Tutorial 7

STAT3015/4030/7030 Generalised Linear Modelling

The Australian National University

Week 7, 2017

Overview

Identify probability distribution of Y

(a) Let Y be the number of iPhone owners out of a random sample of 100 that download an app today.

Identify probability distribution of Y

(a) Let Y be the number of iPhone owners out of a random sample of 100 that download an app today.

- Binomial. This is a discrete variable and can be viewed as counting the number of successes/failures in 100 independent trials.

Identify probability distribution of Y

(a) Let Y be the number of iPhone owners out of a random sample of 100 that download an app today.

- Binomial. This is a discrete variable and can be viewed as counting the number of successes/failures in 100 independent trials.

(b) Let Y be the number of babies born on a single day in Canberra.

Identify probability distribution of Y

(a) Let Y be the number of iPhone owners out of a random sample of 100 that download an app today.

- Binomial. This is a discrete variable and can be viewed as counting the number of successes/failures in 100 independent trials.

(b) Let Y be the number of babies born on a single day in Canberra.

- Poisson. This is a discrete variable counting the number of events in a fixed interval.

Identify probability distribution of Y

(c) Let Y be the initial weight (in kilograms) of a randomly selected male enrolling in the Biggest Loser diet.

Identify probability distribution of Y

(c) Let Y be the initial weight (in kilograms) of a randomly selected male enrolling in the Biggest Loser diet.

- Normal. Reasonable to think that weights are symmetrically distributed around a mean and more likely to be close to that mean than far from it. Distribution could be skewed as well.

Identify probability distribution of Y

(c) Let Y be the initial weight (in kilograms) of a randomly selected male enrolling in the Biggest Loser diet.

- Normal. Reasonable to think that weights are symmetrically distributed around a mean and more likely to be close to that mean than far from it. Distribution could be skewed as well.

(d) Let Y be the initial weight (in kilograms) of a randomly selected person on the Biggest Loser diet.

Identify probability distribution of Y

(c) Let Y be the initial weight (in kilograms) of a randomly selected male enrolling in the Biggest Loser diet.

- Normal. Reasonable to think that weights are symmetrically distributed around a mean and more likely to be close to that mean than far from it. Distribution could be skewed as well.

(d) Let Y be the initial weight (in kilograms) of a randomly selected person on the Biggest Loser diet.

- Other. Would expect two modes, one for men and one for women.

Identify probability distribution of Y

(e) Let Y be 1 if a randomly selected person who saw the movie The Wolverine liked it, and 0 if not.

Identify probability distribution of Y

(e) Let Y be 1 if a randomly selected person who saw the movie The Wolverine liked it, and 0 if not.

- Bernoulli. Y is either 0 or 1.

Identify probability distribution of Y

(e) Let Y be 1 if a randomly selected person who saw the movie The Wolverine liked it, and 0 if not.

- Bernoulli. Y is either 0 or 1.

(f) Let Y be 1 if you flip a coin and it lands tails, and 2 if it lands heads.

Identify probability distribution of Y

(e) Let Y be 1 if a randomly selected person who saw the movie The Wolverine liked it, and 0 if not.

- Bernoulli. Y is either 0 or 1.

(f) Let Y be 1 if you flip a coin and it lands tails, and 2 if it lands heads.

- Other. Y follows a Bernoulli random variable plus a constant of 1.

Choose the appropriate link function

(a) Suppose a probability distribution has an unknown mean μ that is restricted to be greater than 1. Why is $g(\mu) = \log(\mu)$ NOT a sensible link function? What would be a more reasonable link function?

Choose the appropriate link function

(a) Suppose a probability distribution has an unknown mean μ that is restricted to be greater than 1. Why is $g(\mu) = \log(\mu)$ NOT a sensible link function? What would be a more reasonable link function?

If μ is always greater than 1, then $g(\mu) = \log(\mu) > 0$. This is not sensible because the range of $X\boldsymbol{\beta}$ is the entire real line and we would like the ranges of $g(\mu)$ and $X\boldsymbol{\beta}$ to match up.

A more reasonable link function would be $\log(\mu - 1)$.

