Logistic regression: Miscellaneous topics

Patrick Breheny

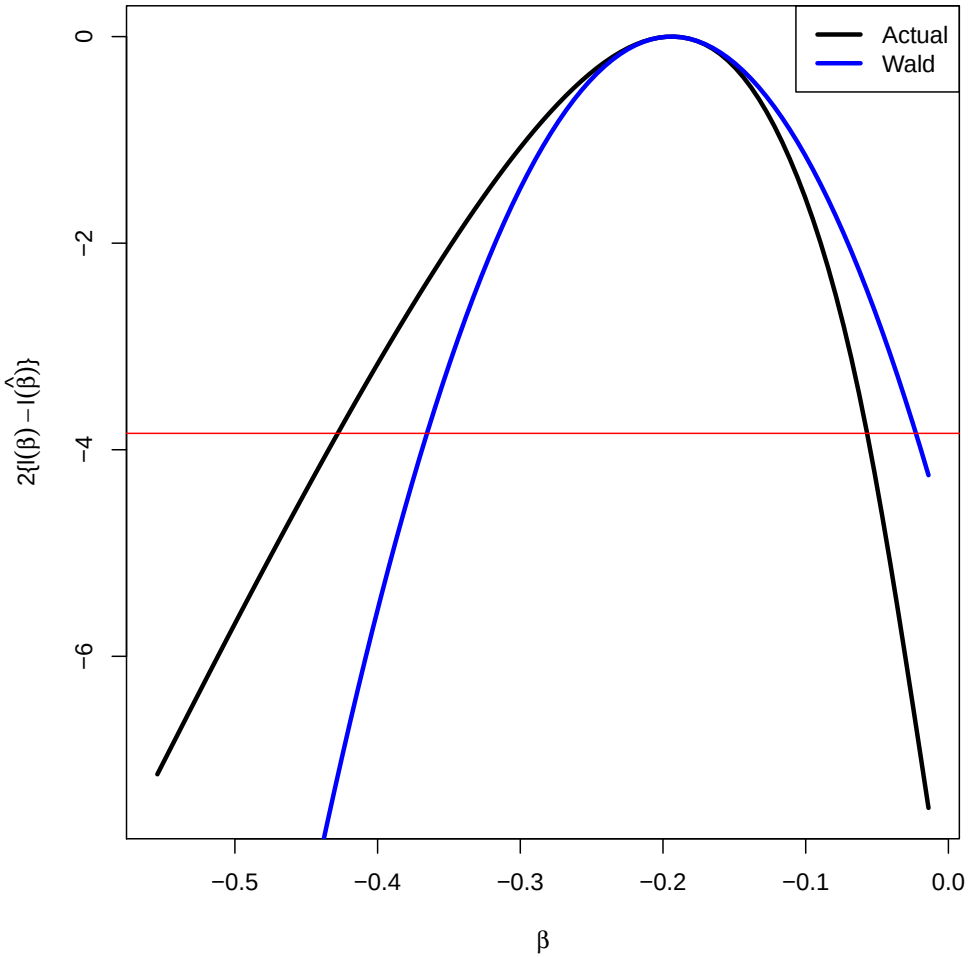
April 11

Introduction

- We have covered two approaches to inference for GLMs: the Wald approach and the likelihood ratio approach
- I claimed that the likelihood ratio approach is better; we will now take a closer look at the two approaches

Wald vs. Likelihood ratio

Estimating the effect of age upon survival for females in the Donner party:

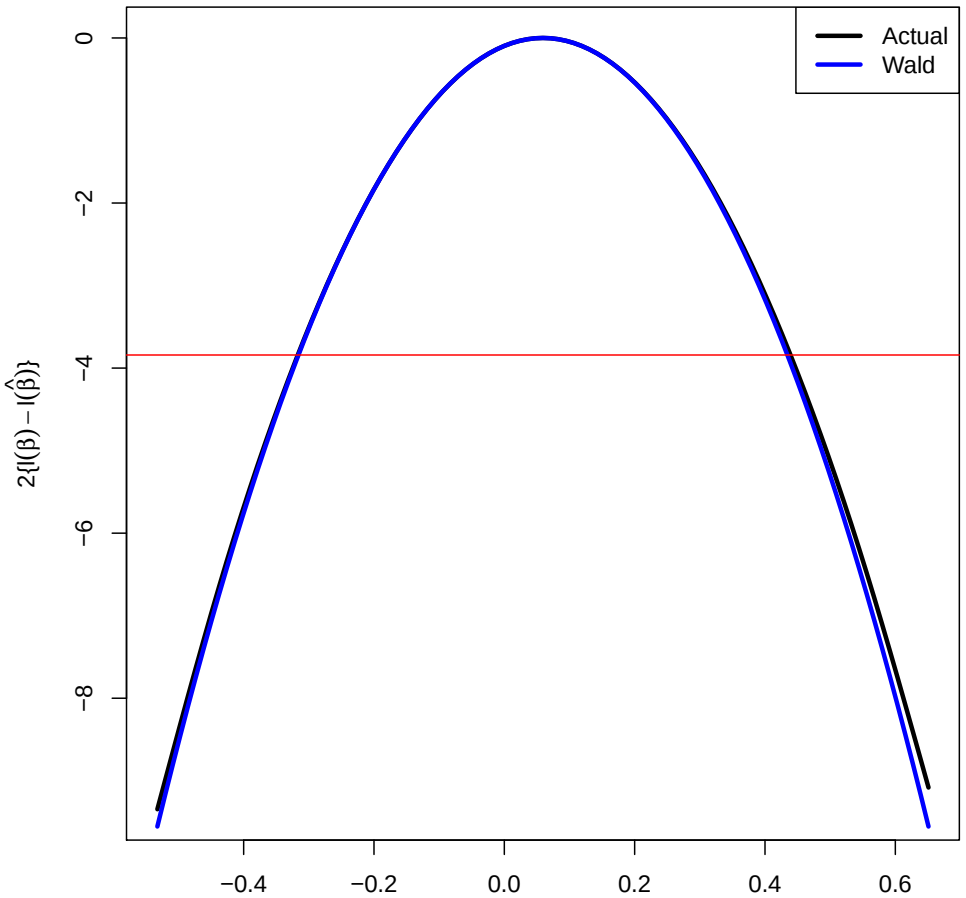


Comments

- As you can see, the Wald approach is incapable of capturing asymmetry in the likelihood function, and must therefore always produce symmetric confidence intervals about the MLE
- The likelihood ratio is not subject to this restriction (but of course must refit a new model at all the different values for β)
- This impacts hypothesis testing as well
- Recall that the likelihood is only approximately (asymptotically) normal – and there are only 15 adult females in this data set

