ADVANCED BAYESIAN MODELING

Logistic Regression Modeling

Data Model

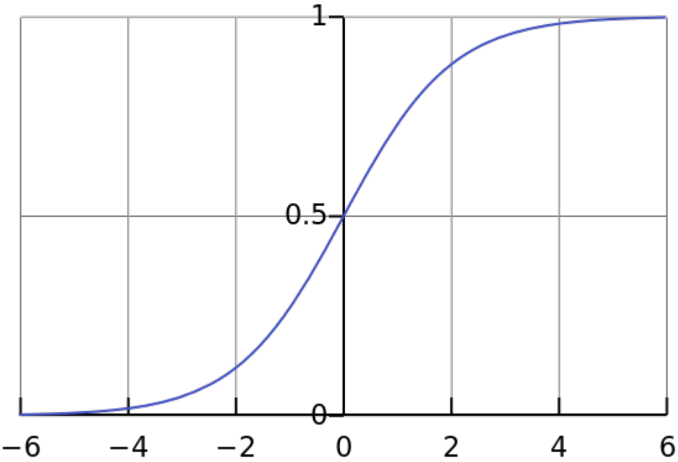
$$X_i \;=\; (x_{i1}, \ldots, x_{ik}), \quad y_i \qquad\qquad i = 1, \ldots, n$$

X has k explanatory variables (usually including an intercept)

$$y_i \mid \beta, X_i \quad \sim \quad \text{indep. Bin}(n_i, p_i) \qquad\qquad n_i \text{ fixed and known}$$

$$\text{logit}(p_i) \;=\; X_i\beta \qquad\qquad p_i \;=\; \text{logit}^{-1}(X_i\beta) \;=\; \frac{e^{X_i\beta}}{1+e^{X_i\beta}}$$

The inverse logit:



From: Logistic regression. (2017, July 8). In *Wikipedia, The Free Encyclopedia*. Retrieved August 4, 2017 from `https://en.wikipedia.org/w/index.php?title=Logistic_regression&oldid=789649742`

Priors

In general logistic regression, conjugate priors are not very useful.

Flat priors are often a poor choice (later).

Consider instead proper priors from a standard (non-conjugate) distribution, but reasonably vague ...

Suggestion for weakly informative priors (see BDA3, Sec. 16.3):

1. Center/standardize explanatory variables:
   - Center indicator variables that don't define the intercept.
   - Center and scale quantitative variables to 0.5 sample standard deviation.
   - Use these centered/scaled variables to form any higher-order terms (e.g., interactions).

2. Choose independent scaled-t_1 priors for coefficients in β:
   - $t_1\big(0, 2.5^2\big)$ for variables not defining the intercept
   - $t_1\big(0, 10^2\big)$ for variables defining the intercept

The $t_1(\mu, \sigma^2)$ distribution, also called the *Cauchy*, is a generalization of the t-distribution with one degree of freedom.

Like the normal, it is symmetric about μ and unimodal, but has heavier tails, which may be useful if a few coefficients turn out to be large.

In JAGS:

```
dt(mu, sigmasqinv, 1)
```

Note: Parameterization uses *inverse* of σ^2.

Why not use flat priors?

Because, in certain cases, they would lead to improper posteriors.

Occurs when there is *separation*: A linear combination exists that predicts y perfectly (except possibly along some boundary).

Can occur even when the columns of X are linearly independent.

Details: BDA3, Sec. 16.3

Overdispersion

Mean and variance of binomial are closely linked:

$$\mathrm{E}(y_i \mid \beta, X_i) = n_i\, p_i \qquad\qquad \mathrm{var}(y_i \mid \beta, X_i) = n_i\, p_i(1 - p_i)$$

Common problem: Data exhibit variances that are incompatible with the means – typically larger than they should be: **overdispersion**.

Possible causes:

- Missing explanatory variables
- Dependent (underlying) Bernoulli variables
- Different probabilities in (underlying) Bernoulli distributions

The **chi-square discrepancy** measures overdispersion:

$$T(y, \theta, X) \;=\; \sum_{i=1}^{n} \frac{(y_i - n_i p_i)^2}{n_i p_i (1 - p_i)}$$

(When the model is correct and the n_is are large enough, it has an approximate χ^2_n distribution.)

When there is overdispersion, it tends to be larger.

Can be used to compute a posterior predictive p-value for overdispersion:

$$\Pr\bigl(T(y^{\textsf{rep}}, \theta, X) \;\geq\; T(y, \theta, X) \mid y\bigr)$$

One possible way to account for overdispersion:

Add a separate random effect to the linear combination of each observation.

$$\mathrm{logit}(p_i) \;=\; X_i\beta + \epsilon_i \qquad\qquad \epsilon_i \;\sim\; \mathrm{iid}\ \mathrm{N}(0, \sigma_\epsilon^2)$$

Then add a reasonably noninformative prior on σ_ϵ^2 (e.g., flat on σ_ϵ).

Remark:

Overdispersion is generally a detectable problem only if the n_is are not too small.

In particular, it is irrelevant to a purely binary response ($n_i = 1$), in which case the model should be checked in other ways (e.g., trying additional explanatory variables).