

ST346 Chapter 2

Contents

Preface	3
2 Binomial models	4
2.1 Bernoulli distribution	4
2.2 Latent linear model	6
2.3 Binomial outcomes	9
2.4 Scaled binomial distribution	10
2.5 Implementation in R	10
2.6 Logistic regression	12
2.7 Tolerance distributions	18

Preface

These slides are a slight adaptation from the original slides developed by Prof Martyn Plummer for the module.

If you find any typos, please inform the module leader.

These materials are solely for your own use and you must not distribute these in any format.

Do not upload these materials to the internet or any filesharing sites nor provide them to any third party or forum.

All rights are reserved.

Chapter 2 Binomial models

2.1 Bernoulli distribution

One of the simplest probability distribution is named after the Swiss mathematician Jacob Bernoulli (1655–1705).

Suppose $Y \in \{0, 1\}$ and let $\mu = \mathbb{E}(Y)$, then

$$\mathbb{P}(Y = 1) = \mu, \quad \mathbb{P}(Y = 0) = 1 - \mu.$$

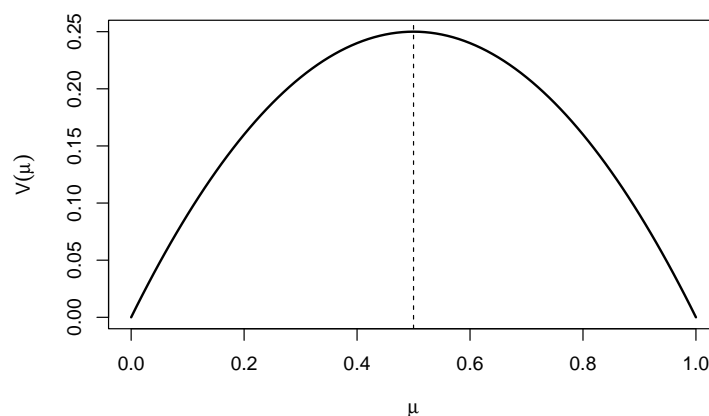
We write $Y \sim \text{Bernoulli}(\mu)$. Then, Y has probability mass function

$$p(y) = \mu^y (1 - \mu)^{1-y}.$$

The **variance** of the Bernoulli distribution depends on the mean μ . We write $\mathbb{V}ar(Y) = V(\mu)$.

As $Y \in \{0, 1\}$ we have $Y^2 = Y$ and so

$$\begin{aligned} V(\mu) &= \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 \\ &= \mathbb{E}(Y) - \mathbb{E}(Y)^2 \\ &= \mu - \mu^2 = \mu (1 - \mu) \end{aligned}$$



Note: Heteroscedasticity is built into the Bernoulli distribution.

- This will not work with the Bernoulli distribution.
- Any transformation of Y to a new variable (say Y^*) will map all $Y = 0$ values to the same value (say $Y^* = a$) and all $Y = 1$ values to the same value (say $Y^* = b$).
- The mean of the transformed variable is

- The variance of the transformed variable is

Whatever we do, the variance depends on the mean.

2.2 Latent linear model

2.2.1 Definition

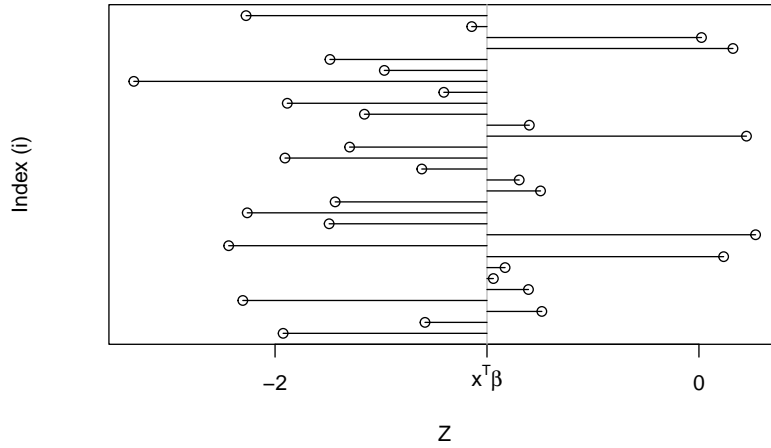
We can use a normal linear model by introducing **latent variables**. Latent variables are random variables that are not observed. Let $Z_i \sim \mathcal{N}(\mathbf{x}_i^T \boldsymbol{\beta}, 1)$ where Z_1, \dots, Z_n are independent. Then, set

$$Y_i = \begin{cases} 1 & \text{if } Z_i \geq 0, \\ 0 & \text{if } Z_i < 0. \end{cases}$$

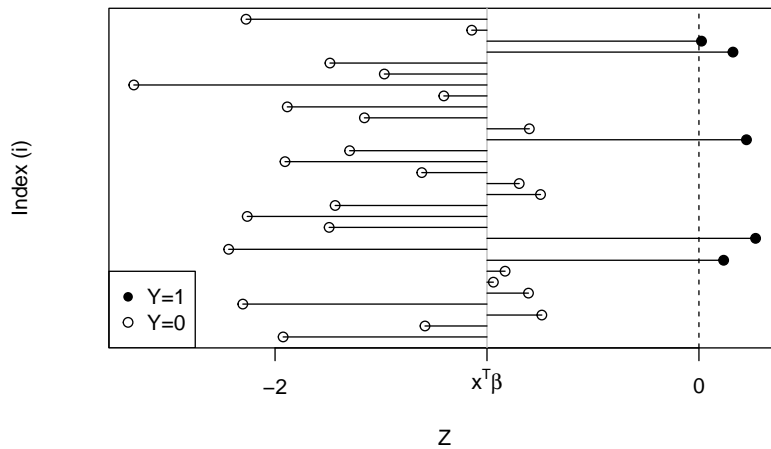
2.2.2 Illustration

The graph shows an iid random sample of $Z \sim \mathcal{N}(\mathbf{x}^T \boldsymbol{\beta}, 1)$ of size $n = 30$.

