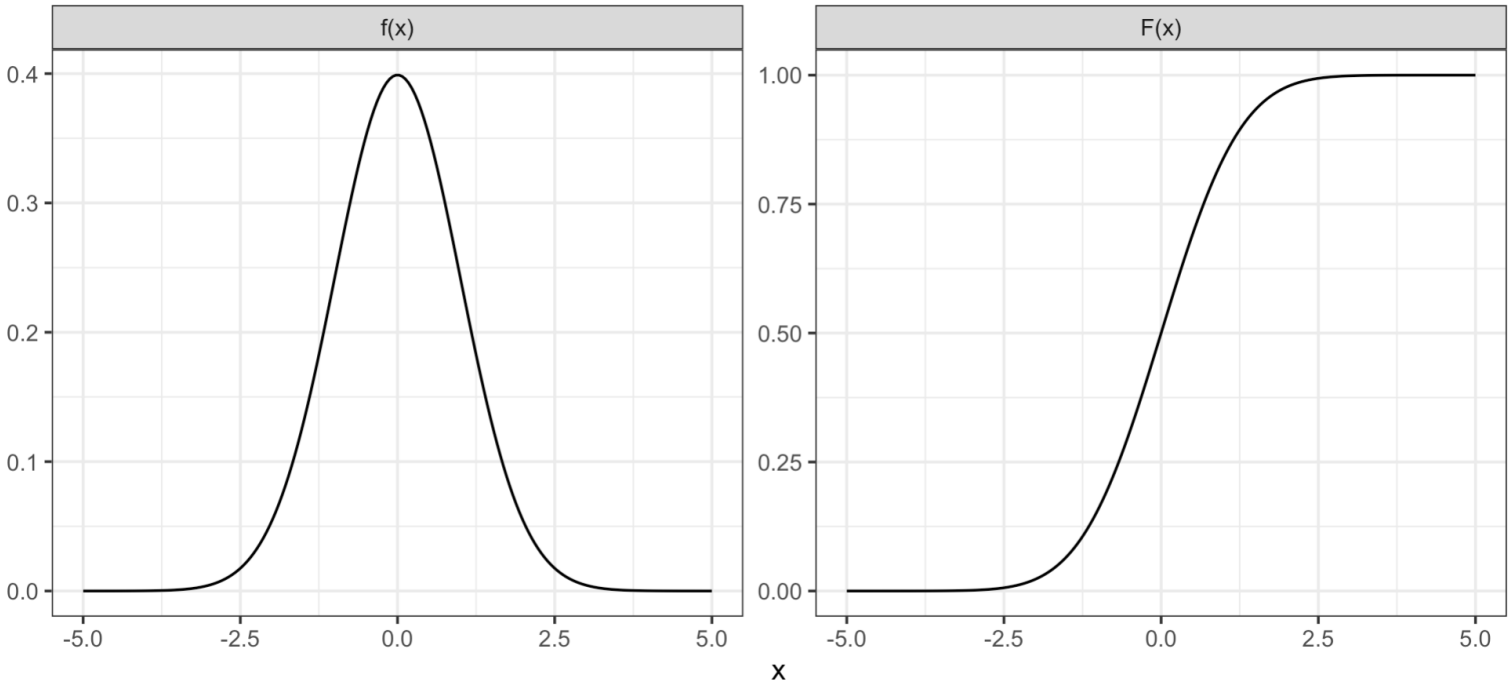
- If X has a logistic distribution with mean zero and scale parameter equal to one, then the cumulative distribution function of Y is

$$F(x) = \mathsf{P}(X \leq x) = \frac{1}{1 + e^{-x}}$$



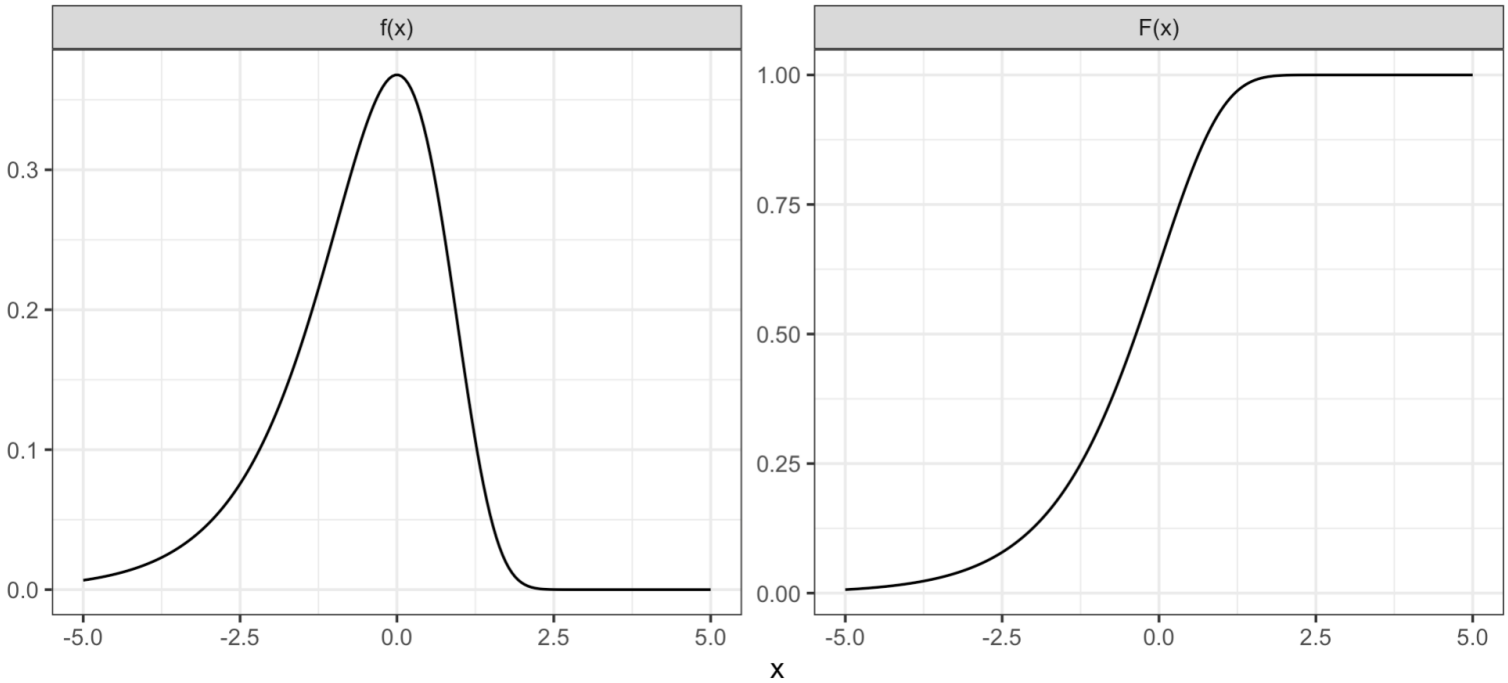
The Binomial GLM: Alternative Link Functions

