Inference: Likelihood ratio vs. Wald approaches

Patrick Breheny

March 19

Introduction: The Wald approach

- Thus far, all our inferences have been based on the result:

$$\widehat{\boldsymbol{\beta}} \dot{\sim} \mathrm{N}\left(\boldsymbol{\beta}, \phi(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\right)$$

- This approach has the great advantage of simplicity: all you need to know is $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathrm{Var}}(\widehat{\boldsymbol{\beta}})$ and you may construct by hand all the tests and confidence intervals you need for any element of $\boldsymbol{\beta}$ or any linear combination of the elements of $\boldsymbol{\beta}$ (these are called "Wald tests", "Wald confidence intervals", etc.)
- Recall, however, that the result on the previous slide is based on an approximation to the likelihood at the MLE, and this approximation may be poor at $\boldsymbol{\beta}$ values far from $\widehat{\boldsymbol{\beta}}$

Likelihood ratios

- A competing approach is based on likelihood ratios
- We consider the univariate case first, comparing the likelihood at an arbitrate value θ with that of the MLE $\hat{\theta}$:

$$\lambda = \frac{L(\theta)}{L(\hat{\theta})}$$

- **Theorem:** As $n \to \infty$ with iid data, subject to the usual regularity conditions,

$$-2\log\lambda \xrightarrow{\text{d}} \chi^2_1$$

Likelihood ratios for regression

- This result can be extended to multivariate and non-iid cases as well; consider two models:

$$\text{Full: } \boldsymbol{\beta} = (\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)})$$
$$\text{Reduced: } \boldsymbol{\beta} = (\boldsymbol{\beta}_0^{(1)}, \boldsymbol{\beta}^{(2)})$$

  where $\boldsymbol{\beta}_0^{(1)}$ is a specified vector of constants

- Letting λ denote the likelihood ratio comparing the reduced model to the full model, we have

$$-2\log\lambda \;\dot\sim\; \chi^2_q,$$

  where q is the length of $\boldsymbol{\beta}^{(1)}$ (typically, the number of parameters assumed to be zero)

Likelihood ratio tests and confidence intervals

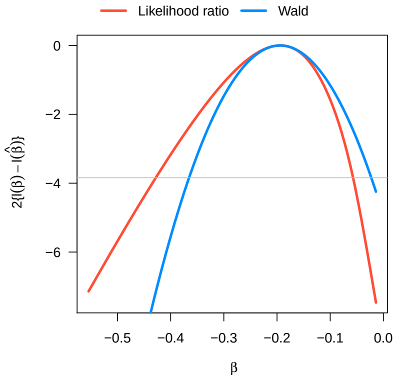
- This result allows us to carry out hypothesis tests by calculating $p = \Pr(\chi^2_q \geq 2\log(\lambda))$
- It also allows us to construct $(1-\alpha)$ confidence intervals by inverting the above test – *i.e.*, finding the set of parameter values $\boldsymbol{\beta}_0^{(1)}$ such that

$$-2\log\frac{L(\widehat{\boldsymbol{\beta}}|\boldsymbol{\beta}^{(1)} = \boldsymbol{\beta}_0^{(1)})}{L(\widehat{\boldsymbol{\beta}})} \leq \chi^2_{1-\alpha,q},$$

where $\chi^2_{1-\alpha,q}$ is the $(1-\alpha)$ quantile of the χ^2 distribution with q degrees of freedom

Wald vs. Likelihood ratio

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



95% confidence intervals:

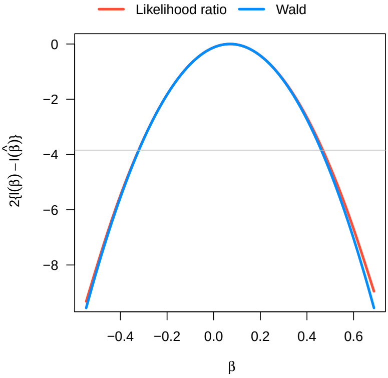
$$\text{Wald: } (-0.365, -0.023)$$
$$\text{LR: } (-0.428, -0.057)$$

Remarks

- 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 (the downside, of course, is that we must refit a new model at all the different values for β)
- This impacts hypothesis testing as well: for testing the interaction term, the Wald test gives $p = 0.087$ while the LRT gives $p = 0.048$

Wald vs. Likelihood ratio

For the donner data, $n = 45$ and $p = 3$; when n is larger, the agreement is much better (here, $n = 100$, $p = 2$):

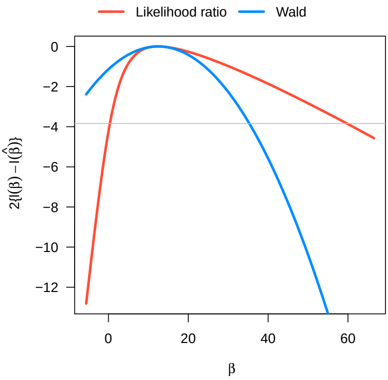


95% confidence intervals:

$$\text{Wald: } (-0.321, 0.461)$$
$$\text{LR: } (-0.322, 0.468)$$

Wald vs. Likelihood ratio

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



95% confidence intervals:

$$\text{Wald: } (-10.4, 35.3)$$
$$\text{LR: } (0.336, 59.7)$$

Likelihood ratio vs. Wald: Summary

- The Wald approach enjoys popularity due to its simplicity (likelihood ratio confidence intervals are obviously difficult to construct by hand)
- The two approaches often agree quite well
- However, there are also situations where the two disagree dramatically
- Tests and confidence intervals based on likelihood ratios are more accurate, and should be trusted over the Wald approach