(Here we have chosen $\mathbf{x}^T \boldsymbol{\beta} = -1$.)

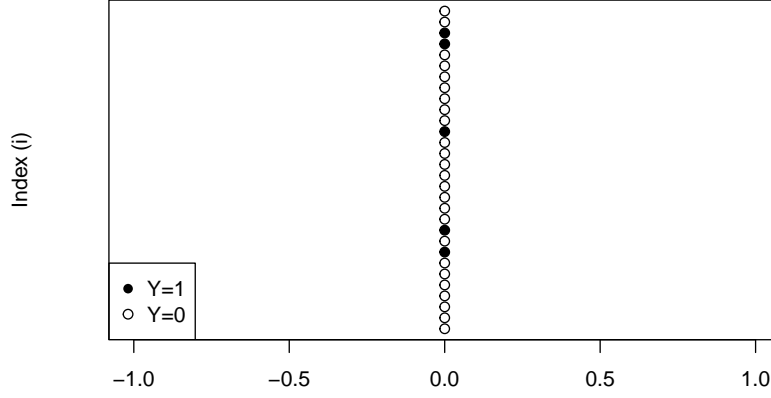


- For values of $Z_i \geq 0$, the corresponding $Y_i = 1$ (black circles).
- For values of $Z_i < 0$, the corresponding $Y_i = 0$ (white circles).



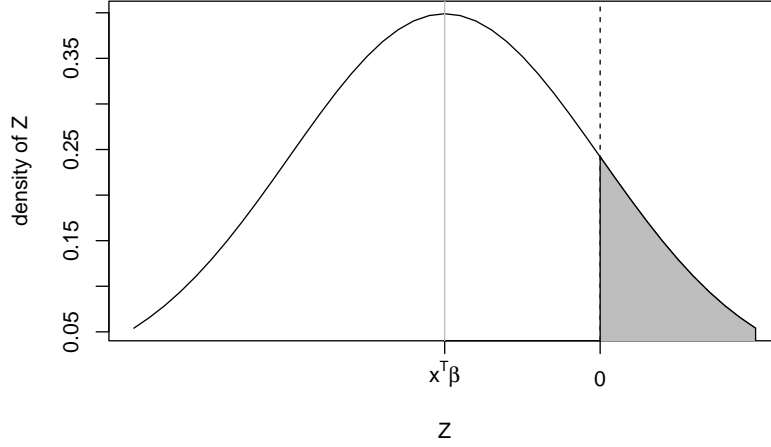
- We do not observe Z_1, \dots, Z_n , only Y_1, \dots, Y_n , represented here by the colour of the points.

- We need to use a probability model for Y_i based on the latent value of Z_i .



Consider the density of $Z_i \sim \mathcal{N}(\mathbf{x}_i^T \boldsymbol{\beta}, 1)$ shown below.

The shaded tail area in the figure below shows $\mu_i = \mathbb{P}(Y_i = 1) = \mathbb{P}(Z_i \geq 0)$.



We have

$$\begin{aligned}
 \mu_i = \mathbb{P}(Y_i = 1) &= \mathbb{P}(Z_i \geq 0) \\
 &= \mathbb{P}(\mathcal{N}(0, 1) \geq -\mathbf{x}_i^T \boldsymbol{\beta}) \\
 &= \mathbb{P}(\mathcal{N}(0, 1) \leq \mathbf{x}_i^T \boldsymbol{\beta}) \\
 &= \Phi(\mathbf{x}_i^T \boldsymbol{\beta}).
 \end{aligned}$$

where Φ (“Phi”) is the cumulative distribution function of the standard normal distribution.

To avoid writing $\mathbf{x}_i^T \boldsymbol{\beta}$ we introduce a new notation

$$\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$$

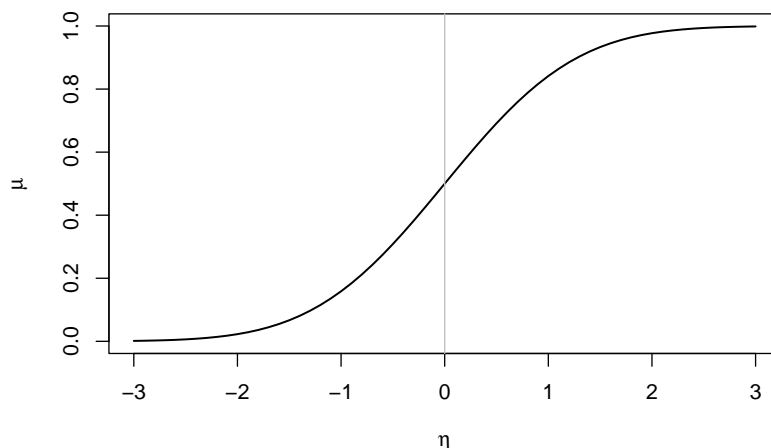
and new terminology.

