

6-8-2021

Penalized Likelihood Estimation of Gamma Distributed Response Variable via Corrected Solution of Regression Coefficients

Rasaki Olawale Olanrewaju

University of Ibadan, Nigeria, rasakiolawale@gmail.com

Follow this and additional works at: <https://digitalcommons.wayne.edu/jmasm>



Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Olanrewaju, R. O. (2020). Penalized Likelihood Estimation of Gamma Distributed Response Variable via Corrected Solution of Regression Coefficients. *Journal of Modern Applied Statistical Methods*, 19(1), eP2891. <https://doi.org/10.22237/jmasm/1608552720>

This Emerging Scholar is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in *Journal of Modern Applied Statistical Methods* by an authorized editor of DigitalCommons@WayneState.

Penalized Likelihood Estimation of Gamma Distributed Response Variable via Corrected Solution of Regression Coefficients

Cover Page Footnote

I thank Adewale Lukman (Ph.D.) and Mr. Oladebo Omotayo for their useful suggestions that led to the improvement of this paper.

EMERGING SCHOLAR

Penalized Likelihood Estimation of Gamma Distributed Response Variable via Corrected Solution of Regression Coefficients

Rasaki Olawale Olanrewaju
University of Ibadan
Ibadan, Nigeria

A Gamma distributed response is subjected to regression penalized likelihood estimations of Least Absolute Shrinkage and Selection Operator (LASSO) and Minimax Concave Penalty via Generalized Linear Models (GLMs). The Gamma related disturbance controls the influence of skewness and spread in the corrected path solutions of the regression coefficients.

Keywords: gamma, generalized linear models (GLMs), least absolute shrinkage and selection operator (LASSO), minimax concave penalty (MCP), skewness

Introduction

The classical regression model has been widely used in various disciplines with major interest on the response variable (Y) subjected to a set of linearized predictor (covariate) functions (Dette et al. 2006). Estimating the regression coefficients requires different types of techniques based on the conditional distribution of the random component (ε) or the distribution of the response variable and the form(s) of the covariate measurements (Berk et al. 2013).

In a conventional classical regression approach, where the response variable is to be subjugated to Gaussian distribution with covariates (X_i) of the continuous type data, the Generalized Linear Models (GLMs) technique of parameter estimation via an identity smooth link function of $\eta_i = g(\mu_i) = \mu_i$ and identity invertible link function of $\mu_i = g^{-1}(\eta_i) = \eta_i$. Because the link and invertible functions

are identity (that is, transfer function not required), the GLMs technique of estimation coincides with General Linear Model (GLM); a technique of estimation for Gaussian conditional distribution of the random component or response variable (Fox, 2008; Antoniadis et al., 2014). Unlike the GLM, when the distribution of the random component or the response variable belongs to the exponential family coupled with any data form of the covariates, the GLMs via Maximum Likelihood Estimation (MLE), Quasi-Likelihood Estimation (QLE), Iterative Re-weighted Least Square (IRLS) etc. would be the ideal technique to diagnose and estimate the embedded parameters (Watkins, 2001).

The aforementioned techniques (GLM and GLMs) pigeonholed a vital trait in high-dimensional, ultra-dimensional and high-throughput data in epidemiology, clinical studies, survival studies, genetics, bioinformatics and at times in finance where $d \geq n$, that is, the number of covariates or intended regression coefficients to be estimated is greater than the considered or available sample size or sample space. (Zhou & Shen, 2001; Fan & Lv, 2008; Fan et al., 2009).

Sparse regression estimation via penalized regularization and variable selection technique of important variables (seeking for smaller subset vital covariates) among others was forth put to checkmate this challenge (Zhou & Shen, 2001; Tibshirani, 1996). Saldana & Feng (2018) buttressed the assertion of Yu & Feng (2014) that the sparse estimation and variable selection technique in a parametric context was based on penalized (pseudo-) likelihood approach. According to Fan et al. (2014), the approach was based on two methods – the least absolute penalization method of Least Absolute Shrinkage and Selection Operator (LASSO) proposed by (Tibshirani, 1996 & Tibshirani, 1997) and folded concave penalization either by Minimax Concave Penalty (MCP) (Zhang, 2010) or by Smoothed Clipped Absolute Deviation (SCAD) (Fan & Li, 2001). The LASSO shrinkage least penalization which encompasses the ridge regression, strongly rely on irreparable necessary condition for its selection consistent for penalizing the absolute size of the regression coefficients. Its solution paths are piecewise linear in nature to allow and enable algorithm type to compute the solution path efficiently (Fan et al., 2014; Zhao & Yu, 2006). Unlike the LASSO approach, the folded concave penalty (MCP) requires reparable condition for selection of variable consistency to correct the bias in LASSO (Nardi & Rinaldo., 2012).