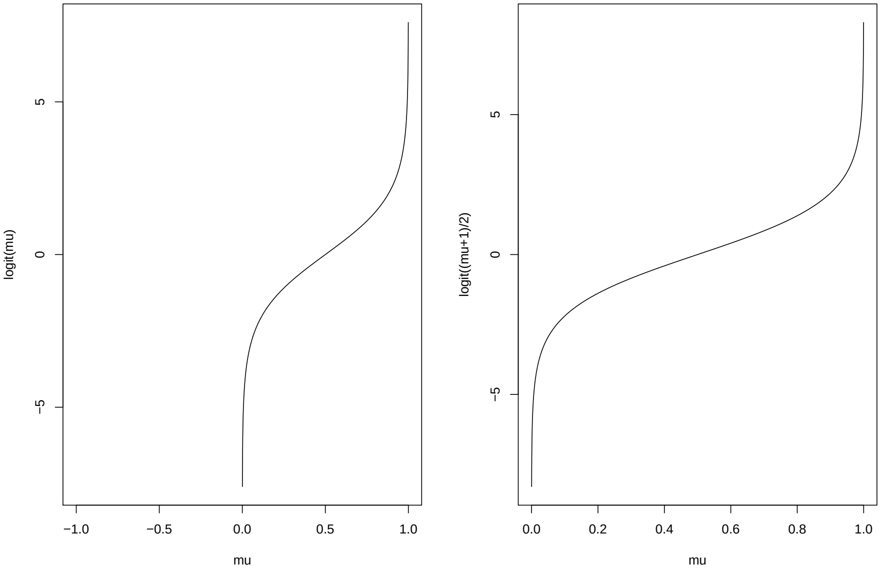
Choose the appropriate link function

(b) Suppose that a probability distribution has an unknown mean μ that is restricted to be between -1 and 1. What would be a reasonable link function for μ?

Choose the appropriate link function

(b) Suppose that a probability distribution has an unknown mean μ that is restricted to be between -1 and 1. What would be a reasonable link function for μ?

A reasonable link function would be $g(\mu) = \text{logit}\left(\frac{\mu+1}{2}\right)$


logit(mu)
5
0
−5
−1.0
−0.5
0.0
0.5
1.0
mu
logit((mu+1)/2)
5
0
−5
0.0
0.2
0.4
0.6
0.8
1.0
mu

Delta Method

The delta method is a statistical approach to derive an approximate probability distribution for a function of an asymptotically normal estimator using the Taylor series approximation.

If a sequence of random variables $Y_1, \cdots, Y_n$ satisfying

$$\sqrt{n}(Y_i - \theta) \xrightarrow{D} \mathcal{N}(0, \sigma^2),$$

where θ and σ^2 are finite valued constants, then

$$\sqrt{n}\,(g(Y_i) - g(\theta)) \xrightarrow{D} \mathcal{N}(0, [g'(\theta)]^2\sigma^2).$$

Confidence interval for $g^{-1}(X^T\boldsymbol{\beta})$

When we want to calculate a 95% confidence interval for a function of the parameters $\boldsymbol{\beta}$, say $\mu = g^{-1}(X^T\boldsymbol{\beta})$, we can firstly compute a confidence interval for $X^T\boldsymbol{\beta}$ as $\{L, U\}$, and then apply the function $g^{-1}()$ to both bounds L and U.

The desired confidence interval is given by $\{g^{-1}(L), g^{-1}(U)\}$. Part (c) of Question 2 involves this method.

(Proof is on Page 41 of the lecture brick on Wattle.)

A Binomial example

Example 1 of the lecture notes considers anaesthetic data where Y represents proportions of patients responding to the stimulus within each group. With the fact that Y follows a Binomial$(n, p)/n$=n distribution and the table on Page 33:

Distribution	$E(Y) = \mu$	$Var(Y)$	$b(\mu)$	$V(\mu)$	ϕ
Normal(μ, σ^2)	μ	σ^2	μ	1	σ^2
Binomial(n, p)	np	$np(1-p)$	$\log\left(\frac{\mu}{n-\mu}\right)$	$\frac{\mu(n-\mu)}{n}$	1
Binomial$(n, p)/n$	p	$\frac{p(1-p)}{n}$	$\log\left(\frac{\mu}{1-\mu}\right)$	$\mu(1-\mu)$	$\frac{1}{n}$
Poisson(λ)	λ	λ	$\log(\mu)$	μ	1
Poisson$(\lambda T)/T$	λ	$\frac{\lambda}{T}$	$\log(\mu)$	μ	$\frac{1}{T}$
Gamma(α, β)	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$	$-\mu^{-1}$	μ^2	$\frac{1}{\alpha}$

we want to find $Var(g(Y))$.

A Binomial example (cont.)

- We have assumed that $g(\mu) = X^T\boldsymbol{\beta}$ and want to estimate $\boldsymbol{\beta}$. However, we don't know the population mean μ.
- We have a unbiased estimator for μ, namely the sample mean (denoted by Y here).
- We can use the GLM technique, regress $g(Y)$ on observed covariates; OR, use a weighted least-squares model.

Following the Delta Method, we have

$$\begin{aligned} Var\{g(Y)\} &\approx Var(Y)g'(\mu)^2 = \phi Var(\mu)g'(\mu)^2 \\ &= \frac{1}{n(\mu(1-\mu))}. \end{aligned}$$

(calculation left as your exercise)

Iteratively Re-weighted Least-Squares (IRLS) algorithm

We apply the IRLS algorithm as follows:

1. Set $\hat{Y} = Y$
2. Apply a weight according to $\omega_i^2 = 1/Var(g(Y) \approx n(\mu(1-\mu))$
3. Perform a weighted regression of $g(Y)$ on covariates to obtain estimates $\hat{\beta}$
4. Calculate new fitted values $\hat{Y} = g^{-1}(X^T\hat{\beta})$ and repeat steps 2 and 3 till successive values of the parameter estimates (or equivalently values of the weights) do not change substantially.