

Bayesian Logistic Regression

Let Y be an indicator for presence of disease $\rightarrow Y | p \sim \text{Bin}(p, 1)$

Logistic regression models the log-ODDS (also called logit) of the mean outcome using a linear predictor.

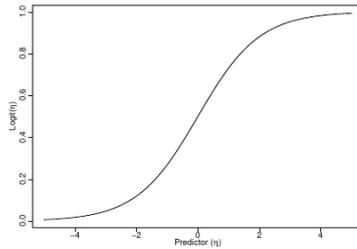
$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$

$$p = P(D|x) = \frac{e^\eta}{1 + e^\eta}$$

$$\eta = \text{logit}(p) = \log\left(\frac{p}{1-p}\right)$$

where $\eta = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$

is called the linear predictor



$$\eta = \beta_0 + \beta_1 x, \text{ with } x = \begin{cases} 0 & \text{if subject is not exposed} \\ 1 & \text{if subject is exposed} \end{cases}$$

$$\text{If } x = 0 \Rightarrow \eta = \log\left(\frac{p}{1-p} | x = 0\right) = \beta_0$$

$$\text{If } x = 1 \Rightarrow \eta = \log\left(\frac{p}{1-p} | x = 1\right) = \beta_0 + \beta_1$$

$$\beta_1 = \log\left(\frac{p}{1-p} | x = 1\right) - \log\left(\frac{p}{1-p} | x = 0\right) = \log(OR)$$

$$OR = \exp\{\beta_1\}$$

In a case-control study β_0 is not interpretable. But the estimate of $\log(OR)$ can be interpreted correctly, given the symmetry of roles of disease/exposure

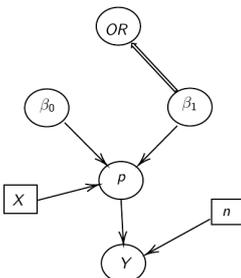
H_0 : $OR = 1$; $\log(OR) = 0$; No association

H_1 : $OR \neq 1$; $\log(OR) \neq 0$; Significant association

The Frequentist approach to test the above is to fit a logistic regression in R using the function `glm()` and assess the OR.

In Bayesian formulation we need to identify the parameters in our prior model, and the parameters we wish to estimate to form a posterior distribution to assess the OR. We can compute the probability that $OR > 1$, < 1 or $\neq 1$ using a small range.

Ex. Below we model 2 parameters, β_0 and β_1 , in a DAG model:



Because there's no constraint on the beta regression parameters, it is common to assume a normal distribution.

The magnitude of variance reflects the amount of uncertainty

In the below code:

- We loop to specify each observation
- $Y.D[i] =$
 - 1 if subject i has the disease
 - 0 if subject i does not have the disease
- $X.E[i] =$
 - 1 if subject i has been exposed
 - 0 if subject i has not been exposed

```
# Analysis with JAGS
library(rjags)

model.1 <- "model{
### data model
for(i in 1:N){
  Y.D[i] ~ dbin(p[i], 1)
  logit(p[i]) <- beta_0 + beta_1*X.E[i]
}
OR <- exp(beta_1)
pos.prob <- step(beta_1)
### prior
beta_1 ~ dnorm(0,0.0001)
beta_0 ~ dnorm(0,0.0001)
  0 i th subject unexposed
}"

# In R Data stored as a list
data.1 <- list(N = 331, X.E = X.E, Y.D = Y.D)

# Compile model (data is part of it!); `adapt` for 2,000 samples
model_odds <- jags.model(textConnection(model.1), data = data.1,
n.adapt = 2000)

update(model_odds, n.iter = 5000) # 5,000 burn-in samples

# Get 10,000 samples from the posterior distribution of OR, beta_0, beta_1
test_odds <- coda.samples(model_odds, c("OR", "beta_1", "beta_0"), n.iter = 10000)

plot(test_odds)

summary(test_odds)
```

In the output, we'll often look for the point estimate to be the posterior mean. However, when posterior density is skewed, posterior median is sometimes preferable.

Asymptotic Approximation

When the sample size is large enough, one can use an asymptotic approximation of the posterior distribution of the parameters.

When the sampling distribution is a member of the exponential family (Normal, Binomial, Poisson, or Gamma), then the posterior distribution of the parameters is (approximately) normally distributed.

$$\beta|y_1, \dots, y_n \sim N\left(\hat{\beta}, I(\hat{\beta})^{-1}\right) \quad [1]$$

$\hat{\beta}$ is the ML (maximum likelihood) estimate of β

$I(\hat{\beta})^{-1}$ is the Fisher Information matrix, evaluated in the ML estimate of β

Based on the Maximum likelihood theory, asymptotically:

$$\hat{\beta}(y_1, \dots, y_n)|\beta \sim N\left(\beta, I(\beta)^{-1}\right) \quad [2]$$

We don't need a prior in [1] because we have a lot of data. Results in [1] and [2] are close because the prior is weighted low because of the high sample size.

Gibbs Sampling for Logistic Regression

Θ the vector of parameters $\Theta = (\beta_0, \beta_1)$

The Likelihood function is $P(y_1, \dots, y_n|\Theta)$

Conditional distributions: Compute the conditional distributions of each parameter:

$$P(\beta_1|y_1, \dots, y_n, \beta_0) \text{ AND } P(\beta_0|y_1, \dots, y_n, \beta_1)$$

Initialization (Step 0): have initial values β_{00}, β_{10}

Sampling scheme (Step 1):

- Sample β_{01} from $P(\beta_0|y_1, \dots, y_n, \beta_{10})$.
- Sample β_{11} from $P(\beta_1|y_1, \dots, y_n, \beta_{01})$.