Both the LASSO and MCP use a penalty called turning parameter otherwise known as the regularized parameter η in correcting the irreparable and concavity. Different algorithms have been proposed for calculating the solution path of the estimates of the associated regression coefficients ($\hat{\beta}$). Among the algorithms was the Least Square Angle Regression (LARS) and homotopy algorithm by Osborne

et al. (2000) for updating the solution path. Zou & Li (2008) improved on the LARS algorithm by adding a Local Linear Approximation (LLA) to the penalized likelihood for a best and robust solution. In addition, Wu (2011) initiated an Ordinary Differential Equation (ODE) based solution path algorithm which forgone the maximum penalized likelihood estimate for quasi-penalized likelihood estimate to update the solution path by LARS-algorithm via some variable selection criteria such as Extended BIC (EBIC), Penalized BIC (PBIC) etc. by varying the regularized parameter in LASSO and MCP. Unlike the LARS-algorithm, Breheny & Huang (2011) adopted an adaptive rescaling method, the method allowed the range of the turning parameter η to be widen similarly to that of the linear regression. Moreover, Xu et al. (2017) used the quadratic approximation to update the path solution of the regression coefficients with random component that was best explained by Poisson and Binomial distributions. They came-up with a new solution path correction called approximate path for penalized likelihood estimators because it turns to be faster than the conventional coordinate descent method. They varied the turning parameter in LASSO and MCP before reaching the number of vital covariates to neutralize the problem.

The intent of this study is to improve on Xu et al. (2017)'s work by considering the gamma probability distribution random component due to its capability, not to only control, absolve, influence spread and skewness but to also streamline high-dimensional readings and measurements to $(0, \infty)$ as they are mostly represented in a positive continuous form. Furthermore, the approximate path for penalized likelihood estimators will be employed via the quadratic Newton-Raphson convergence technique to correct the updated solution path of $(\hat{\beta})$ by a stepwise and pre-selected turning parameter in the least absolute shrinkage penalty of LASSO and folded concave penalty of MCP.

Generalized Linear Models and Penalized Likelihood Estimations

Given a linear regressors function (a linear predictor) with response variable (Y_i) of a known conditional distribution

$$Y_i = \theta_i = \beta'X_i = \beta_0 + \beta_1X_{i1} + \beta_2X_{i2} + \cdots \beta_nX_{in} + \varepsilon_i \quad (1)$$

or in matrix form $Y = X\beta + \varepsilon$ where $Y = (y_1, \dots, y_n)^T$ is the response vector matrix, $\beta = (\beta_1, \dots, \beta_n)^T$ is vector matrix of the true parameter of the regression coefficients, X is the $n \times d$ design matrix of the covariates while $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$ is the vector

matrix of the random components that is gamma independently distributed with a smooth and invertible linearizing (link function) $g(\cdot)$, $\mu_i = E(Y_i)$ to the linear predictors

$$g(\mu_i) = \theta_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots \beta_n X_{in} \quad (2)$$

For an invertible link function (mean function)

$$\mu_i = g^{-1}(\theta_i) = g^{-1}(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots \beta_n X_{in}) \quad (3)$$

Fox (2008), Feng & Yu (2013), and Xu et al. (2017) defined the likelihood of any exponential family with n -dimensional observations (y_i, x_i) written in a canonical form in a GLMs context as

$$f(y; x, \beta) = c(y) \exp \left[\frac{y(\theta) - b(\theta)}{\alpha(\phi)} \right] \quad (4)$$

Where $\theta = X'\beta$, $b(\theta)$ is the natural parameter, $y(\theta)$ is the canonical form, $\alpha(\phi)$ nuisance parameter (constant) and $\phi \in (0, \infty)$.

The log-likelihood of $f(y; x, \beta)$ with respect to β of the n -dimensional proportionality in (4), that is, $\ln f(y; x, \beta) \propto \ell(\beta; y, x)$

$$\ell(y; \beta) = \frac{\sum_{i=1}^n [y_i(\theta_i) - b(\theta_i)]}{n} \quad (5)$$

Saldana & Feng (2018) and Xu et al. (2017) defined the objective function Penalized Likelihood Estimator (PLE) for estimating n -dimensional vector of ' β ' via least absolute penalization or folded concave penalization function as

$$L(\hat{\beta}) = \hat{\beta}(\eta) = \arg \min_{\beta} \left[\eta \|\beta\| - \frac{1}{n} \sum_{i=1}^n \ell(y; \beta) \right] \quad (6)$$

$$\hat{\beta}(\eta) = \arg \min_{\beta} [-\ell(y; \beta) + f_{\eta}(\beta)] = \frac{1}{n} \left[\sum_{i=1}^n [y_i \theta(\beta)_i - b(\theta(\beta)_i)] \right] + \eta \|\beta\|$$

LIKELIHOOD ESTIMATION OF Γ DISTRIBUTED RESPONSE VARIABLE

where $f_\eta(\cdot)$ is the LASSO or MCP function $\ni \eta > 0$ is the regularized parameter that controls the strength of the log-likelihood for a unique minimizer ($\hat{\beta}$) that must meet the first-order Karush-Kuhn-Tucker (KKT) conditions as follows:

$$\begin{cases} \frac{\partial \ell}{\partial \beta_0} = 0 \\ \frac{\partial \ell}{\partial \beta_i} = \eta \operatorname{sgn}(\hat{\beta}_k) & \text{for } i = 1, \dots, n \ni \hat{\beta}_k \neq 0 \\ \frac{\partial \ell}{\partial \beta_i} < \eta & \text{for } i = 1, \dots, n \ni \hat{\beta}_k = 0 \end{cases} \quad (7)$$

with a grid penalty of

$$\eta \geq \eta_{\max} = \operatorname{Max} \left| \frac{\partial \ell}{\partial \beta_i} \right| \quad \beta_i = 0 \quad 1 \leq i \leq n \quad (8)$$

Equation (8) can then be updated via

$$C_a = \left\{ 1 \leq i \leq n; \left| \frac{\partial \ell}{\partial \beta_i} \right| \geq \eta_a \right\} \cup \{0\}$$

Using an approximation quadratic stepwise and size iteration of $\Delta_a = \eta_{a+1} + \eta_k$, then

$$\hat{\beta}_{C_a}^{(a+1,0)} = \hat{\beta}_{C_a}^{(a)} + H^{(a)} \cdot \Delta_\eta + \frac{1}{2} Z^{(a)} \cdot \Delta_a^2 \quad (9)$$

where $H^{(a)}$ and $Z^{(a)}$ are the first and second derivatives of $\hat{\beta}$ w.r.t. η .