We call this quantity – represented by the Greek letter “eta” – the **linear predictor**.

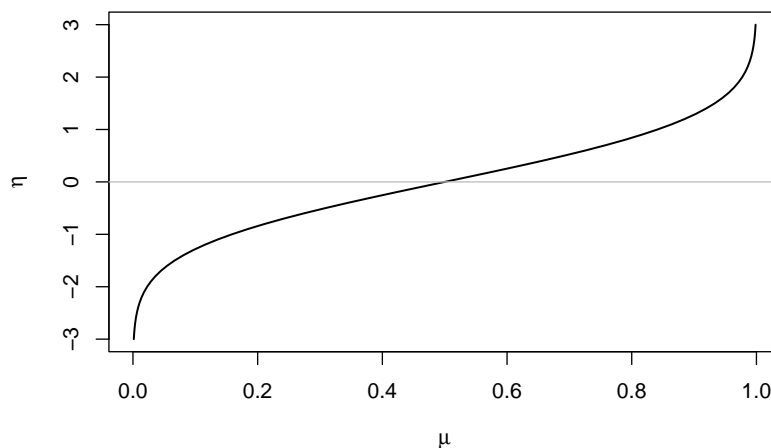
The expectation μ_i of the outcome variable Y_i is not the same as the linear predictor $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$. The two quantities are related by a smooth increasing function, namely Φ as

$$\mu_i = \Phi(\eta_i).$$

The **mean function** maps the linear predictor η onto the mean μ .



The **link function** is the inverse of the mean function. It maps the mean onto the linear predictor.



For the model considered here the link function is the inverse of Φ , that is

$$\eta_i = \Phi^{-1}(\mu_i).$$

This link function is called the probit function. The name comes from an abbreviation of “**probability unit**”.

Exercise 3

Work through Sections 1 - 3 (inclusive) from Computer Practical 1.

2.3 Binomial outcomes

- The name “binomial” comes from the Latin for “two names”.
- A binomial outcome is used whenever the response is a choice between two categories:
 - Yes vs No
 - Success vs Failure
- For a group of observations with the same predictor variables, we can add together the responses into a single outcome value for the whole group (e.g `turbines` data).

	Hours	Turbines	Fissures
1	400	39	0
2	1000	53	4
3	1400	33	2
4	1800	73	7
5	2200	30	5
6	2600	39	9

The **binomial distribution** is defined via a sequence of m independent Bernoulli random variables with a common mean.

$$Z_1, \dots, Z_m \sim \text{Bernoulli}(\mu).$$

Likelihood:

$$\begin{aligned}
 L(\mu | z_1, \dots, z_n) &= \prod_{i=1}^m p(z_i | \mu) \\
 &= \prod_{i=1}^m \mu^{z_i} (1 - \mu)^{1-z_i} \\
 &= \mu^{\sum_{i=1}^m z_i} (1 - \mu)^{m - \sum_{i=1}^m z_i} \\
 &= \mu^y (1 - \mu)^{m-y}
 \end{aligned}$$

where $y = \sum_{i=1}^m z_i$.

Note: y is **sufficient** for μ .

The sum of m Bernoulli trials with common mean μ has a binomial distribution:

$$Y \sim \text{Binomial}(m, \mu).$$

Probability mass function:

$$p(y) = \binom{m}{y} \mu^y (1 - \mu)^{m-y}$$

for $y \in \{0, 1, \dots, m\}$.

Then

$$\begin{aligned}
 \mathbb{E}(Y) &= m\mu, \\
 \text{Var}(Y) &= m\mu(1 - \mu).
 \end{aligned}$$

2.4 Scaled binomial distribution

In GLMs we use a different version of the binomial distribution. Let

$$Y^* = \frac{1}{m} \sum_{i=1}^m Z_i.$$

Then Y^* has a **scaled** binomial distribution with **probability mass function**

$$p(y^*) = \binom{m}{my^*} \mu^{my^*} (1 - \mu)^{m(1-y^*)} \quad \text{for } y^* \in \{0, \frac{1}{m}, \frac{2}{m}, \dots, 1\}.$$

Then,

$$\begin{aligned} \mathbb{E}(Y^*) &= \mu, \\ \mathbb{V}ar(Y^*) &= \frac{1}{m} \mu(1 - \mu) = \frac{1}{m} V(\mu). \end{aligned}$$

for **variance function** $V(\mu) = \mu(1 - \mu)$.

Log-likelihood:

$$\begin{aligned} l(\mu) &= \log(p(y | \mu)) \\ &= m \left(y^* \log(\mu) + (1 - y^*) \log(1 - \mu) + \dots \right) \end{aligned}$$

where we discard terms depending only on data (y^*, m) .

This formulation allows us to use

- the proportion $Y^* = Y/m$ as the response variable;
- sample size m as a weight.

2.5 Implementation in R

Let's consider the turbines data from the `GLMsData` package.

We can fit the probit binomial GLM in two ways.