- The *probit link* stems from the normal distribution, and means **prob**ability un**it**
- We denote write is as $g(\pi_i) = \Phi^{-1}(\pi_i)$, where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal cdf

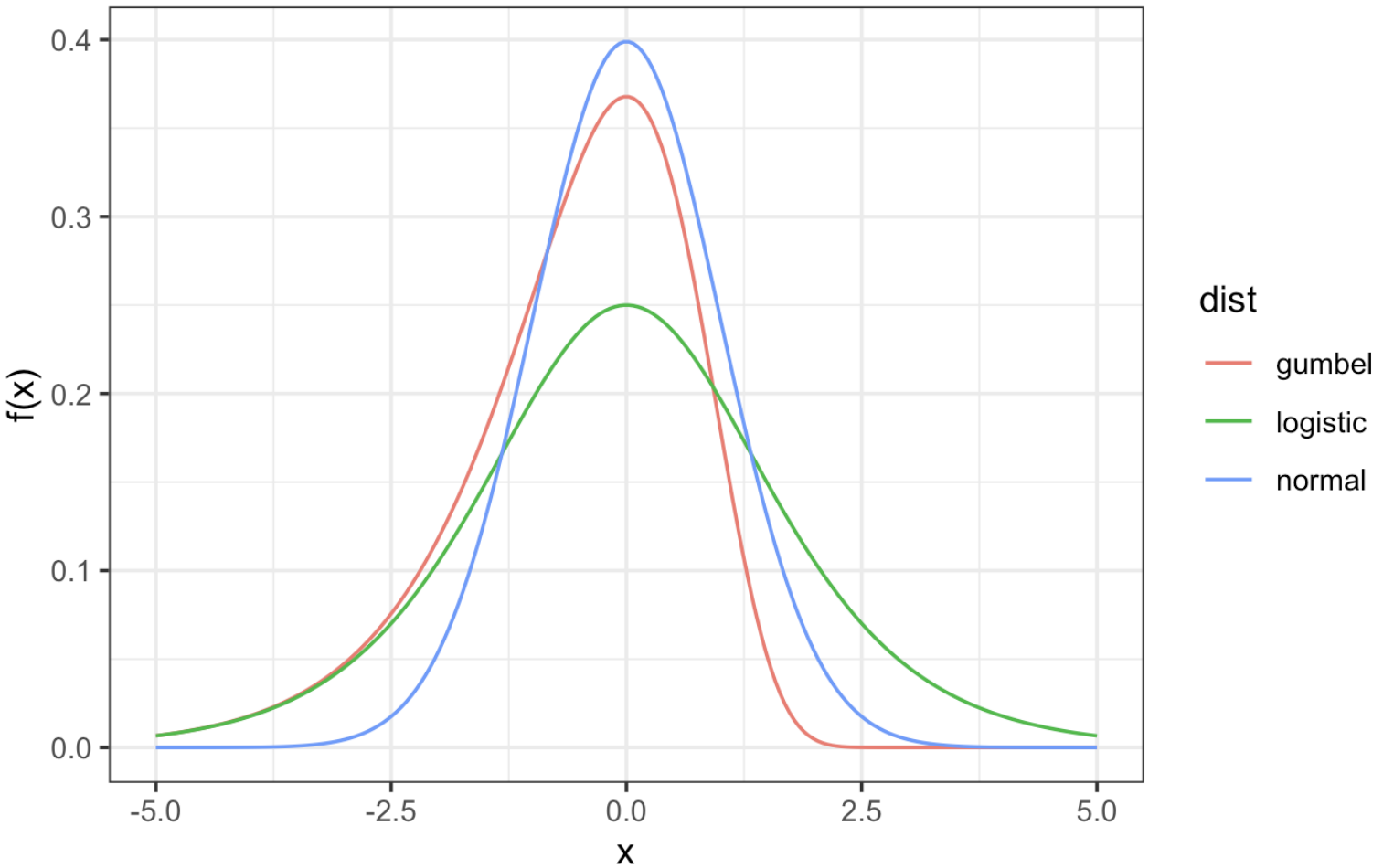


The Binomial GLM: Alternative Link Functions

- The *complementary log-log* stems from the Gumbel distribution
- It is asymmetric, in comparison to the probit and logit links
- We write it as $g(\pi_i) = \log(-\log(1-\pi_i))$



The Binomial GLM: Alternative Link Functions



The Binomial GLM: Alternative Link Functions