Gamma Probability Density Function

Let y be a $1 \times n$ -dimensional vector for a response variable that follows a gamma distribution of a continuous exponential family with probability density function.

$$f(y) = \left(\frac{y}{\omega}\right)^{\alpha-1} \times \frac{e\left(-\frac{y}{\omega}\right)}{\omega\Gamma(\alpha)} \quad \text{for } y \in (0, \infty); \quad \Gamma(\alpha) = \int_0^{\infty} e^{-x} x^{\alpha-1} dx$$

with scale parameter $\omega > 0$; shape parameter $\alpha > 0$; where the scale parameter influences the spread (center of location) of the distribution while the shape parameter controls the skewness parameter of the distribution such that as the shape parameter increases as the distribution symmetric expands. It is to be recalled that in a GLMs context $E(Y) = \omega\alpha; V(Y) = \phi\mu_i^2 \ni \phi = 1/\alpha$.

The Gamma LASSO Penalization Likelihood Estimate

Note the canonical link and inverse functions for gamma density function are μ^{-1} & $\eta_i^{-1} = (\beta' X_i)$ respectively.

$$\begin{aligned} L(\beta) &= -\frac{1}{n} \sum_{i=1}^n \left\{ y_i (\beta' X_i) - (\beta' X_i)^{-1} \right\} + \eta \sum_{k=1}^d |\beta_k| \\ &= \frac{1}{n} \sum_{i=1}^n \left\{ (\beta' X_i)^{-1} - y_i (\beta' X_i) \right\} + \eta \sum_{k=1}^d |\beta_k|. \end{aligned}$$

The KKT conditions are:

$$\left\{ \begin{array}{l} \frac{1}{n} \sum_{i=1}^n \left[-\left(\frac{\hat{\beta}'}{x_i} \right)^{-2} - y_i \right] - x_{ik} = \frac{1}{n} \sum_{i=1}^n \left[-\left(\frac{x_i}{\hat{\beta}'} \right)^2 - y_i \right] - x_{ik} = \eta \operatorname{sgn}(\hat{\beta}_k); \hat{\beta}_k \neq 0 \\ \left| \frac{1}{n} \sum_{i=1}^n \left[-\left(\frac{x_i}{\hat{\beta}'} \right)^2 - y_i \right] - x_{ik} \right| \leq \eta; \hat{\beta}_k = 0 \\ \hat{\beta}_0 = \frac{\sum_{i=1}^n (\hat{\beta}')^2}{\sum_{i=1}^n (y_i x_i^2)} \end{array} \right.$$

LIKELIHOOD ESTIMATION OF Γ DISTRIBUTED RESPONSE VARIABLE

Setting C_a as defined

$$C_a = \left\{ k : \left| \frac{1}{n} \sum_{i=1}^n \left[y_i - \left(\hat{\beta}^{(a)} x_i \right)^{-1} x_{ik} \right] \right| \geq \eta_a \right\} \cup \{0\}$$

Updating gives

$$Var^{(a)} = diag \left\{ \frac{e^{\frac{2 \left(\hat{\beta}^{(a)} x_1 \right)}{\alpha}}}{\alpha}, \dots, \frac{e^{\frac{2 \left(\hat{\beta}^{(a)} x_n \right)}{\alpha}}}{\alpha} \right\}$$

Recall that the $V(Y) = \phi \mu_i^2$ in the GLM context, where $\phi = \frac{1}{\omega}$

$$h_{initial \neq 0}^{(a)} = - \left[X'_{c_a - \{0\}} Var^{(a)} X_{c_a - \{0\}} \right]^{-1} \text{sgn} \left(\hat{\beta}_{c_a - \{0\}}^{(a)} \right)$$

$$\sum^{(a)} = diag \left[Var^{(a)} X_{c_a - \{0\}} h_{initial \neq 0}^{(a)} \right]$$

$$Z_{initial \neq 0}^{(a)} = - \left[X'_{c_a - \{0\}} Var^{(a)} X_{c_a - \{0\}} \right]^{-1} \times X'_{c_a - \{0\}} \sum^{(a)} X_{c_a - \{0\}} \times h_{initial \neq 0}^{(a)}$$

The Correction via Newton-Raphson Approximation for Convergence

$$\frac{\partial L^{(a)}}{\partial \beta_{c_a}} = \frac{1}{n} X'_{c_a} \left(\frac{e^{\frac{2 \left(\hat{\beta}^{(a)} x_1 \right)}{\alpha}}}{\alpha} - Y \right) + \eta_{(a)} \text{sgn} \left(0, \hat{\beta}_{c_a - \{0\}}^{(a)} \right)'$$

$$\frac{\partial^2 L^{(a)}}{\partial \beta_{c_a} \partial \beta'_{c_a}} = \frac{1}{n} X'_{c_a} Var^{(a)} X_{c_a}$$

$$\hat{\beta}_{c_a}^{(a,k+1)} = \hat{\beta}_{c_a}^{(a,k)} - \left(\frac{\partial^2 L^{(a)}}{\partial \beta_{c_a} \partial \beta'_{c_a}} \right)^{-1} \left(\frac{\partial L^{(a)}}{\partial \beta_{c_a}} \right) \quad (10)$$

Equation (10) is the corrected solution path of the coefficients vector for LASSO penalty. $\hat{\beta}_{c_a}^{(a,k+1)}$ is to be iterated until the optimal solution of vector $\hat{\beta}_{c_a}$ is reached. The Newton-Raphson was chosen because it converges faster than the coordinate descent method.