Repeat (Step 2): Replace β_{00}, β_{10} by β_{01}, β_{11} and repeat **Step 1**

Repeat **Step 1** and **Step 2** $s + m$ times.

Posterior Inference

Obtain $s + m$ sample for each parameter

Throw away the first s values (burn-in)

Use the m samples with Monte Carlo algorithms to carry out inference:

1. Point Estimates - Posterior means or medians

$$E(\beta_j) = \frac{1}{m} \sum_{h=s+1}^{m+s} \beta_{jh}$$

2. Transformations - Any function of the parameters

$$E(g(\beta_j)) = \sum_{h=s+1}^{m+s} g(\beta_{jh}) / m$$

3. Posterior Density - Density estimation methods to estimate the posterior density of parameters

4. Tail probabilities

$$P(\beta_j > b_0) = \frac{1}{m} \sum_{h=s+1}^{m+s} \{\beta_{jh} > b_0\}$$

Marginal and Conditional Independence

Two random variables, X and Y, are independent (denoted by $X \perp Y$) if and only if:

$P(X, Y) = P(X)P(Y)$; Joint p.m.f. for joint density = Marginal p.m.f. or marginal density

Given three random variables, X, Y and Z, are conditionally independent given Z (denoted by $Y \perp X | Z$) if and only if:

$$P(X, Y | Z) = P(X | Z) P(Y | Z)$$

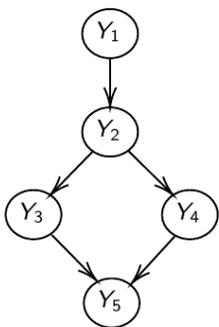
There is no relation between marginal and conditional independence (Simpson Paradox)

The directed graph specifies conditional and marginal independence through two properties; The **Local Markov Property (LMP)** specifies that we need to specify only the distribution of nodes given their parents. The **Global Markov Property (GMP)** is used internally to derive the conditional distributions of each node given all others.

Local Markov Property (LMP)

Descendants of Y [D(Y)] are all nodes reached from Y with a directed path

Non-descendants of Y [NS(Y)] - all nodes minus the descendant nodes



- **Local Markov property (LMP):**

$$Y \perp \{ND(Y) - \text{par}[Y]\} | \text{par}[Y]$$

- $Y_3 \perp \{Y_1, Y_4\} | Y_2$; $Y_5 \perp \{Y_1, Y_2\} | Y_3, Y_4$

- **Factorization of joint probability**

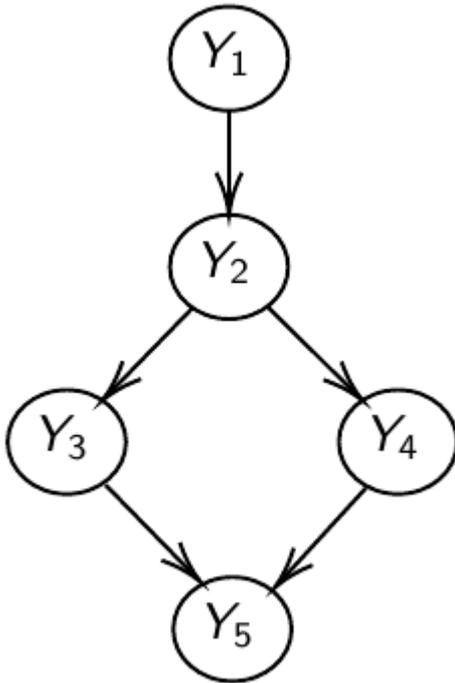
Factorization Identity

$$P(V) = \prod_{v \in \mathcal{G}} P(v | \text{par}[v])$$

Global Markov Property (GMP)

The node is independent of everything else in the network given its Markov Blanket.

For each node, list the **Markov blanket**: Markov blanket of A: the parents of A, the children of A and the parents of the children of A



Y_1 independent of Y_3, Y_4, Y_5 , given Y_2
 $MB(Y_1) = \{ Y_2 \}$; $MB(Y_2) = \{ Y_1, Y_3, Y_4 \}$

The LMP is used to specify the priors and the likelihood
 The GMP is used to generate the conditional distributions to run the Gibbs Sampling

Bayesian Hypothesis Testing: Prior Odds

On each hypothesis we have a prior probability:

$$P(H_0) = P(M_0) \text{ and } P(H_a) = P(M_1)$$

Use the data to compute the posterior probability of each hypothesis:

$$\frac{P(M_0|y_1, \dots, y_n)}{P(M_1|y_1, \dots, y_n)} = \frac{\frac{P(y_1, \dots, y_n|M_0)P(M_0)}{P(y_1, \dots, y_n)}}{\frac{P(y_1, \dots, y_n|M_1)P(M_1)}{P(y_1, \dots, y_n)}} = \frac{P(y_1, \dots, y_n|M_0)P(M_0)}{P(y_1, \dots, y_n|M_1)P(M_1)}$$

This equates to:

Posterior ODDs = Bayes Factor * Prior Odds

We accept the hypothesis with maximum posterior probability (0 - 1 loss)

The Bayes Factor is the ratio of the likelihood functions computed for models M_0 and M_1

If $P(M_0) = P(M_1)$ then the posterior probability of M_1 is larger than the posterior probability of M_0

We can also assign weights to adjust the probability of making errors:

Loss	Accept M_0	Accept M_1
M_0 true	0	α
M_1 true	β	0

Where alpha would be the loss incurred if M_1 is accepted when M_0 is true and beta is the loss incurred if M_0 is accepted when M_1 is true.

Revision #9

Created 27 January 2023 17:33:40 by Elkip

Updated 10 February 2023 19:07:55 by Elkip