

Logistic regression with distributed data

Estimates and confidence intervals for the parameters of the logistic regression model $P(Y = 1) = \exp(X\beta) / (1 + \exp(X\beta))$ are obtained by making use of the Newton-Raphson algorithm using local gradients and Hessian matrices. This allows for the recreation of the estimates and confidence intervals from the centralised setting. For the procedure, all nodes must have the same number of predictors. The total number of observations (individuals) is N and the number of predictors is p .

After an initial iteration $t = 0$ consisting of averaging the local estimates of each node at the coordination center, the aggregate estimate is updated using the gradients and Hessian matrices of the data nodes. The process may be repeated until convergence (which yields the centralised estimate).

Assuming K nodes and a coordinating node are involved; each node k identifies the local (binary) outcome vector $Y^{(k)}$ and the local predictor matrix $X^{(k)}$. A column with the quantity 1 for each observation is added in the local predictor matrix to account for the intercept estimate.

Example.

Suppose the following dataset at node k , with 3 observations ($N^{(k)} = 3$) and 2 predictors ($p = 2$).

Premature_birth	gestational_age	age_admission
0	42	56
0	38	43
1	37	25

The local outcome vector is $Y^{(k)} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ and the local predictor matrix $X^{(k)} = \begin{bmatrix} 1 & 42 & 56 \\ 1 & 38 & 43 \\ 1 & 37 & 25 \end{bmatrix}$.

Data node (initial phase: iteration $t = 0$)

1. Each node computes its local estimate $\hat{\beta}^{(k)}$ using the standard procedure of maximizing its local likelihood function (e.g. using R's glm function). This estimate is saved as a csv file and sent to the coordination node, along with the local sample size.

Example (continued).

The following quantities **are shared** to the coordinating node:

$$\hat{\beta}^{(k)} = \begin{bmatrix} -280.796 \\ 10.385 \\ -3.195 \end{bmatrix}$$

$$N^{(k)} = 3$$

The exported csv will share the following table from node k :

coefs,	n
- 280.796,	3
10.385,	NA
- 3.195,	NA

Coordinating node (initial phase, iteration $t = 0$)

2. The coordinating node averages the local estimates to generate the simple averaging estimate: $\hat{\beta}_{NR, t=0} = \hat{\beta}^{SA} = \left(\sum_{k=1}^K N^{(k)} \hat{\beta}^{(k)} \right) / \sum_{k=1}^K N^{(k)}$, then sends that quantity back to the data nodes.

Data node (iteration phase, $t = 1, \dots, T$)

3. Each data node computes its local gradient and Hessian evaluated at the latest β estimate, $\hat{\beta}_{NR, t-1}$:

- $D_{NR, t}^{(k)} = X^{(k)T} \left\{ Y^{(k)} - s^{(k)}(\hat{\beta}_{NR, t-1}) \right\}$,
- $V_{NR, t}^{(k)} = X^{(k)T} W^{(k)}(\hat{\beta}_{NR, t-1}) X^{(k)}$.

Where

- $s^{(k)}(\beta) = \left(\frac{\exp(\beta^T x_1^{(k)})}{1 + \exp(\beta^T x_1^{(k)})}, \dots, \frac{\exp(\beta^T x_{N^{(k)}}^{(k)})}{\exp(\beta^T x_{N^{(k)}}^{(k)})} \right)^T$,
- $W^{(k)}(\beta) = \text{diag} \left(\left\{ \frac{\exp(\beta^T x_i^{(k)})}{1 + \exp(\beta^T x_i^{(k)})} \right\} \left\{ 1 - \frac{\exp(\beta^T x_i^{(k)})}{1 + \exp(\beta^T x_i^{(k)})} \right\} \right)$.

The quantity $s^{(k)}(\hat{\beta}_{NR, t-1})$ represents the outcome predictions using the current β estimate and $W^{(k)}(\hat{\beta}_{NR, t-1})$ is a diagonal matrix whose entries can be interpreted as outcome predictions variances.

The quantities $D_{NR, t}^{(k)}$ and $V_{NR, t}^{(k)}$ are sent to the coordinating center.

Example (continued).

Suppose the coordinating node shared $\hat{\beta}_{NR, t=0} = \hat{\beta}^{SA} = \begin{bmatrix} -20 \\ 5 \\ -4 \end{bmatrix}$.

The following quantities **are shared** to the coordinating node (iteration $t = 1$):

$$D_{NR, t=1}^{(k)} = \begin{bmatrix} -0.1192 \\ -4.5297 \\ -5.1257 \end{bmatrix}, \quad V_{NR, t}^{(k)} = \begin{bmatrix} 0.1050 & 3.9898 & 4.5147 \\ 3.9898 & 151.6107 & 171.5595 \\ 4.5147 & 171.5595 & 194.1331 \end{bmatrix}.$$

The exported csv will share the following table from node k :

gradient,	hessian_intercept	hessian_pred1	hessian_pred2
- 0.1192,	0.1050,	3.9898,	4.5147
- 4.5297,	3.9898,	151.6107,	171.5595
- 5.1257,	4.5147,	171.5595,	194.1331

Coordinating node (iteration phase, $t = 1, \dots, T - 1$)

4. The coordinating center computes the global gradient and Hessian

- $\bar{D}_{NR, t} = \sum_{k=1}^K D_{NR, t}^{(k)}$,
- $\bar{V}_{NR, t} = \sum_{k=1}^K V_{NR, t}^{(k)}$.

It then computes the Newton-Raphson iteration $\hat{\beta}_{NR, t} = \hat{\beta}_{NR, t-1} + (\bar{V}_{NR, t})^{-1} \bar{D}_{NR, t}$ and sends the new estimate to the data nodes.

Steps 3. and 4. are repeated until convergence or a fixed number of times (T).

Coordinating node (last iteration T)

5. The coordinating center computes the global gradient and Hessian

- $\bar{D}_{NR, T} = \sum_{k=1}^K D_{NR, T}^{(k)}$,
- $\bar{V}_{NR, T} = \sum_{k=1}^K V_{NR, T}^{(k)}$.

It then computes the Newton-Raphson iteration $\hat{\beta}_{NR, T} = \hat{\beta}_{NR, T-1} + (\bar{V}_{NR, T})^{-1} \bar{D}_{NR, T}$.

The (estimated) covariance matrix of the coefficients is $\hat{\Sigma} = (\bar{V}_{NR, T})^{-1}$.

Lower and upper bounds for the confidence intervals of the model parameters are also calculated at the coordinating node:

$$CI(\beta_j) = \left[\hat{\beta}_{NR, T} \pm z_{\frac{\alpha}{2}} \sqrt{\hat{\Sigma}_{j+1, j+1}} \right]$$

The outputs of the procedure are the parameters estimates (including intercept), and the upper and lower bounds of the confidence intervals for the model parameters.