Stat314/461Term 4:
Logistic Regression from a Bayesian viewpoint

Patrick Graham

September, 2021

Logistic regression

Regression modelling for a binary response variable

$$Y_i \sim \text{Bernoulli}(\theta_i);\;\; i = 1, \ldots, N$$

$$\log\left(\frac{\theta_i}{(1-\theta_i)}\right) = \beta_0 + X_{1i}\beta_1 + X_{2i}\beta_2 + X_{3i}\beta_3 + ...$$

or equivalently

$$\theta_i = \frac{\exp(\beta_0 + X_{1i}\beta_1 + X_{2i}\beta_2 + X_{3i}\beta_3 + ...)}{1 + \exp(\beta_0 + X_{1i}\beta_1 + X_{2i}\beta_2 + X_{3i}\beta_3 + ...)}$$

This model ensures $0 < \theta_i < 1$, for all i, or, equivalently, for all combinations of the X's.

We are modelling the relationship of the success probabilities to the covariates $X_1, X_2, \ldots ...$ i.e "regression."

Interpretation of logistic regression coefficients (i)

Let

$$\theta = \frac{\exp(\beta_0 + X_1\beta_1 + X_2\beta_2 + X_3\beta_3 + ...)}{1 + \exp(\beta_0 + X_1\beta_1 + X_2\beta_2 + X_3\beta_3 + ...)}$$

$$\log\left(\frac{\theta}{1-\theta}\right) = \beta_0 + X_1\beta_1 + X_2\beta_2 + X_3\beta_3 + ... \tag{1}$$

Effect on $\mathrm{logit}(\theta)$ on changing from, say, X_1 to $X_1 + 1$ is

$$\begin{aligned} \beta_0 + &(X_1 + 1)\beta_1 + X_2\beta_2 + X_3\beta_3 + ...) && (2)\\ &- \beta_0 + (X_1)\beta_1 + X_2\beta_2 + X_3\beta_3 + ...) && (3)\\ &= \beta_1 && (4) \end{aligned}$$

Interpretation of logistic regression coffecients (ii)

$$\log\left(\frac{\theta}{1-\theta}|(X_1+1)\right) - \log\left(\frac{\theta}{1-\theta}|X_1\right) = \beta_1 \tag{5}$$

Hence

$$\exp\left[\log\left(\frac{\theta}{1-\theta}|(X_1+1)\right) - \log\left(\frac{\theta}{1-\theta}|X_1\right)\right] = \exp(\beta_1) \tag{6}$$

$$\frac{[\theta/(1-\theta)]\,|(X_1+1)}{[\theta/(1-\theta)]\,|X_1} = \exp(\beta_1) \tag{7}$$

exponentiated logistic regression coefficients are interpretable as odds ratios (except for intercept)

$\mathrm{invlogit}(\beta_0)$ is the probability of the event when $X_1 = X_2 = \ldots, = 0$

Example

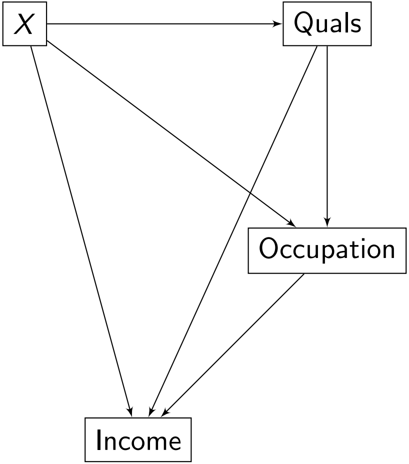
Example: $\Pr(income > \$1250/week | age, sex, quals)$
- from the income survey data

$$p(\beta|\mathbf{Y}) \propto p(\beta) \prod_i \theta_i^{Y_i}(1-\theta_i)^{(1-Y_i)}$$

$$\theta_i = \frac{\exp(\beta_0 + X1i\beta_1 + X_{2i}\beta_2 + X_{3i}\beta_3 + ...)}{1 + \exp(\beta_0 + X_{1i}\beta_1 + X_{2i}\beta_2 + X_{3i}\beta_3 + ...)}$$

and X_1=age, X_2 = sex, X_3, X_4, X_5, X_6 =indicators for qualification categories.

Directed Acyclic Graph for Income model



Likelihood for the logistic regression example (i)

X fixed; Y random (or analysing conditional on X)
really just a sequence of Bernoulli trials but with different success probabilities on each trial.
Don't panic!
Let $\mathbf{X}_i = (1, X_{1i}, X_{2i}, \ldots)'$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \ldots)'$

$\mathbf{X}'\boldsymbol{\beta} = \sum_{j=0}^{i=K} Xj_i\beta_j$ where $X_{0i} = 1$

$$\theta_i = \mathrm{expit}(X_i'\beta) = \mathrm{invlogit}(X_i'\beta) = \frac{\exp(X_i'\beta)}{1 + \exp(X_i'\beta)}.$$

Likelihood for the logistic regression example (ii)

