A non-convex regularization approach for stable estimation of loss development factors

Himchan Jeong, PhD, FSA

(joint work with Hyunwoong Chang and Emiliano A. Valdez)

Simon Fraser University

IDSC 2021

18 June, 2021

Linear Models and Penalization

- OLS: No penalization,

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \frac{1}{2}||y - X\beta||^2$$

- Ridge: L_2 penalization,

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \frac{1}{2}||y - X\beta||^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

- LASSO: L_1 penalization,

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \frac{1}{2}||y - X\beta||^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

- LAAD: Adjusted L_1 penalization,

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \frac{1}{2}||y - X\beta||^2 + r \sum_{j=1}^{p} \log(1 + |\beta_j|)$$

Our contributions can be summarized as follows:

Literature Review and Contributions

Our contributions can be summarized as follows:

- We focus on the use of maximum a posterior (MAP) estimate with LAAD penalty and its convergence analysis under mild conditions.

Our contributions can be summarized as follows:

- We focus on the use of maximum a posterior (MAP) estimate with LAAD penalty and its convergence analysis under mild conditions.
- It is also shown that the proposed penalty can be effectively used in an actuarial application, insurance claim reserving with aggregate loss triangle.

Analytic Solution for Univariate Case

To better understand the characteristic of a model with LAAD penalty, let us consider a simple example when $p = 1$ and $||X|| = 1$. In this case, it is sufficient to solve the following:

$$\hat{\theta}_j = \underset{\theta_j}{\operatorname{argmin}} \frac{1}{2}(z_j - \theta_j)^2 + r \log(1 + |\theta_j|), \tag{1}$$

where $z = X'y$.

Analytic Solution for Univariate Case

Theorem 1

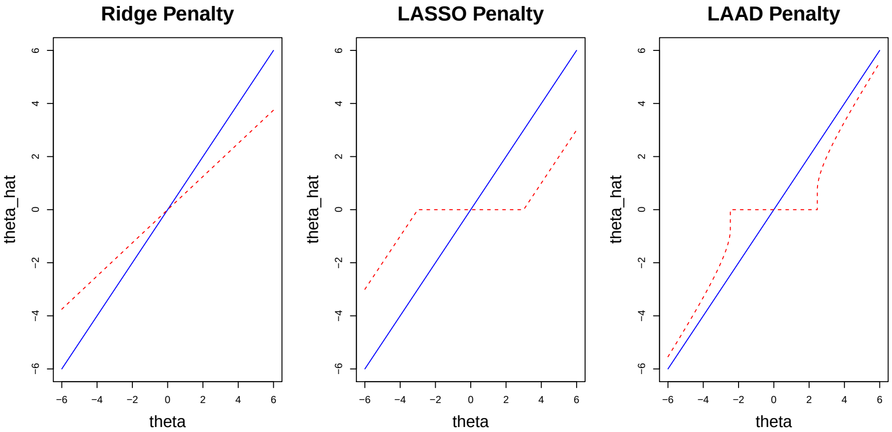
Let us set $\ell(\theta|r,z) = \frac{1}{2}(z-\theta)^2 + r\log(1+|\theta|)$*. Then the corresponding minimizer will be given as* $\hat{\theta} = \theta^* \cdot \mathbb{1}_{\{|z| \geq z^*(r) \vee r\}}$*, where*

$$\theta^* = \frac{1}{2}\left[z + sgn(z)\left(\sqrt{(|z|-1)^2 + 4|z| - 4r} - 1\right)\right],$$

and $z^*(r)$ *is the unique solution of*

$$\Delta(z|r) = \frac{1}{2}(\theta^*)^2 - \theta^* z + r\log(1+|\theta^*|) = 0.$$

Estimate Behavior for Different Penalties



Issues in Penalization

- Shrinkage: whether the method distorts the estimated parameter value(s),
- Variable selection: whether the method allows us to rule out some insignificant variables from the model

	Shrinkage	Variable Selection
OLS	NO	NO
Ridge	More if $\lvert\beta\rvert \uparrow$	NO
LASSO	Constant	YES
LAAD	More if $\lvert\beta\rvert \downarrow$	YES

Coordinate Descent Algorithm for the Multivariate Case

Since an analytic solution is obtained in the case of univariate penalized least squares, one can implement **coordinate descent algorithm** in the multivariate case,

- which starts with an initial set of estimates,
- and then successively optimize along each coordinate or blocks of coordinates.

Convergence of Coordinate Descent Algorithm

The following theorem provides us sufficient conditions that our optimization problem converges with coordinate descent algorithm.

