Profile Likelihood and RVM Confidence Intervals under Separation

Dennis Zhou

Aug 6, 2025

1 Simulating Separation in Logistic Regression

We simulate a binary outcome $y_i \in \{0, 1\}$ with a single covariate $x_i \in \mathbb{R}$ such that:

$$y_i = \begin{cases} 1 & \text{if } x_i > 0 \\ 0 & \text{if } x_i \leq 0 \end{cases}$$

This produces **perfect separation**, where a linear function of x perfectly predicts the outcome y. We generate data as:

$$x_i = \text{linspace}(-2, 2, n), \quad y_i = \mathbb{1}(x_i > 0)$$

2 Logistic Regression and Wald Confidence Interval

We fit a logistic regression model:

$$\log\left(\frac{\Pr(Y = 1)}{\Pr(Y = 0)}\right) = \beta_0 + \beta_1 x$$

Under separation, the maximum likelihood estimate (MLE) for β_1 diverges toward $+\infty$, but if the model is fit anyway, the Wald confidence interval is computed as:

$$\hat{\beta}_1 \pm z_{\alpha/2} \cdot \text{SE}(\hat{\beta}_1)$$

This interval is based on the normal approximation and is generally unreliable under separation due to:

- Inflated standard errors,
- Asymmetry of the likelihood,
- Infinite or undefined MLE.

```
                           Logit Regression Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  100
Model:                          Logit   Df Residuals:                       98
Method:                           MLE   Df Model:                            1
Date:                Wed, 06 Aug 2025   Pseudo R-squ.:                   1.000
Time:                        15:20:39   Log-Likelihood:            -1.8764e-08
converged:                      False   LL-Null:                       -69.315
Covariance Type:            nonrobust   LLR p-value:                 5.308e-32
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const       1.882e-09   7300.284   2.58e-13      1.000   -1.43e+04    1.43e+04
x1           914.9820   3.61e+05      0.003      0.998   -7.07e+05    7.09e+05
==============================================================================

Complete Separation: The results show that there iscomplete separation or perfect prediction.
In this case the Maximum Likelihood Estimator does not exist and the parameters
are not identified.
```

Figure 1: Logit fit result with complete seperation detected

Wald Confidence Interval	
β_0	β_1
$[-14308.29,\ 14308.29]$	$[-707345.56,\ 709175.52]$

Table 1: Wald CI for intercept and slope under separation

3 Profile Likelihood-Based Confidence Interval

To address issues with the Wald CI, we compute a **profile likelihood** for β_1:

$$\ell_p(\beta_1) = \max_{\beta_0} \sum_{i=1}^{n} \left[y_i \log\left(\Lambda(\beta_0 + \beta_1 x_i)\right) + (1 - y_i) \log\left(1 - \Lambda(\beta_0 + \beta_1 x_i)\right) \right]$$

where $\Lambda(z) = \frac{1}{1+e^{-z}}$ is the logistic function.
We define the **likelihood ratio statistic**:

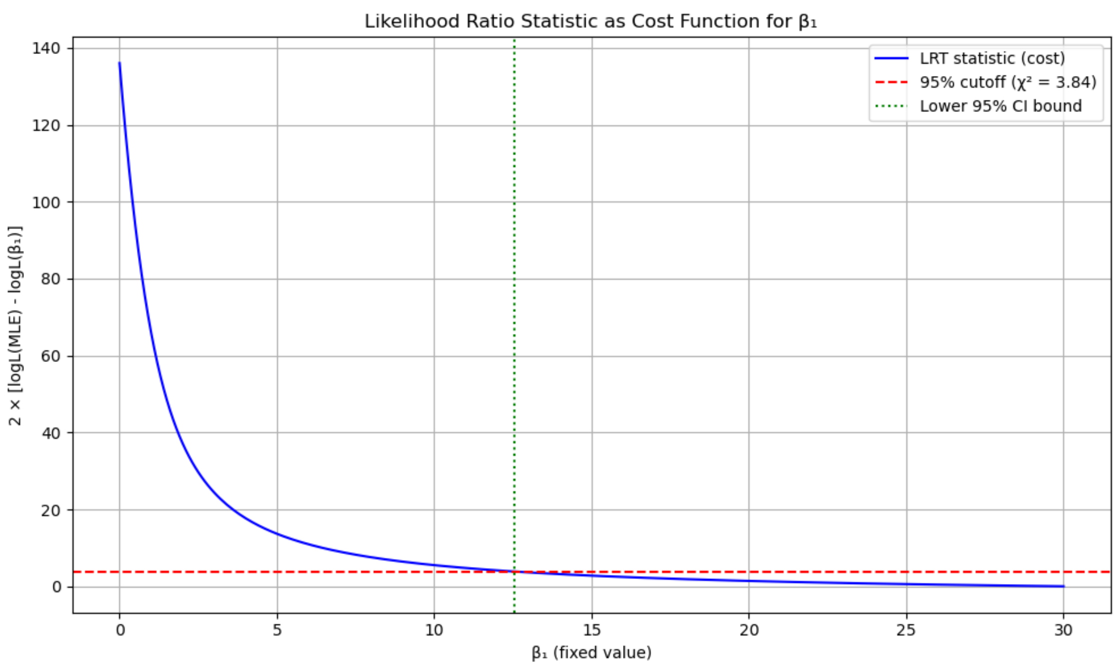
$$\lambda(\beta_1) = 2\left[\ell_p(\hat{\beta}_1) - \ell_p(\beta_1)\right]$$