The Gamma MCP Penalization Likelihood Estimate

Zhang (2010), Zhang & Huang (2008) defined the penalty for MCP to be

$$f_{\eta}(\beta) = p_{\eta,\gamma}(t) = \eta \int_0^t \left(1 - \frac{x}{\eta\gamma}\right)_+ \partial t$$

where $P_{\eta\gamma}(t)$ is the Cumulative Distribution (CD) of X and $\gamma > 0$ measures the concavity of the penalty and η is the regularized parameter.

$$L(\beta) = \frac{1}{n} \sum_{i=1}^n \left\{ y_i (\beta' X_i) - (\beta' X_i)^{-1} \right\} + \eta \sum_{k=1}^d \int_0^t \left(1 - \frac{x}{\eta\gamma}\right)_+ \partial t$$

$$\begin{cases} \frac{1}{n} \sum_{i=1}^n \left[-\left(\frac{x_i}{\hat{\beta}'} \right)^2 - y_i \right] - x_{ik} = \eta \left(1 - \frac{|\hat{\beta}_k|}{\eta\gamma} \right) \text{sgn}(\hat{\beta}_k) \\ \frac{1}{n} \sum_{i=1}^n \left[-\left(\frac{x_i}{\hat{\beta}'} \right)^2 - y_i \right] - x_{ik} = 0 \\ \left| \frac{1}{n} \sum_{i=1}^n \left[-\left(\frac{x_i}{\hat{\beta}'} \right)^2 - y_i \right] - x_{ik} \right| \leq \eta \\ \hat{\beta}_0 = \frac{\sum_{i=1}^n (\hat{\beta}')^2}{\sum_{i=1}^n (y_i x_i^2)} \end{cases}$$

Defining the active set

$$C_a = \{C_{a-1} \cup \mathbf{N}_a\} - \mathbf{M}_a \ni$$

$$\mathbf{N}_a = \left\{ k = 1, \dots, d \setminus C_{a-1} : \left| \frac{1}{n} \sum_{i=1}^n \left[- \left(\frac{x_i}{\hat{\beta}'} \right)^2 - y_i \right] - x_{ik} \right| > \eta_a \right\}$$

$$\mathbf{M}_a = \left\{ k \ni C_{a-1} \cap C_{a-2} : \text{sgn}(\hat{\beta}_k^{(a-1)}) \text{sgn}(\hat{\beta}_k^{(a-2)}) < 0 \right\}$$

For rescaling of γ

$$\mathbf{H}_{in \neq 0}^{(a)} = \left(\frac{1}{n} X'_{C_a} \text{Var}^{(a)} X'_{C_a - \{0\}} - \Psi_{\text{inf}} \right)^{-1} \text{sgn}(\hat{\beta}_{C_a - \{0\}}^{(a)})$$

$$\mathbf{Z}_{in \neq 0}^{(a)} = \left(\frac{1}{n} X'_{C_a} \text{Var}^{(a)} X'_{C_a - \{0\}} - \Psi_{\text{inf}} \right)^{-1} \cdot X'_{C_a - \{0\}} \Sigma^{(a)} X_{C_a - \{0\}} \mathbf{H}_{in \neq 0}^{(a)}$$

$\ni \text{Var}^{(a)}$ and $\Sigma^{(a)}$ are as defined initially; $in = \text{initial value}$,

$$\Psi = \left(\frac{1}{\gamma_1}, \dots, \frac{1}{\gamma_n} \right), \text{sgn}(\hat{\beta}_{C_a - \{0\}}^{(a)}) = \begin{pmatrix} \text{sgn}(\hat{\beta}_{C_a,1}^{(a)}) \mathbf{I} \left\{ \left| \hat{\beta}_{C_a,1}^{(a)} \right| \eta^{(a)} \gamma \right\} \\ \vdots \\ \text{sgn}(\hat{\beta}_{C_a,na}^{(a)}) \mathbf{I} \left\{ \left| \hat{\beta}_{C_a,nk}^{(a)} \right| \eta^{(a)} \gamma \right\} \end{pmatrix}$$

$$\text{sgn}(\hat{\beta}_k^{(a)}) = \frac{1}{n} \sum_{i=1}^n \left[- \left(\frac{x_i}{\hat{\beta}'} \right)^2 - y_i \right] - x_{ik}$$