Theorem 2

If $||X_j|| = 1$ *for all* $j = 1, \ldots, p$ *and* $r \leq 1$*, then the solution from coordinate descent algorithm with function* $l : \mathbb{R}^p \to \mathbb{R}$ *converges to* $\hat{\beta}$ *where*

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \frac{1}{2}||y - X\beta||^2 + r\sum_{j=1}^{p} \log(1 + |\beta_j|).$$

Loss Development with Aggregate Triangular Claims Data

	DL 1	DL 2	DL 3	DL 4	DL 5	DL 6	DL 7	DL 8	DL 9	DL 10
AY 1	$Y_{1,1}$	$Y_{1,2}$	$Y_{1,3}$	$Y_{1,4}$	$Y_{1,5}$	$Y_{1,6}$	$Y_{1,7}$	$Y_{1,8}$	$Y_{1,9}$	$Y_{1,10}$
AY 2	$Y_{2,1}$	$Y_{2,2}$	$Y_{2,3}$	$Y_{2,4}$	$Y_{2,5}$	$Y_{2,6}$	$Y_{2,7}$	$Y_{2,8}$	$Y_{2,9}$	
AY 3	$Y_{3,1}$	$Y_{3,2}$	$Y_{3,3}$	$Y_{3,4}$	$Y_{3,5}$	$Y_{3,6}$	$Y_{3,7}$	$Y_{3,8}$		
AY 4	$Y_{4,1}$	$Y_{4,2}$	$Y_{4,3}$	$Y_{4,4}$	$Y_{4,5}$	$Y_{4,6}$	$Y_{4,7}$			
AY 5	$Y_{5,1}$	$Y_{5,2}$	$Y_{5,3}$	$Y_{5,4}$	$Y_{5,5}$	$Y_{5,6}$				
AY 6	$Y_{6,1}$	$Y_{6,2}$	$Y_{6,3}$	$Y_{6,4}$	$Y_{6,5}$					
AY 7	$Y_{7,1}$	$Y_{7,2}$	$Y_{7,3}$	$Y_{7,4}$						
AY 8	$Y_{8,1}$	$Y_{8,2}$	$Y_{8,3}$							
AY 9	$Y_{9,1}$	$Y_{9,2}$								
AY 10	$Y_{10,1}$									

- Given dataset can also be expressed as $\mathcal{D}_{1:10} = \{Y_{ij} : 1 \leq i \leq 10 \text{ and } 1 \leq j \leq \min(10, 11-i)\}$.
-

Loss Development with Aggregate Triangular Claims Data

	DL 1	DL 2	DL 3	DL 4	DL 5	DL 6	DL 7	DL 8	DL 9	DL 10
AY 1	$Y_{1,1}$	$Y_{1,2}$	$Y_{1,3}$	$Y_{1,4}$	$Y_{1,5}$	$Y_{1,6}$	$Y_{1,7}$	$Y_{1,8}$	$Y_{1,9}$	$Y_{1,10}$
AY 2	$Y_{2,1}$	$Y_{2,2}$	$Y_{2,3}$	$Y_{2,4}$	$Y_{2,5}$	$Y_{2,6}$	$Y_{2,7}$	$Y_{2,8}$	$Y_{2,9}$	$Y_{2,10}$
AY 3	$Y_{3,1}$	$Y_{3,2}$	$Y_{3,3}$	$Y_{3,4}$	$Y_{3,5}$	$Y_{3,6}$	$Y_{3,7}$	$Y_{3,8}$	$Y_{3,9}$	
AY 4	$Y_{4,1}$	$Y_{4,2}$	$Y_{4,3}$	$Y_{4,4}$	$Y_{4,5}$	$Y_{4,6}$	$Y_{4,7}$	$Y_{4,8}$		
AY 5	$Y_{5,1}$	$Y_{5,2}$	$Y_{5,3}$	$Y_{5,4}$	$Y_{5,5}$	$Y_{5,6}$	$Y_{5,7}$			
AY 6	$Y_{6,1}$	$Y_{6,2}$	$Y_{6,3}$	$Y_{6,4}$	$Y_{6,5}$	$Y_{6,6}$				
AY 7	$Y_{7,1}$	$Y_{7,2}$	$Y_{7,3}$	$Y_{7,4}$	$Y_{7,5}$					
AY 8	$Y_{8,1}$	$Y_{8,2}$	$Y_{8,3}$	$Y_{8,4}$						
AY 9	$Y_{9,1}$	$Y_{9,2}$	$Y_{9,3}$							
AY 10	$Y_{10,1}$	$Y_{10,2}$								

- Given dataset can also be expressed as $\mathcal{D}_{1:10} = \{Y_{ij} : 1 \leq i \leq 10 \text{ and } 1 \leq j \leq \min(10, 11 - i)\}$.
- One needs to predict ultimate claim (lower triangle) described as $\mathcal{D}_{10+k} = \{Y_{ij}^{(n)} : 1 + k \leq i \leq 10 \text{ and } j = 11 + k - i\}$.