The **one-sided 95% lower confidence bound** for β_1 is given by:

$$\lambda(\beta_1) \leq \chi^2_{1,0.95}$$

This yields an interval:

$$\beta_1 \in [\beta_L, \infty), \quad \text{where } \beta_L = \inf\left\{\beta_1 : \lambda(\beta_1) \leq \chi^2_{1,0.95}\right\}$$



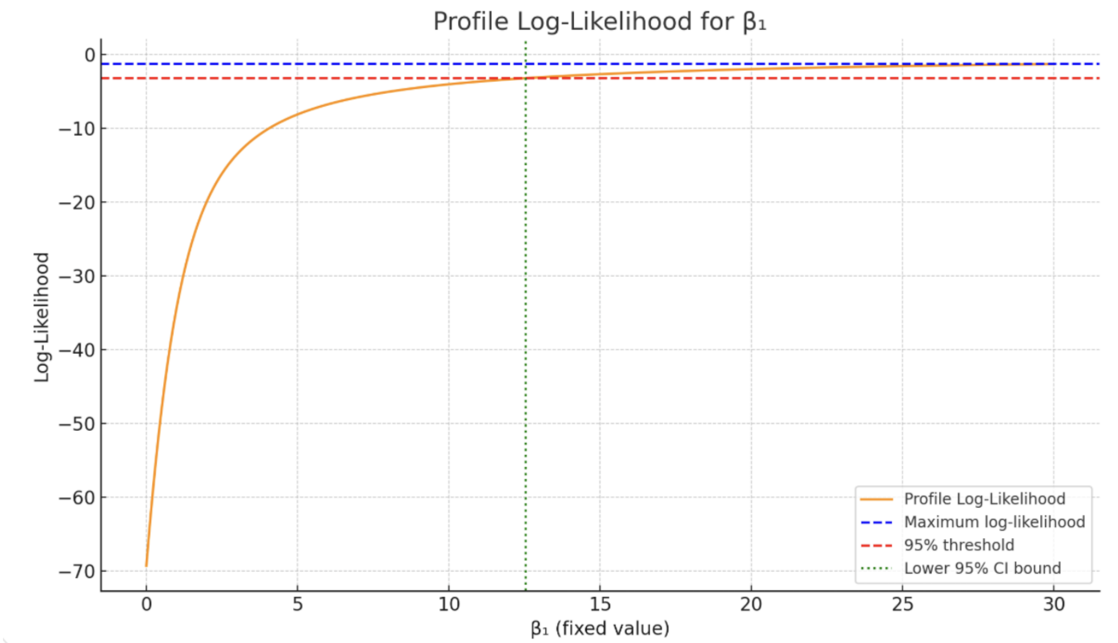


Figure 2: profile log-likelihood and β_1 selection

Orange curve: For each fixed value of β_1, this is the maximized *profile log-likelihood*, defined as:

$$\ell_p(\beta_1) = \max_{\beta_0} \ell(\beta_0, \beta_1)$$

This curve reflects the explanatory power of different β_1 values with respect to the data.