The Correction via Newton-Raphson Approximation for Convergence

$$\begin{aligned} \frac{\partial L^{(a)}}{\partial \beta_{C_a}} &= -\frac{1}{n} X'_{C_a} Var^{(a)} \left[\left(Var^{(a)} \right)^{-1} \left(Y - \frac{e^{2\left(\hat{\beta}_{C_a}^{(a)} x_i\right)}}{\alpha} \right) - X_{C_a} \hat{\beta}_{C_a}^{(a)} \right] + \eta_{(a)} \operatorname{sgn}\left(0, \hat{\beta}_{C_a \setminus \{0\}}^{(a)}\right) \\ &\quad \times \left(1 - \frac{\left| \hat{\beta}_{C_a \setminus \{0\}}^{(a)} \right|}{\eta_{(a)} \gamma} \right) \\ \frac{\partial L^{(a)}}{\partial \beta_{C_a} \partial \beta'_{C_a}} &= \frac{1}{n} X'_{C_a} Var^{(a)} - \Psi_{\inf} \\ \hat{\beta}_{C_a}^{(a,k+1)} &= \hat{\beta}_{C_a}^{(a,k)} - \left(\frac{\partial L^{(a)}}{\partial \beta_{C_a} \partial \beta'_{C_a}} \right)^{-1} \left(\frac{\partial L^{(a)}}{\partial \beta_{C_a}} \right) \end{aligned} \quad (11)$$

$\hat{\beta}_{C_a}^{(a,k+1)}$ is to be iterated until optimal solution of the vector matrix is reached. Equation (11) is the corrected solution path of the coefficients vector for MCP penalty

Selection Criterion of the Turning Parameter (η)

The execution of the penalized likelihood estimate solely boils down to the choice the turning point η in LASSO and (η, γ) in MCP. According to Yu & Feng (2014), Chen & Chen (2008), they claimed that can be achieved via information criteria such as AIC, BIC, HQ etc. and proposed an addition extra penalty to BIC called Extended BIC (EBIC) for $0 \leq \gamma < 1$ such that

$$EBIC_{\gamma}(s) = -2 \log L(\hat{\theta}_{(s)}) + v_{(s)} \log(n) + 2\gamma \log \left(\frac{d}{k} \right) \quad (12)$$

where $s \subset \{1, \dots, d\}$, $\theta_{(s)}$ is the parameter θ whose components outside s being set to be zero or some pre-selected values, $\hat{\theta}_{(s)}$ is the maximum likelihood estimator of $\theta_{(s)}$ while $v_{(s)}$ is the number of component(s).

Simulation

A simulation study of eight (8) covariates (d) were generated for a six (6) sample size (n) such that the response variable was generated from a gamma distribution with shape parameter 0.5, 1, 2 and 3.2 respectively, that is $Y \sim \text{Gamma}(\alpha)$. The size of the responses corresponds to the sample size to satisfy the condition of $d > n$. The gamma responses and the covariates were first subjected to the GLMs to ascertain the influence of shape parameter (α) on corrected solution of the regression coefficients $\hat{\beta}$ and to the corresponding link function after which the result of different path solutions of penalized LASSO and MCP were obtained at different level of $\alpha = 0.5, 1, 2$, and 3.2 .

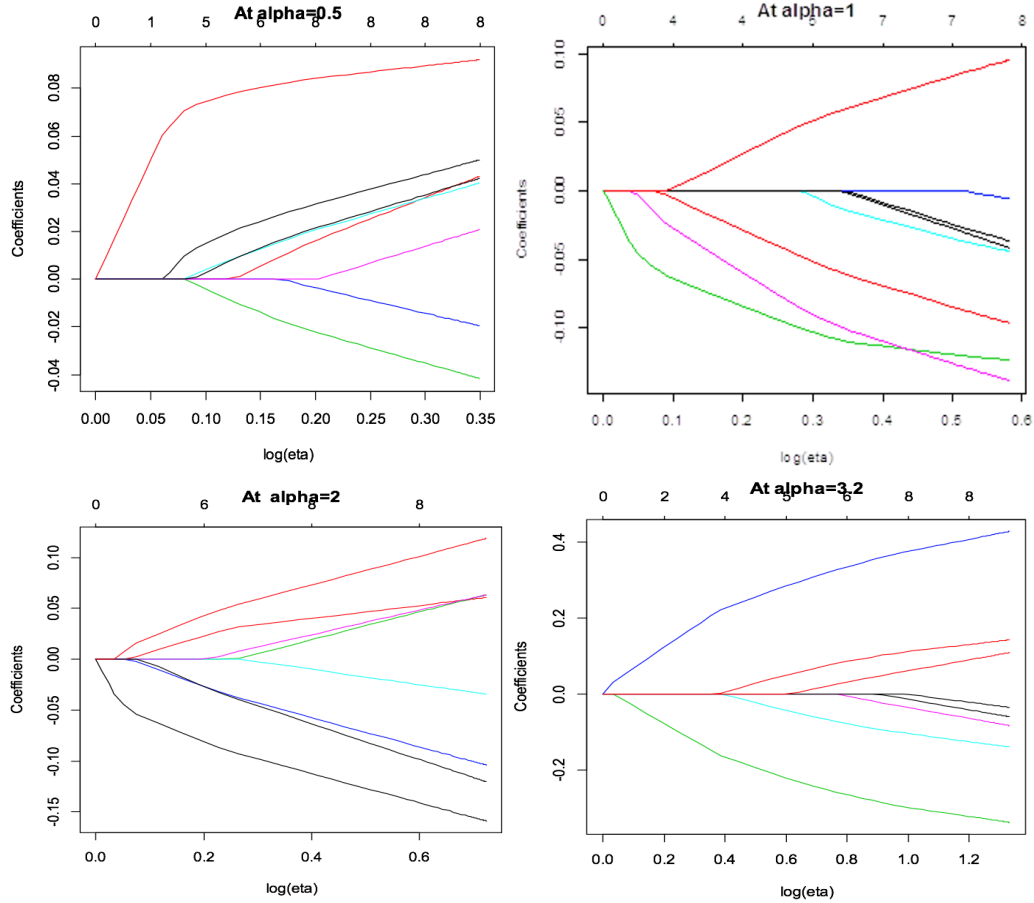


Figure 1. Different structure of the coefficients at four varied level of shape parameters.