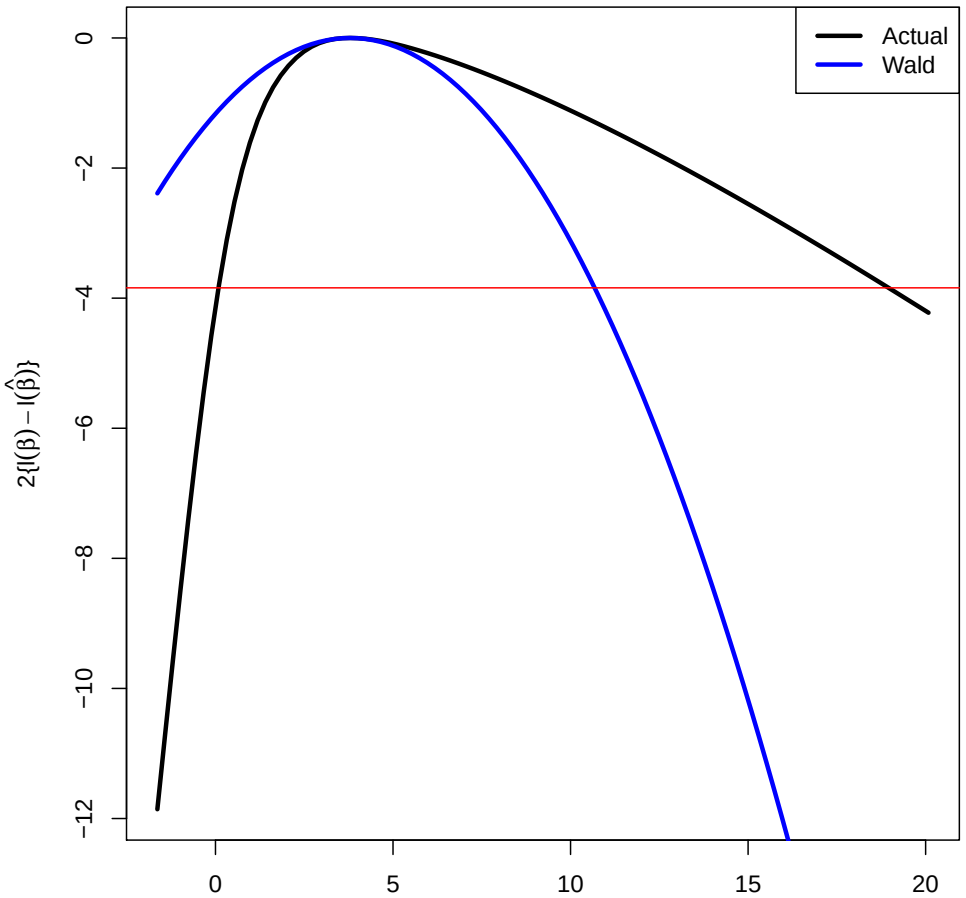
Wald vs. Likelihood ratio

When n is larger, the agreement is much better (here, $n = 100$, $p = 2$):



Wald vs. Likelihood ratio

When n is smaller, the agreement is even worse (here, $n = 6$, $p = 2$):



Complete separation

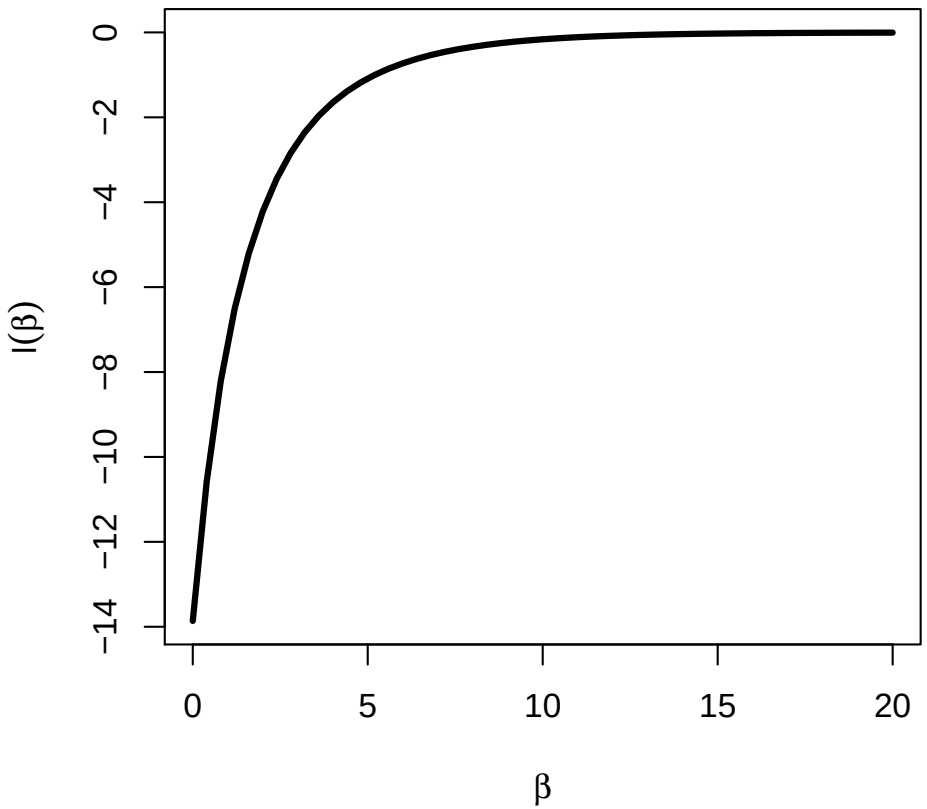
- Just as in univariate statistics, when n is large we can often ignore the fact that our data is discrete and use a normal approximation
- When n is small, however, problems can arise
- Consider the following data:

x	y
-1.64	0
-0.80	0
-0.46	0
-0.46	0
-0.34	0
0.12	1
0.62	1
0.64	1
0.73	1
1.10	1

Complete separation (cont'd)