The first approach uses the proportion y/m as response and weights m :

```
glm.out1 <- glm(Fissures/Turbines ~ Hours, weights=Turbines,
               family=binomial(link="probit"), data=turbines)
```

The second approach uses the vector $(y, m - y)$ as the response without any weights:

```
turbines <- transform(turbines, NoFissures = Turbines - Fissures)
glm.out2 <- glm(cbind(Fissures, NoFissures) ~ Hours,
               family=binomial(link="probit"), data=turbines)
```

We use the `summary()` function from the `faraway` package to produce an abbreviated summary of the two models.

```
library(faraway)
summary(glm.out1)
```

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.2758e+00	1.9742e-01	-11.5278	< 2.2e-16
Hours	5.7832e-04	6.2597e-05	9.2388	< 2.2e-16

```
n = 11 p = 2
```

```
Deviance = 9.81484 Null Deviance = 112.67005 (Difference = 102.85521)
```

```
summary(glm.out2)
```

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.2758e+00	1.9742e-01	-11.5278	< 2.2e-16
Hours	5.7832e-04	6.2597e-05	9.2388	< 2.2e-16

```
n = 11 p = 2
```

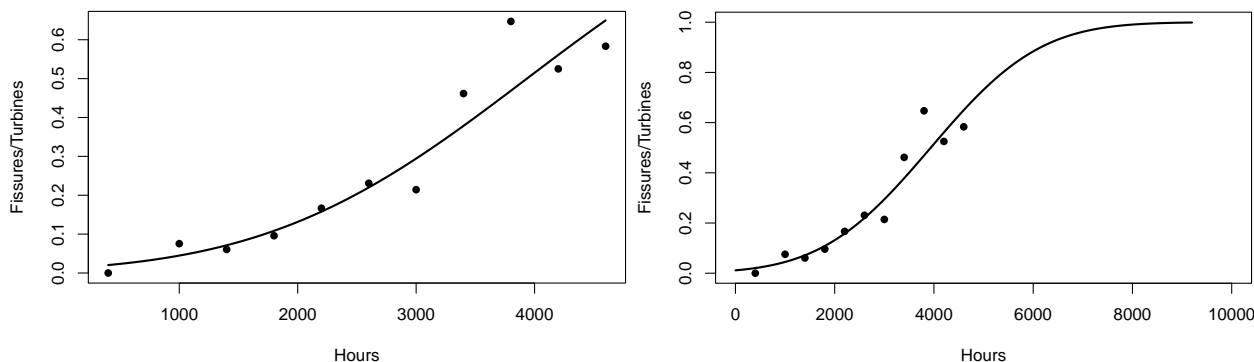
```
Deviance = 9.81484 Null Deviance = 112.67005 (Difference = 102.85521)
```

Both models give the same estimates, standard errors, deviance, etc.

Let's explore the fitted model graphically. The fitted line is non-linear so we need to use the `predict()` function with `type = "response"` to indicate that we would like to get the predictions in the scale of the response. If we examine the fitted curve for very small and very large number of hours we observe horizontal asymptotes at 0 and at 1 (beware of extrapolation!).

```
plot(Fissures/Turbines ~ Hours, data=turbines, pch=16)
hvec <- seq(from=min(turbines$Hours), to=max(turbines$Hours), length=101)
ypred <- predict(glm.out1, newdata=list(Hours=hvec), type="response")
lines(hvec, ypred, lwd=2)
```

```
plot(Fissures/Turbines ~ Hours, data=turbines,
     xlim=c(0, 10000), ylim=c(0,1), pch=16)
hvec2 <- seq(from=0, to=2*max(turbines$Hours), length=201)
ypred2 <- predict(glm.out1, newdata=list(Hours=hvec2), type="response")
lines(hvec2, ypred2, lwd=2)
```



2.6 Logistic regression

To interpret the parameters of a binary regression model we need to be able to translate between the linear predictor and the mean.

We do this via the link function $\eta_i = g(\mu_i)$ and mean function $\mu_i = h(\eta_i)$.

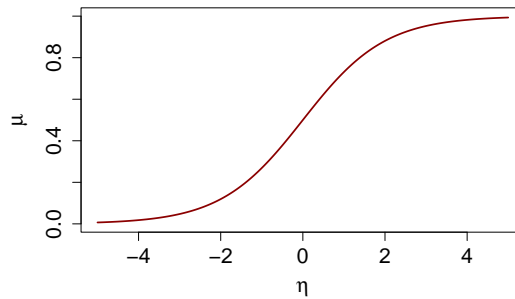
But the probit link has no closed form.

The **logit** link function is an alternative that has a simple enough closed form.

2.6.1 The logit link function

The **logistic** function is defined as

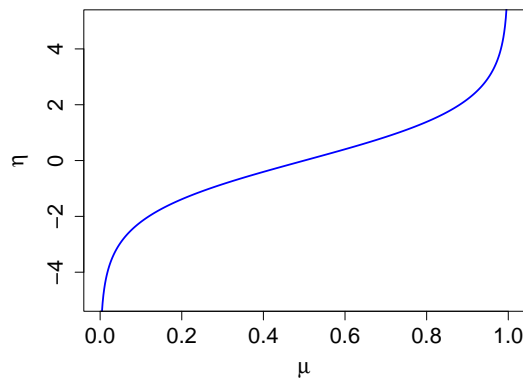
$$h(\eta) = \frac{\exp(\eta)}{1 + \exp(\eta)} = \frac{1}{1 + \exp(-\eta)}.$$



It is monotonically increasing with horizontal asymptotes at zero and 1 and thus is suited to modelling probabilities.

The corresponding link function is the **logit link** function:

$$g(\mu) = \log\left(\frac{\mu}{1 - \mu}\right) = \eta.$$



A GLM for binary response data using a logistic mean function and logit link is known as a **logistic** regression model.