Four shape parameters of (0.5, 1, 2, and 3.2) with their corresponding rates at (1.02, 1.44, 1.99, and 2.44) respectively were considered. This was due to Delignette-Muller & Dutang (2017) assertion that if $0 < \alpha < 1$, it is assumed that skewness threat is minimal, if $\alpha = 1$ the gamma distribution can be substituted for exponential distribution and if $\alpha > 3$, it means the response data has been highly captivated by the threat of skewness. Figure 1 is when shape parameters $\alpha_i i = 0.5, 1, 2, 3.2$ respectively. It is to be noted that one of the corrected solutions of the regression coefficients was far apart positively skewed when shape parameter (α) 0.5 and 1 respectively. More so, the corrected solutions of the coefficients are more evenly when shape parameters (α) are 2 and 3.2, which hinted that the positively skewed of the response variable has been controlled. The shape parameter must have controlled the skewness in the distribution of the response variable for a conformable and similar size, shape, effective and relative position of the response distribution for symmetric expansion and effect on the path solution of the regression coefficients.

Table 1. LASSO & MCP information criteria of at varied level of shape parameters.

Pen		$\alpha = 0.5$		$\alpha = 1$		$\alpha = 2$		$\alpha = 3.2$	
		LASSO	MCP	LASSO	MCP	LASSO	MCP	LASSO	MCP
1	<i>eta</i>	0.0977	0.0977	0.0614	0.0614	0.0890	0.0890	0.0218	0.0216
	<i>EBIC</i>	(5.2721)	(5.2544)	(3.7933)	(3.7933)	(4.3000)	(4.3000)	(2.3692)	(2.3692)
2	<i>eta</i>	0.0673	0.0673	0.0320	0.0320	0.0890	0.0890	0.0203	0.0200
	<i>EBIC</i>	(5.2313)	(5.2313)	(3.7431)	(3.7431)	(4.2324)	(4.2324)	(2.9990)	(2.9833)
3	<i>eta</i>	0.0509	0.0509	0.0320	0.0320	0.0890	0.0890	0.0277	0.0267
	<i>EBIC</i>	(5.2043)	(5.2042)	(3.7400)	(3.7400)	(4.2977)	(4.2977)	(3.0469)	(3.0462)
4	<i>eta</i>	0.0509	0.0509	0.0351	0.0351	0.0890	0.0890	0.0289	0.0297
	<i>EBIC</i>	(5.2085)	(5.2085)	(3.7410)	(3.7410)	(4.2323)	(4.2323)	(2.4673)	(2.4690)
5	<i>eta</i>	0.0509	0.0509	0.0464	0.0464	0.0890	0.0890	0.0310	0.0302
	<i>EBIC</i>	(5.2052)	(5.2052)	(3.7442)	(3.7442)	(4.2327)	(4.2327)	(2.5270)	(2.5231)
6	<i>eta</i>	0.0509	0.0509	0.0614	0.0614	0.0890	0.0890	0.0245	0.0242
	<i>EBIC</i>	(5.2056)	(5.2056)	(3.7591)	(3.7591)	(4.2356)	(4.2356)	(3.0056)	(3.0004)
7	<i>eta</i>	0.0811	0.0811	0.0614	0.0614	0.0890	0.0890	0.0890	0.0890
	<i>EBIC</i>	(5.2135)	(5.2135)	(3.8026)	(3.8022)	(4.2323)	(4.2323)	(2.7468)	(2.7461)
8	<i>eta</i>	0.0509	0.0509	0.0320	0.0320	0.0890	0.0890	0.0316	0.0313
	<i>EBIC</i>	(5.2040)	(5.2040)	(3.7400)	(3.7400)	(4.2431)	(4.2432)	(3.0010)	(2.9902)

From Table 1, when the shape parameter equals 0.5, the minimal EBIC for both LASSO and MCP of the corrected solutions of the coefficients was (5.2040) if covariate x_8 is penalized, when the shape parameter equals 1, the least EBIC for

LIKELIHOOD ESTIMATION OF Γ DISTRIBUTED RESPONSE VARIABLE

both LASSO and MCP was (3.7400) provided either covariate x_3 or x_8 is been penalized. In addition, when the shape parameter equals 2, the minimum EBIC for both LASSO and MCP was (4.2323) if and only if covariate x_4 or x_7 is been penalized. Furthermore, when shape parameter equals 3.2 (already hinted by the control of positively skewed of the response variable from Fig. 1) the diminutive size (smallest) value of EBIC for both LASSO and MCP of the corrected solution of the whole system was (2.3692) if x_1 is penalized for $\eta > 0$ and $\gamma = [0,1)$ values in MCP iteration. So,

$$\left\{ \begin{array}{l} Y_{LASSO} = 3.15 - 0.02x_2 + 0.19x_3 - 0.43x_4 \\ \quad + 0.02x_5 + 0.11x_6 - 0.07x_7 - 0.26x_8 \\ \text{for } \alpha = 3.2, \omega = 1.2; \eta = 0.0218; \text{ Adjusted } R^2 = 85.78; \\ MSE = 0.0156 \\ \\ Y_{MCP} = 3.16 - 0.03x_2 + 0.20x_3 - 0.44x_4 \\ \quad + 0.02x_5 + 0.11x_6 - 0.07x_7 - 0.26x_8 \\ \text{for } \alpha = 3.2, \omega = 1.2; \eta = 0.0216, \gamma = [0,1); \text{ Adjusted } R^2 = 86.05; \\ MSE = 0.0154 \end{array} \right. \quad (13)$$