- If we try to fit a logistic regression model to this data, we find that the algorithm will not converge and we get an error message in SAS or R
- The reason is that all of the events occur when x is large and don't occur when x is small
- To put it another way, we can draw a line in the x's and separate the $y = 0$'s from the $y = 1$'s
- This phenomenon is referred to as *complete separation* (or more generally, as the problem of *monotone likelihood*)

Monotone likelihood



Ramifications

- What it means is that the MLE occurs at infinity (or $-\infty$)
- This has a number of ramifications:
  - Numerical algorithms will fail
  - Weights will go to zero
  - Standard errors will go to infinity

Complete separation: Practical aspects

- This has a number of complicated ramifications for inference lie beyond the scope of this course
- Practically speaking, the ramifications are that the data do not allow you to estimate a certain parameter in the way that the model is currently specified
- This can often occur when models are overparameterized – in models with many explanatory variables, complete separation occurs whenever a linear predictor completely separates the outcome
- In linear regression, estimates are only undefined if $\mathbf{X}$ is not full rank; in logistic regression, complete separation represents an additional restriction on the complexity of the design matrix

Linear vs. logistic regression

- Note that this phenomenon does not occur for linear regression
- In linear regression, each outcome contains a continuous amount of information about μ_i
- In logistic regression, outcomes are discrete, and it is easy for two outcomes to both have $y = 1$ even though $\mu_1 = .99$ and $\mu_2 = .51$

Information

- To quantify this, let's compare the Fisher information contained in a single observation in each type of regression:

$$\mathbf{J}_1 = \sigma^2 \mathbf{x}\mathbf{x}^T \qquad \text{(Linear regression)}$$

$$\mathbf{J}_1 = w\mathbf{x}\mathbf{x}^T \qquad \text{(Logistic regression)}$$

- Compared to an observation for linear regression (with $\sigma^2 = 1$), an equivalent observation from logistic regression carries only w as much information
- Note that the maximum value $w = \pi(1-\pi)$ can take on is 0.25, and if π is close to 0 or 1, the observation contains almost no information

Example

- To see how this works, consider our alcohol metabolism data set
- In that data set, we had a continuous measurement of metabolism, but what if we only had a discrete measurement, `HighMet`, an indicator for whether metabolism was above 2 or not
- Consider two models with the same systematic component:

$$\eta = \beta_0 + \beta_1 \texttt{Gastric} + \beta_2 \texttt{Male} + \beta_3 \texttt{Alcoholic},$$

but one is a linear regression model for `Metabol` while the other is a logistic regression model for `HighMet`

Results

The p-values for the regression coefficients:

	Linear	Logistic
Gastric	< 0.0001	0.1165
Male	0.0056	0.1267
Alcoholic	0.8453	0.8153

Results (cont'd)

- Furthermore, as you may recall, linear regression indicated a significant interaction between `Sex` and `Gastric`
- In the logistic regression model with the same data, this interaction cannot even be estimated due to complete separation in the male group
- The take-home message here is that you need far larger sample sizes when your outcome is discrete than when your outcome is continuous
- This is worth keeping in mind, as non-statisticians often categorize continuous outcomes when they analyze data

Categorical analyses

- One final miscellaneous topic: when all explanatory variables are categorical, logistic regression has much in common with classical methods for categorical data analyses: χ^2 tests, Mantel-Haenszel estimators, tests for homogeneity, etc.
- The methods often share much in common, such as enforcing (or testing) the constancy of odds ratios across levels of a confounder
- As a result, the tests are often very similar, although not exactly the same (the difference arising from whether or not the number of events is treated as fixed or random)

t tests

- Suppose we have a discrete outcome and a continuous exposure: instead of logistic regression, we also test for association using a t-test
- In general, there is a tradeoff involved here:
  - If the distribution of the exposure is normal, the t-test is more powerful because it utilizes additional information
  - On the other hand, it the distribution is far from normal, logistic regression is more powerful
- Keep in mind, however, that there are advantages for interpretation if the outcome is in fact treated as the random variable (such as the ability to estimate and report probabilities and odds ratios)