Blue dashed line: This is the global maximum of the log-likelihood, i.e., the value achieved by the best possible model, corresponding to $\hat{\beta}_1 \to \infty$ (due to separation).

Red dashed line: This represents the 95% confidence interval cutoff in the log-likelihood space, corresponding to a drop of

$$\frac{1}{2}\chi^2_{1,0.95} = 1.92$$

from the maximum:

$$\ell_p(\beta_1) \geq \ell_{\max} - 1.92$$

Green vertical line: This marks the smallest β_1 value where the profile log-likelihood remains above the cutoff, which is the final lower bound estimation:

$$\boxed{\beta_1 > \beta_{1,\text{lower}} \quad \text{with 95\% confidence}}$$

4 Robust Venzon–Moolgavkar Algorithm

The Robust Venzon–Moolgavkar (RVM) algorithm is a numerically stable method for computing profile likelihood-based confidence intervals, particularly in models with nonstandard behavior such as separation in logistic regression, near-boundary parameters, or weak identifiability.

4.1 Recall Likelihood Ratio Statistic

The standard profile likelihood confidence interval for a scalar parameter of interest θ is defined as the set of values for which the likelihood ratio statistic does not exceed a threshold based on the χ^2 distribution:

$$\lambda(\theta) = 2\left[\ell(\hat{\theta}) - \ell_p(\theta)\right] \leq \chi^2_{1,1-\alpha}$$

where:

- $\ell(\hat{\theta})$ is the maximized log-likelihood,
- $\ell_p(\theta) = \max_\psi \ell(\theta, \psi)$ is the profile log-likelihood, optimized over nuisance parameters ψ,
- $\chi^2_{1,1-\alpha}$ is the $(1-\alpha)$ quantile of the chi-squared distribution with 1 degree of freedom.

The classical algorithm by Venzon and Moolgavkar (1988) attempts to find the endpoints of this confidence region by solving the equation $\lambda(\theta) = \chi^2_{1,1-\alpha}$ via root-finding (e.g., Newton–Raphson). However, this method can fail in practice due to poor curvature, non-convexity, or flat regions in the log-likelihood surface.

4.2 RVM Algorithm and Trust Region Approach

The RVM algorithm extends the classical method by introducing a *trust region* strategy to ensure convergence and robustness. Specifically, at each iteration, the algorithm:

1. Approximates the likelihood ratio statistic $\lambda(\theta)$ using a local quadratic model.
2. Solves a constrained optimization problem to update θ with d within a neighborhood:

   $$\min_d \left(\lambda(\theta_k + d) - \chi^2_{1,1-\alpha}\right)^2 \quad \text{subject to} \quad \|d\| \leq \Delta_k$$

   where Δ_k is the trust region radius.
3. Updates the step size Δ_k depending on whether the new step reduces the residual effectively.
4. Iterates until the residual $|\lambda(\theta_k) - \chi^2_{1,1-\alpha}|$ is below a specified tolerance.

This approach ensures that updates are only taken where the quadratic approximation of the likelihood is valid, thus avoiding divergence or cycling in difficult likelihood surfaces.

4.3 Advantages of the RVM Method

Compared to classical methods, the RVM algorithm offers several advantages:

- **Robustness:** Works reliably under nonconvex, flat, or irregular likelihood functions.
- **Boundary handling:** Performs well even when the MLE lies on the boundary of the parameter space (e.g., separation).
- **No second derivatives:** Avoids explicit Hessian computation, relying only on function evaluations and directional curvature.
- **Confidence interval validity:** Yields accurate one-sided and two-sided profile likelihood confidence intervals.

4.4 Applications

The RVM algorithm is particularly useful in the context of:

- Logistic regression with complete or quasi-complete separation,
- Models with weakly identified parameters,
- Highly skewed or non-quadratic likelihoods,
- Complex parametric models where standard Wald-type intervals fail.

It is implemented in the `ci-rvm` Python package and can be integrated with custom likelihood functions.

5 Conclusion

Method	Description	Reliability under Separation
Wald CI	Normal-based, symmetric	Poor
Profile Likelihood CI	Inverts likelihood ratio	Good
RVM Profile CI	Trust-region robust profile CI	Excellent

Profile likelihood and RVM approaches provide robust alternatives to classical normal-theory inference, especially when MLEs are non-existent or unreliable.

6 Appendix