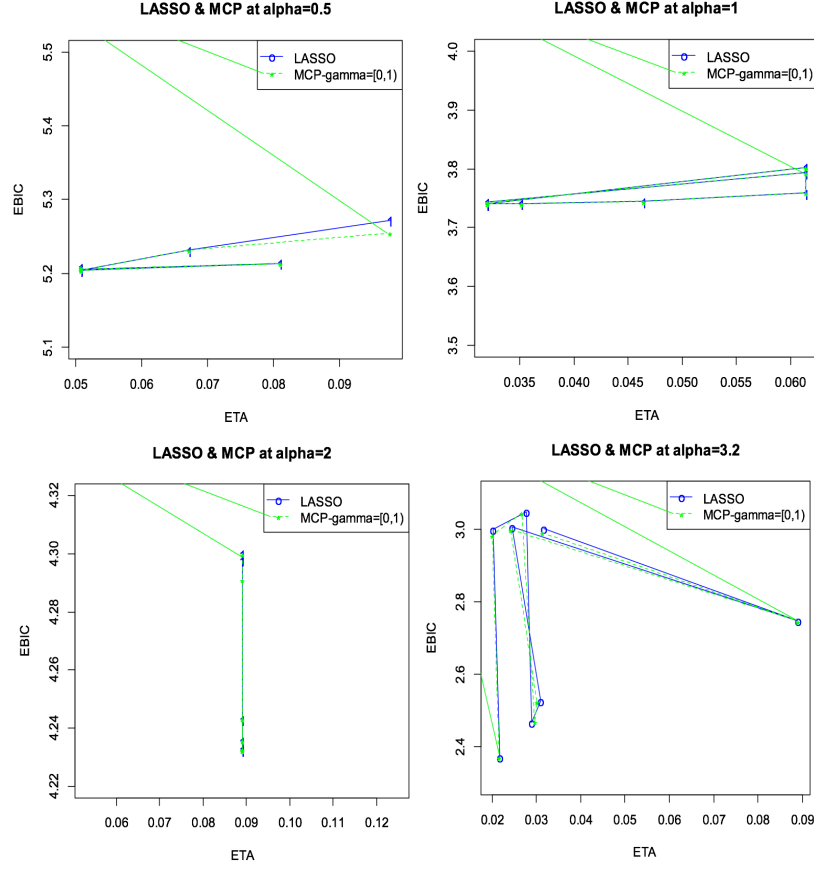


Figure 2. EBIC of LASSO and MCP at different established of turning parameters.

Table 2. Penalized regression coefficients of LASSO at $\alpha = 3.2$, $\omega = 1.2$, diminutive size of EBIC.

<i>Pen</i>	<i>Inter</i>	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
1	3.15	-	-0.02	0.19	-0.43	0.02	0.11	-0.07	-0.26
2	3.15	0.06	-	0.15	-0.4	0.03	0.08	-0.04	-0.32
3	3.21	0.17	-0.14	-	-0.44	-0.08	0.08	-0.15	-0.3
4	2.95	-0.11	0.04	0.26	-	-0.07	-0.1	-0.01	-0.29
5	3.14	0.05	-0.06	0.14	-0.44	-	0.1	-0.07	-0.31
6	3.21	0.1	-0.14	0.16	-0.48	-0.02	-	-0.1	-0.32
7	3.14	0.05	0.01	0.17	-0.41	0.02	0.09	-	-0.31
8	3.11	-0.13	-0.09	0.32	-0.47	0.06	0.11	-0.14	-

Table 3. Penalized regression coefficients of MCP at $\alpha = 3.2$, $\omega = 1.2$, diminutive size of EBIC.

LIKELIHOOD ESTIMATION OF Γ DISTRIBUTED RESPONSE VARIABLE

<i>Pen</i>	<i>Inter</i>	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
1	3.16	-	-0.03	0.20	-0.44	0.02	0.11	-0.07	-0.26
2	3.14	0.06	-	0.14	-0.40	0.04	0.08	-0.03	-0.31
3	3.21	0.17	-0.13	-	-0.43	-0.08	0.075	-0.15	-0.30
4	3.00	-0.11	0.03	0.27	-	-0.07	-0.09	-0.01	-0.25
5	3.13	0.05	-0.06	0.13	-0.44	-	0.09	-0.07	-0.31
6	3.21	0.11	-0.14	0.16	-0.48	-0.02	-	-0.10	-0.32
7	3.14	0.05	0.04	0.17	-0.41	0.02	0.09	-	-0.31
8	3.11	-0.13	-0.09	0.32	-0.47	0.06	0.11	-0.14	-

Conclusion

The ability of gamma independently distributed responses to moderate spread and control the influence of skewness in a corrected and updated path solutions of the regression coefficients via penalized regularization and variable selection technique of LASSO and MCP give viable option to handle the threat of known or unknown outlier(s) and oblique position of the response data. This work can be extended to the generalized gamma distribution with kurtosis parameter embedded to control the peakedness or flatness of the graph of the frequency distribution especially with respect to the concentration of values near the mean as compared with the normal distribution in order not to distort corrected regression coefficients from approaching the true values.