We interpret the parameters of a logistic regression model in terms of **log odds ratios**.

2.6.2 Odds of an event occurring

Odds are another way to express probabilities and are used in gambling.

Suppose an event E has known probability μ . You think it will happen; I do not. I bet 1 pound that event E will not occur.

- If E occurs, then you get to keep my stake of 1 pound.
- If E does not occur, then you return my stake and, in addition, give me L pounds.

Suppose we both agree on the probability μ but disagree only in our predictions of what will happen.

If the game is fair, then your expected gain and mine are both equal to 0.

$$\mu \times 1 - (1 - \mu) \times L = 0.$$

Rearranging gives

$$L = \frac{\mu}{1 - \mu}.$$

These are the **odds of E occurring**.

2.6.3 Interpreting parameters for binary predictors

We will use terminology from epidemiology:

- If $Y_i = 1$ the individual is diseased.
- If $Y_i = 0$ the individual is healthy.
- If $X_i = 1$ the individual is exposed.
- If $X_i = 0$ the individual is unexposed.

Let

$$\begin{aligned}\mu^{(0)} &= \mathbb{P}(Y_i = 1 \mid X_i = 0), \\ \mu^{(1)} &= \mathbb{P}(Y_i = 1 \mid X_i = 1).\end{aligned}$$

The odds of disease ($Y_i = 1$) in the unexposed ($X_i = 0$) is $\frac{\mu^{(0)}}{1 - \mu^{(0)}}$.

The odds of disease ($Y_i = 1$) in the exposed ($X_i = 1$) is $\frac{\mu^{(1)}}{1 - \mu^{(1)}}$.

The **odds ratio** is given by

$$\begin{aligned}\Omega &= \frac{\mu^{(1)}}{1 - \mu^{(1)}} \times \frac{1 - \mu^{(0)}}{\mu^{(0)}} \\ &= \frac{\text{Odds of disease for the exposed}}{\text{Odds of disease for the unexposed}}.\end{aligned}$$

If $\mu_i = \mathbb{E}(Y_i)$, then using the logit link

$$g(\mu_i) = \log\left(\frac{\mu_i}{1 - \mu_i}\right) = \alpha + \beta x_i.$$

It follows that

$$\frac{\mu_i}{1 - \mu_i} = \exp(\alpha + \beta x_i).$$

The odds of disease ($Y_i = 1$) in the unexposed ($X_i = 0$) is

$$\frac{\mu^{(0)}}{1 - \mu^{(0)}} = \exp(\alpha + \beta \times 0) = \exp(\alpha).$$

So the intercept α is the log odds of disease when $X_i = 0$.

The odds of disease ($Y_i = 1$) in the exposed ($X_i = 1$) is

$$\frac{\mu^{(1)}}{1 - \mu^{(1)}} = \exp(\alpha + \beta \times 1) = \exp(\alpha) \times \exp(\beta).$$

So the odds ratio is

$$\Omega = \frac{\exp(\alpha) \times \exp(\beta)}{\exp(\alpha)} = \exp(\beta).$$

Hence $\beta = \log(\Omega)$ is the **log odds ratio** for $X_i = 1$ versus $X_i = 0$.

2.6.4 Interpreting parameters for continuous predictors

Suppose X_i is continuous instead of binary. We still assume a logistic regression model with

$$\log\left(\frac{\mu_i}{1 - \mu_i}\right) = \alpha + \beta x_i.$$

Then β is the **log odds ratio** for a **unit increase** in X_i .

- If we compare the odds for $X_i = x + 1$ versus the odds for $X_i = x$, then the odds ratio is $\exp(\beta)$.
- The magnitude of β depends on the units (e.g. cigarettes or packets in the practical class).

2.6.5 Interpreting parameters for categorical predictors

If we have a categorical predictor variable (factor) with K levels $\{1, \dots, K\}$ and we use treatment coding, then the logistic regression model is

$$\log\left(\frac{\mu_i}{1 - \mu_i}\right) = \alpha + \sum_{k=2}^K \beta_k \mathbf{1}_{[X_i=k]}$$

where the original categorical variable has been replaced by $K - 1$ indicator variables.

- The **reference category** ($k=1$) is “absorbed” into the intercept α .
- For $k \in \{2, \dots, K\}$, the parameter β_k is the **log odds ratio** for level k versus level 1.

2.6.6 Interpreting the intercept

For all cases, the intercept parameter α is the log odds of the outcome $Y_i = 1$ when all predictors are at their reference value, that is

- zero for binary variables,
- zero for continuous variables,
- reference level for factors (categorical variables using treatment coding).

2.6.7 Interpreting multiple logistic regression

We may have multiple predictors in a logistic regression model, for example

$$\log\left(\frac{\mu_i}{1 - \mu_i}\right) = \alpha + \beta x_i + \gamma z_i.$$

The interpretation of β is the same as before but assumes holding all other predictor variables fixed (that is z_i in the above example).

2.6.8 Epidemiological example

Suppose we have a single binary predictor $X_i \in \{0, 1\}$ and a binary response $Y_i \in \{0, 1\}$ for N independent observations.

We can summarise the data in a 2×2 table.