If

$$\log\left(\frac{\theta_i}{(1-\theta_i)}\right) = X_i'\beta$$

$$\theta_i = \mathrm{invlogit}(X_i'\beta) \Rightarrow (1-\theta_i) = (1+\exp(X_i'\beta))^{-1}$$

and

$$\begin{aligned} p(\mathbf{Y}|\mathbf{X},\boldsymbol{\beta}) &= \prod_i \theta_i^{Y_i}(1-\theta_i)^{(1-Y_i)} \\ &= \prod_i \left(\frac{\theta_i}{1-\theta_i}\right)^{Y_i}(1-\theta_i) \end{aligned}$$

Likelihood for the logistic regression example (ii)

If

$$\log\left(\frac{\theta_i}{(1-\theta_i)}\right) = X_i'\beta$$

$$\theta_i = \mathrm{invlogit}(X_i'\beta) \Rightarrow (1-\theta_i) = (1+\exp(X_i'\beta))^{-1}$$

and

$$\begin{aligned} p(\mathbf{Y}|\mathbf{X},\boldsymbol{\beta}) &= \prod_i \theta_i^{Y_i}(1-\theta_i)^{(1-Y_i)} \\ &= \prod_i \left(\frac{\theta_i}{1-\theta_i}\right)^{Y_i}(1-\theta_i) \end{aligned}$$

$$\begin{aligned} \log(p(\mathbf{Y}|\mathbf{X},\boldsymbol{\beta})) &= \sum_i Y_i \log\left(\frac{\theta_i}{1-\theta_i}\right) + \log((1+\exp(X_i'\beta))^{-1}) \\ &= \sum_i Y_i \times X_i'\beta - \log(1+\exp(X_i'\beta)) \end{aligned}$$

Prior for logistic regression coefficients (ii)

- We have a vector of parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \ldots)'$. Even if we model these parameters as *a priori* independent they will be correlated in the posterior: (To get the idea try

```
logitmodel <- glm(highincome ~ hoursfactor + sexfactor +
qualfactor,family=binomial(link="logit") )
vcov(logitmodel)
```

which returns the variance *matrix* for the mle.

Prior for logistic regression coefficients (ii)

- It is convenient to treat $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \ldots)'$. as a single entity in computations.
- If we do regard the components of $\boldsymbol{\beta}$ as *a priori* independent and adopt a normal prior we are saying

$$p(\boldsymbol{\beta}) = \prod_{j=0}^{j=K} normal(\beta_j | \mu_j, \sigma_j^2) \tag{8}$$

This is equivalent to saying $\boldsymbol{\beta} \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Omega})$ where

$$\boldsymbol{\mu} = (\mu_0, \mu_1, \ldots \mu_K)' \text{ and } \boldsymbol{\Omega} = \begin{bmatrix} \sigma_0^2 & 0 & 0 & \ldots \\ 0 & \sigma_1^2 & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0\ldots & \sigma_K^2 \end{bmatrix}$$

Prior for logistic coefficients (iii)

- In general the multivariate normal (MVN) distribution is characterised by a mean *vector* and variance *matrix*.
- Off-diagonal elements in the variance matrix are *covariances*. Non-zero covariances indicate correlation between the corresponding elements of the vector being modelled

$$\begin{bmatrix} \sigma_0^2 & \sigma_{01} & \sigma_{02} & \ldots\sigma_{0K} \\ \sigma_{01} & \sigma_1^2 & \sigma_{12} & \ldots\sigma_{1K} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{0K} & \sigma_{1K} & \sigma_{2K}\cdots & \sigma_K^2 \end{bmatrix}$$

$\rho_{12} = \frac{\sigma_{12}}{\sigma_1\sigma_2}$ etc

Prior for logistic coefficients (iv)

If we assume *a priori* independence we can construct the joint prior for $\boldsymbol{\beta}$ by considering each component separately.

- $\pi_0 = \text{invlogit}(\beta_0)$ is the probability of the event if the $X's$ are all zero. Could
  1. set , say, 95% prior limits on π_0
  2. transform to logit scale
  3. construct a normal that matches the transformed limits: $\mu_0 = 0.5(\text{logit}(\pi_{0,low}) + \text{logit}(\pi_{0,high}))$; variance s.t transformed limits are 2.5% and 97.5% quantiles of a normal with mean μ_0
- $\gamma_j = \exp(\beta_j)$ is the odds ratio for the effect of $Xj, j = 1, \ldots, K$.
  1. set 95% prior limits on γ_j (remember 1 is the “null” value)
  2. transform to log-scale
  3. construct a normal that matches the transformed limits (2.5% and 97.5% quantiles of a normal.)

Importance sampling for logistic regression

- If we can:
  1. write a function to compute the log-likelihood for any value of $\boldsymbol{\beta}$
  2. write a function to compute the log prior density for any value of $\boldsymbol{\beta}$

  we can compute the log unnormalised posterior density at any setting of the parameters.
- We can obtain a multivariate normal approximation to the posterior by maximizing the log unnormalised posterior density wrt β to find the posterior mode and the curvature of the log-posterior at the model. The approximating MVN is centred at the posterior mode and and has variance based on the inverse of the curvature.
- Standard optimisation functions can find these quantities. The laplace function from the LearnBayes package is useful; bayesglm from the ARM package can also be used.

Example: logistic regression for analysis of effect of sex and tertiary education on probability of high income

- see logistic regression importance sampling R code for an application based on the income survey data `logisticregression_importance2021.r`