References

- Antoniadis, A., Gijbels, I., & Nikolova, M. (2011). Penalized likelihood regression for generalized linear models with non-quadratic penalties. *Annals of the Institute of Statistical Mathematics*, 63(3), 585-615. <https://doi.org/10.1007/s10463-009-0242-4>
- Berk, R., Brown, L., Buja, A., Zhang, K., & Zhao, L. (2013). Valid post-selection inference. *Annals of Statistics*, 41(2), 802-837. <https://doi.org/10.1214/12-aos1077>
- Breheny, P., & Huang, J. (2011). Co-ordinate descent algorithms for non-convex penalized regression with applications to biological feature selection. *The Annals of Applied Statistics*, 5(1), 232-253. <https://doi.org/10.1214/10-aos388>
- Chen, J., & Chen, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. *Biometrika*, 95(3), 759-771. <https://doi.org/10.1093/biomet/asn034>.

- Delignette-Muller, M. L., & Dutang, C. (2017). *fitdistrplus: An R Package for Fitting Distributions*. *Journal of Statistical Software*, 64(4), 1-34.
<https://doi.org/10.18637/jss.v064.i04>
- Dette, H., Neumeyer, N. & Pilz, K. F. (2006). A simple nonparametric estimator of a monotone regression function. *Bernoulli*, 12(3), 469–490.
<https://doi.org/10.315/bj/1151525131>
- Fan, J., Lingzhou, X., & Hui, Z. (2014). Strong oracle optimality of folded concave penalized estimation. *The Annals of Statistics*, 42(3), 819-849. <https://doi.org/10.1214/13-aos1198>
- Fan, J., R. Samworth, & Wu, Y. (2009). Ultrahigh dimensional feature selection: beyond the linear model. *The Journal of Machine Learning Research*, 10, 2013-2038
- Fan, J., & Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(5), 849-911. <https://doi.org/10.1111/j.1467-9868.2008.00674.x>
- Fan, J., & Li, R. (2001). Variable selection via non-concave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* 96(456), 1348–1360. <https://doi.org/10.1198/016214501753882273>
- Feng, Y., & Yu, Y. (2013). Consistent cross-validation for tuning parameter selection in high-dimensional variable selection. *arXiv.org*.
<http://arxiv.org/abs/1308.5390>
- Fox, J. (2008). *Maximum-Likelihood Estimation of Generalized Linear Models: Applied Regression Analysis and Generalized Linear Models* (2nd Ed.). New York: Springer.
- Osborne, M., Presnell, B., & Turlach, B. (2000): A new approach to variable selection in least squares problems. *IMA Journal of Numerical Analysis*, 20(3), 389–404.
<https://doi.org/10.1093/imanum/20.3.389>
- Nardi, Y., & Rinaldo, A. (2012). The log-linear group-lasso estimator and its asymptotic properties. *Bernoulli*, 18(3), 945–974. <https://doi.org/10.3150/11-bej364>
- Saldana, D. F., & Feng, Y. (2018). SIS: An r package for sure independence screening in ultrahigh-dimensional statistical models. *Journal of Statistical Software*, 83(2). <https://doi.org/10.18637/jss.v083.i02>
- Tibshirani, R. J. (1997). The lasso method for variable selection in the cox model. *Statistics in Medicine*, 16(4), 385-395. [https://doi.org/10.1002/\(sici\)1097-0258\(19970228\)16:4<385::aid-sim380>3.0.co;2-3](https://doi.org/10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3)

LIKELIHOOD ESTIMATION OF Γ DISTRIBUTED RESPONSE VARIABLE

Tibshirani, R. J. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288.

<https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>

Watkins, C. J. (2001). *Introduction to Statistical methodology: Maximum Likelihood*. Retrieved from <http://math.arizona.edu/~jwatkins/statbook.pdf>

Wu, Y. (2011). An ordinary differential equation-based solution path algorithm. *Journal of Non-parametric Statistics*, 23(1), 185–199.

<https://doi.org/10.1080/10485252.2010.490584>

Xu, J., Kenneth, L., Eric, C. C. (2017). Generalized linear model regression under distance-to-set penalties. *31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA*. 1-11.

Yu, Y., & Feng, Y. (2014). APPLE: approximate path for penalized likelihood estimators. *Statistics and Computing*, 24(5), 803-819, 803–819.

<https://doi.org/10.1007/s11222-013-9403-7>

Zhang, C. H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38(2), 894-942. <https://doi.org/10.1214/09-aos729>

Zhang, C.H., & Huang, J. (2008). The sparsity and bias of the lasso selection in high-dimensional linear regression. *Annals of Statistics*, 36(4), 1567-1594.

<https://doi.org/10.1214/07-aos20>

Zou, H., & Li, R. (2008). One-step sparse estimates in non-concave penalized likelihood models. *The Annals of Statistics*, 36(4), 1509-1533.

<https://doi.org/10.1214/00536070000000802>

Zhao, P., & Yu, B. (2006). On model selection consistency of Lasso. *Journal of Machine Learning Research*, 7, 2541–2563.

Zhou, S. & Shen, X. (2001). Spatially adaptive regression splines and accurate knot selection schemes. *Journal of the American Statistical Association*, 96(453), 247-259.

<https://doi.org/10.1198/01621407750332820>