	$X_i = 1$	$X_i = 0$	Total
$Y_i = 1$	$D^{(1)}$	$D^{(0)}$	D
$Y_i = 0$	$H^{(1)}$	$H^{(0)}$	H
Total	$N^{(1)}$	$N^{(0)}$	N

Then

$$\begin{aligned} D^{(0)} &\sim \text{Binomial}(N^{(0)}, \mu^{(0)}), \\ D^{(1)} &\sim \text{Binomial}(N^{(1)}, \mu^{(1)}), \end{aligned}$$

where

$$\begin{aligned} \mu^{(0)} &= \mathbb{P}(Y_i = 1 \mid X_i = 0), \\ \mu^{(1)} &= \mathbb{P}(Y_i = 1 \mid X_i = 1). \end{aligned}$$

Recall, the **odds ratio** is given by

$$\begin{aligned} \Omega &= \frac{\mu^{(1)}}{1 - \mu^{(1)}} \times \frac{1 - \mu^{(0)}}{\mu^{(0)}} \\ &= \frac{\text{Odds of disease for the exposed}}{\text{Odds of disease for the unexposed}}. \end{aligned}$$

$\Omega = 1$ if and only if $\mu^{(1)} = \mu^{(0)}$, so the risk is the same in both exposed and unexposed.

The maximum likelihood estimate is the cross-product ratio

$$\hat{\Omega} = \frac{D^{(1)} H^{(0)}}{H^{(1)} D^{(0)}}.$$

We note the balance between the numerator and denominator

$$\Omega = \frac{\mathbb{P}(Y_i = 1 \mid X_i = 1) \mathbb{P}(Y_i = 0 \mid X_i = 0)}{\mathbb{P}(Y_i = 1 \mid X_i = 0) \mathbb{P}(Y_i = 0 \mid X_i = 1)}.$$

We can use Bayes' theorem to invert the conditional probabilities. For $a, b \in \{0, 1\}$

$$\mathbb{P}(Y_i = a \mid X_i = b) = \mathbb{P}(X_i = b \mid Y_i = a) \times \frac{\mathbb{P}(Y_i = a)}{\mathbb{P}(X_i = b)}.$$

Due to the balance between numerator and denominator in the expression for Ω , the second factor cancels out and

$$\Omega = \frac{\mathbb{P}(X_i = 1 \mid Y_i = 1) \mathbb{P}(X_i = 0 \mid Y_i = 0)}{\mathbb{P}(X_i = 1 \mid Y_i = 0) \mathbb{P}(X_i = 0 \mid Y_i = 1)}.$$

So we can express Ω as the odds ratio of $X_i = 1$ for $Y_i = 1$ versus $Y_i = 0$.

Inversion of the odds ratio is useful in epidemiology.

We can estimate the odds ratio by recruiting participants using their disease status.

- Identify D individuals with a given disease ($Y_i = 1$).
- Find a group of H healthy individuals ($Y_i = 0$).
- Measure the exposure X and classify participants as exposed ($X_i = 1$) or unexposed ($X_i = 0$).

This is called a **case control study**.

2.6.9 The rare outcome interpretation

Suppose the outcome $Y_i = 1$ is rare as $\mathbb{P}(Y_i = 1)$ is small for both groups (exposed and unexposed):

$$\begin{aligned}\mu^{(0)} &\ll 1, \\ \mu^{(1)} &\ll 1.\end{aligned}$$

Then

$$\Omega = \frac{\mu^{(1)} (1 - \mu^{(0)})}{\mu^{(0)} (1 - \mu^{(1)})} \approx \frac{\mu^{(1)}}{\mu^{(0)}}$$

and so the odds ratio is approximately equal to the **relative risk**, or risk ratio.

Exercise 4

Work through Sections 4 - 5 from Computer Practical 1.

2.7 Tolerance distributions

2.7.1 Generalizing the latent linear model

The earlier construction of the latent linear model for a probit regression can be generalized beyond the normal distribution.

We can write the latent variable Z_i as

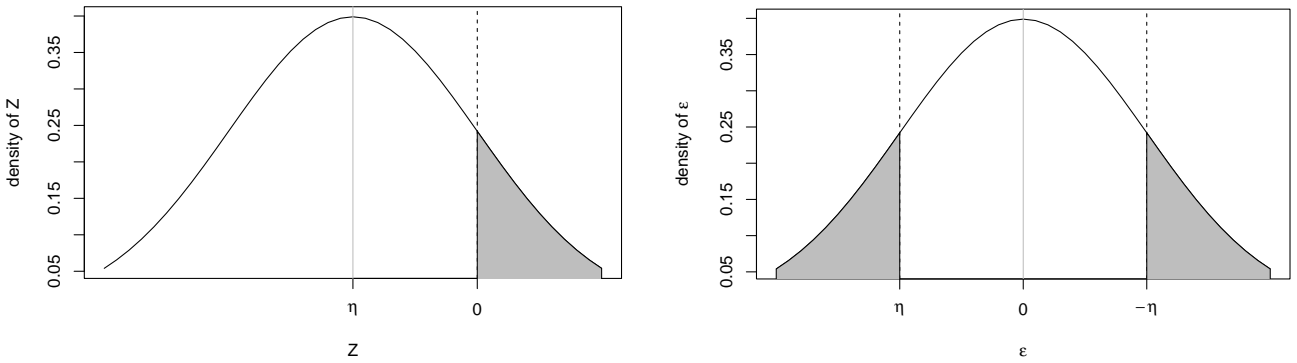
$$Z_i = \eta_i + \epsilon_i$$

where ϵ_i has a continuous, real valued distribution.