Cross-Classified Model for Loss Development Prediction

- One can suggest the following model specification:

$$C_{i,j+1} := \log \frac{Y_{i,j+1}}{Y_{i,j}} \text{ and } \; C_{i,j+1} \sim \mathcal{N}\left(\eta_{j+1}, \sigma^2\right),$$

where η_{j+1} means the incremental development at $j+1^{th}$ year.

Model Specifications

- **Unconstrained model** - a model which minimizes the following for each line of business:

$$\sum_{i=1}^{I}\sum_{j=1}^{I-i}\left(C_{i,j+1}-\eta_{j+1}\right)^2,$$

- **LASSO constrained model** - a model which minimizes the following for each line of business with LASSO penalty:

$$\sum_{i=1}^{I}\sum_{j=1}^{I-i}\left(C_{i,j+1}-\eta_{j+1}\right)^2+\lambda\left[\sum_{j=2}^{J-1}\left|\eta_{j+1}\right|\right],$$

- **LAAD constrained model** - a model which minimizes the following for each line of business with LAAD penalty:

$$\sum_{i=1}^{I}\sum_{j=1}^{I-i}\left(C_{i,j+1}-\eta_{j+1}\right)^2+r\left[\sum_{j=2}^{J-1}\log(1+\left|\eta_{j+1}\right|)\right].$$

Summary of Estimated Incremental Development Factors

	General Liability			Other Casualty		
	Unconstrained	LASSO	LAAD	Unconstrained	LASSO	LAAD
$\exp(\eta_2)$	2.2022	2.3203	2.4066	1.2975	1.3064	1.3570
$\exp(\eta_3)$	1.5681	1.5514	1.5408	1.1052	1.1016	1.0876
$\exp(\eta_4)$	1.3108	1.2956	1.2846	1.0792	1.0754	1.0606
$\exp(\eta_5)$	1.1723	1.1574	1.1458	1.0352	1.0312	1.0157
$\exp(\eta_6)$	1.1569	1.1407	1.1281	1.0298	1.0254	1.0085
$\exp(\eta_7)$	1.0465	1.0299	1.0164	0.9959	1.0000	1.0000
$\exp(\eta_8)$	1.0512	1.0317	1.0163	1.0024	1.0000	1.0000
$\exp(\eta_9)$	1.0106	1.0000	1.0000	0.9929	0.9998	1.0000
$\exp(\eta_{10})$	1.0147	1.0000	1.0000	0.9589	0.9685	1.0000

- Note that η_2 is not penalized for both LASSO and LAAD models to avoid under-reserving due to shrinkage from regularization.
- A natural pattern of incremental LDFs is only observed with LAAD model, $\hat{\eta}_2 \geq \hat{\eta}_3 \geq \cdots \geq \hat{\eta}_L = 0 = \cdots = \hat{\eta}_{10}$.

Model Comparison of Validation Measures

It is also possible to evaluate the performance of prediction based on usual validation measures such as root mean squared error (RMSE) and mean absolute error (MAE) defined as follows:

$$\text{RMSE} =: \sqrt{\frac{1}{9}\sum_{i=2}^{10}(\hat{Y}^{(n)}_{i,12-i} - Y^{(n)}_{i,12-i})^2},$$

$$\text{MAE} =: \frac{1}{9}\sum_{i=2}^{10}|\hat{Y}^{(n)}_{i,12-i} - Y^{(n)}_{i,12-i}|.$$

	General Liability			Other Casualty		
	Unconstrained	LASSO	LAAD	Unconstrained	LASSO	LAAD
RMSE	45447.55	39250.54	36573.14	14075.55	12506.03	9919.94
MAE	28395.27	24200.97	23778.28	10738.65	9478.27	7685.04

Concluding Remarks

In this paper, we introduce LAAD penalty derived from the use of Laplace hyperprior for the λ in Bayesian LASSO which has some good properties such as

- variable selection with reversion to the true regression coefficients,
- analytic solution for the univariate case,
- and an optimization algorithm for the multivariate case which converges under modest condition via coordinate descent.

Concluding Remarks

We also explore a possible use of LAAD penalty in actuarial application, especially calibration of loss development model and tail factor selection.

- According to the results of the empirical analysis, one can see that use of LAAD penalty ended up with reasonable loss development pattern while the other methods deviate from that pattern.