The derived random variable Y_i remains a deterministic function of Z_i given by

$$Y_i = \begin{cases} 1 & \text{if } Z_i \geq 0, \\ 0 & \text{if } Z_i < 0. \end{cases}$$

The shaded tail area shows $\mu_i = \mathbb{P}(Y_i = 1) = \mathbb{P}(Z_i \geq 0)$.



We have

$$\begin{aligned} \mu_i = \mathbb{P}(Y_i = 1) &= \mathbb{P}(Z_i \geq 0) \\ &= \mathbb{P}(\eta_i + \epsilon_i \geq 0) \\ &= \mathbb{P}(\epsilon_i \geq -\eta_i) \\ &= 1 - F(-\eta_i). \end{aligned}$$

where F is the cumulative distribution function of the error term ϵ_i .

The distribution of the latent variable Z (or equivalently the error term ϵ) is called the **tolerance distribution**.

If the tolerance distribution is symmetric, then we can simplify

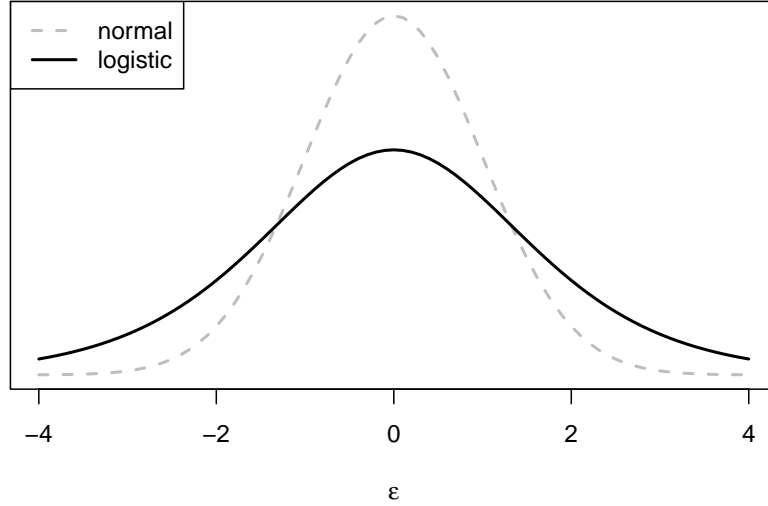
$$\mu = 1 - F(-\eta) = F(\eta)$$

and so the mean/response function is the cumulative distribution function F .

2.7.2 Example tolerance distributions

2.7.2.1 The logistic distribution

The logistic distribution has a symmetric bell-shaped density, like the normal distribution, but with longer tails.



The standard logistic probability density function is given by

$$f(x) = \frac{\exp(x)}{(1 + \exp(x))^2}$$

which gives the cumulative distribution function

$$F(x) = \frac{\exp(x)}{1 + \exp(x)}.$$

Unlike the normal cdf, the logistic cdf can be written in closed form. This allows us to interpret the parameters more easily.

A binary regression model using the logistic distribution as the tolerance distribution is called a **logistic regression model**.

It has **mean function** $h : \mathbb{R} \rightarrow (0, 1)$ with

$$\mu = h(\eta) = \frac{\exp(\eta)}{1 + \exp(\eta)}.$$

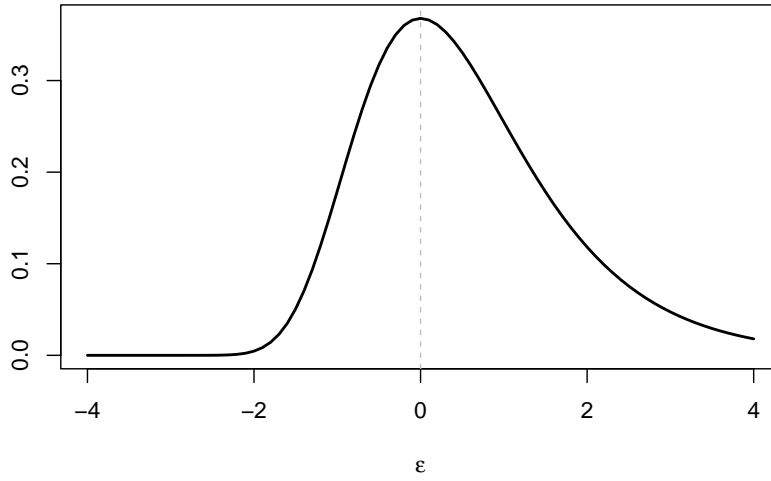
The corresponding **link function** $g : (0, 1) \rightarrow \mathbb{R}$ is given by

$$\eta = g(\mu) = h^{-1}(\mu) = \log\left(\frac{\mu}{1 - \mu}\right).$$

This link function is called the **logit link**.

2.7.2.2 The Gumbel distribution

The Gumbel distribution has a skewed density and is commonly used in applications of extreme value theory.



The probability density function of the standard Gumbel distribution is given by

$$f(x) = \exp(-x - \exp(-x))$$

which leads to the cumulative distribution function

$$F(x) = \exp\left(-\exp(-x)\right).$$

Recall, the Gumbel distribution is asymmetric, so $h(\eta) \neq F(\eta)$.

The formula for the mean function is

$$\mu = h(\eta) = 1 - F(-\eta) = 1 - \exp\left(-\exp(\eta)\right).$$

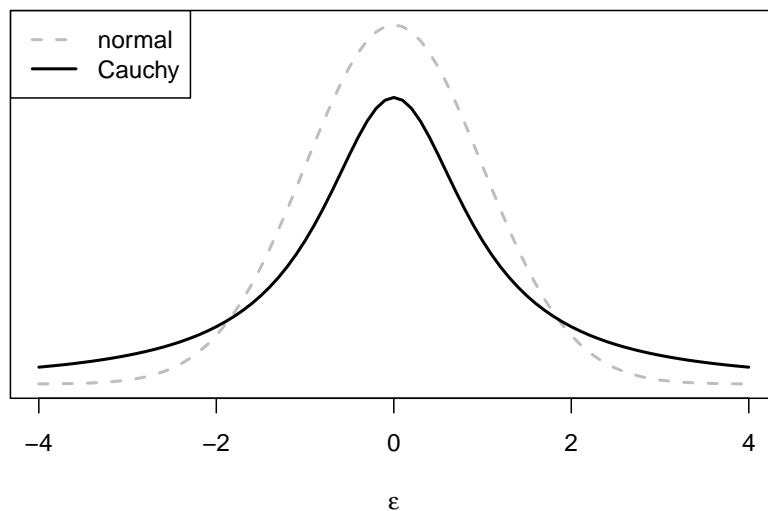
The link function is given by

$$\eta = g(\mu) = h^{-1}(\mu) = \log(-\log(1 - \mu)).$$

This link is called the **complementary log-log link function**.

2.7.2.3 The Cauchy distribution

The Cauchy distribution is the special name given to a t -distribution on 1 degree of freedom.



The probability density function of the Cauchy distribution

$$f(x) = \frac{1}{\pi (1 + x^2)}$$

which corresponds to the cumulative distribution function

$$F(x) = \frac{1}{\pi} \arctan(x) + \frac{1}{2}.$$

The mean and variance of the Cauchy distribution are not defined.

As the Cauchy distribution is symmetric, the mean function is equal to the cdf.

$$h(\eta) = \frac{1}{\pi} \arctan(\eta) + \frac{1}{2}.$$

The link function is

$$g(\mu) = \tan(\pi (\mu - 0.5)).$$

and is called the **cauchit link**.

2.7.2.4 Summary of tolerance distributions

Tolerance distribution	Link function	R link name
Normal	Probit	“probit”
Logistic	Logit	“logit”
Gumbel	Complementary log log	“cloglog”
Cauchy	Cauchit	“cauchit”

2.7.2.5 R illustration of the different link functions

In the following we define the various link functions for a binary GLM and fit a model for each link function to the turbines data.

```
library(GLMsData)
data(turbines)

links <- c(Probit="probit", Logit="logit", CLogLog="cloglog")

fit.turbines <- function(L) {
  glm(Fissures/Turbines ~ Hours, weights=Turbines,
      family=binomial(link=L), data=turbines)}

glm.out <- lapply(links, fit.turbines)
sapply(glm.out, coef)
```

	Probit	Logit	CLogLog
(Intercept)	-2.2758074623	-3.9235965551	-3.6032798443
Hours	0.0005783211	0.0009992372	0.0008104936

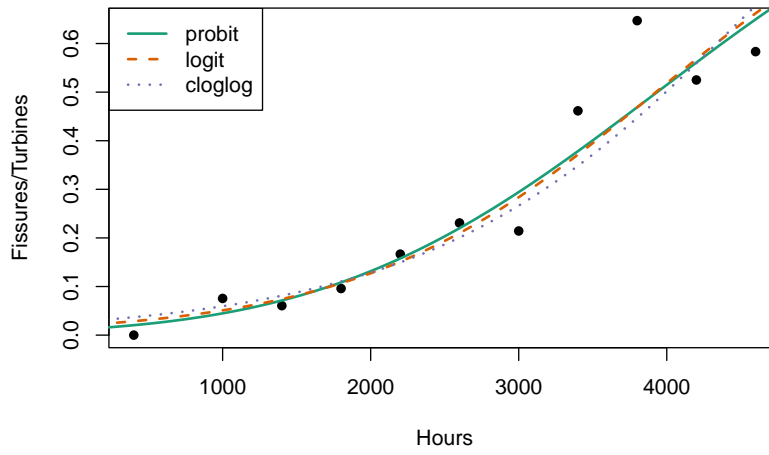
We note differences between coefficients, for example between the model using the probit link and the one using the logit link.

However, while the coefficients differ substantially, the models with different link functions may still produce very similar predictions.

```
newHours <- seq(0, 5000, length=100)

predict.turbines <- function(model) {
  predict(model, type="response", newdata=data.frame(Hours = newHours))}
predicted <- sapply(glm.out, predict.turbines)

plot(Fissures/Turbines ~ Hours, data=turbines, pch=16)
library(RColorBrewer)
palette <- brewer.pal(3, "Dark2")
matlines(newHours, predicted, lty=1:3, lwd=2, col=palette)
legend("topleft", lty=1:3, lwd=2, col=palette, legend=links)
```



Exercise 5

1. For each of the tolerance distributions presented verify
 - $F'(x) = f(x)$ (the derivative of its cdf is its pdf);
 - $g(h(\eta)) = \eta$ and $h(g(\mu)) = \mu$ (the link function is the inverse of the mean function).
2. Repeat the experiment from Section 2.7.2.5 but this time include the cauchit link function.