Asian Journal of Fuzzy and Applied Mathematics (ISSN: 2321 – 564X)
Volume 8– Issue 3, December 2020

Non-Convex Penalized Estimation of Count Data Responses *via* Generalized Linear Model (GLM)

Rasaki Olawale Olanrewaju[1] and Johnson Funminiyi Ojo[2]

[1] University of Ibadan
Department of Statistics (Ibadan, Nigeria)
Email: rasakiolawale [AT] gmail.com

[2] University of Ibadan
Department of Statistics (Ibadan, Nigeria)
Email: jfunminiojo [AT] yahoo.co.uk

ABSTRACT— ***This study provided a non-convex penalized estimation procedure via Smoothed Clipped Absolute Deviation (SCAD) and Minimax Concave Penalty (MCP) for count data responses to checkmate the problem of covariates (d) exceeding the sample size (n). The Generalized Linear Model (GLM) approach was adopted in obtaining the penalized functions needed by the MCP and SCAD non-convex penalizations of Binomial, Poisson and Negative-Binomial related count responses regression. A case study of the colorectal cancer with six (6) covariates against sample size of five (5) was subjected to the non-convex penalized estimation of the three distributions. It was revealed that the non-convex penalization of Binomial regression via MCP and SCAD best explained four un-penalized covariates needed in determining whether surgical or therapy ideal for treating the turmoil.***

Keywords— Count Data, Minimax Concave Penalty (MCP), Non-convex penalization, Smoothed Clipped Absolute Deviation (SCAD).

---

1. INTRODUCTION

Linear regression analysis of high-throughput and high dimensional data in bioinformatics, neurosciences, clinical studies etc. often impede with the number of covariates (d) exceeding the number of the sample size (n) [14, 17]. Efficient methods of variable selection via shrinking of covariate(s) and sparse estimation of regression coefficients techniques have been propounded based on penalized likelihood function (loss function) and regularizing parameter $''\eta''$. Among the techniques proposed to circumvent aforesaid challenge is the folded concave penalty function of Least Absolute Shrinkage and Selection Operator (LASSO) proposed by [11] and either by Smoothed Clipped Absolute Deviation (SCAD) or minimax concave loss function of Minimax Concave Penalty (MCP) as proponed by [3] and [15] respectively.

Unlike the convex penalty, where the likelihood functions of LASSO, SCAD and MCP influence biasedness in their parameter estimations via absolute values, non-convex (non-concave) penalty of MCP and SCAD relieves the absolute value constrain in concave LASSO in order to eliminate the biasedness influence [18, 6]. Non-convex penalty satisfies a wide range of statistical properties; ability not to only fix finite estimates of regression coefficients (β_i) but also to estimate true zero regression coefficients with their probabilities approximately equals one to confirm their sufficiency. In addition, the ability of non-convex MCP and SCAD to estimate non-zero coefficients as if the true sparsely is known [10].

In this paper, the Probability Mass Functions (PMFs) of count data responses (dependent variables) of Binomial, Poisson and Negative-Binomial distributions will be assumed to follow a linear regression responses when $d > n$. The non-convex penalized estimations of these count linear regression responses will be subdued to loss (penalized) functions of SCAD and MCP via Generalized Linear Model (GLM). Moreover, the solution of the penalized regression coefficients will be via proximal coordinate iterative procedure because of its tractable global solution faster rate of convergence for a pre-selected regularized parameter $''\eta''$ and high dimensional selection criteria.

2. SPECIFICATION OF THE NON-CONVEX PENALIZED ESTIMATION VIA GLM

Given independent random variables and a set of covariates (y_i, x_i) $i = 1, \ldots\ldots, n$ for a random sample from a linear regression

$$Y = X\beta + \varepsilon \tag{1}$$

Where $Y = (y_1, y_2, \ldots, y_n)^T$ is the response vector matrix, X is the n by d design matrix, $\beta = (\beta_0, \beta_1, \ldots, \beta_n)^T$ is the vector of regression coefficients while $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^T$ is the vector matrix of the random component of a Generalized Linear Models (GLM) such that its Probability Density Function PDF or Probability Mass Function (PMF) belongs to the exponential family;

$$f(y; x, \beta) = f(y_i; \theta_i, \phi) = c(y; \theta_i)\exp\left[\frac{y_i(\theta_i) - b(\theta_i)}{a(\phi)}\right] \tag{2}$$

Where $(\theta_1, \theta_2, \ldots, \theta_n)^T = X\beta = b(\theta_i)$ is the natural or canonical parameter, $y_i(\theta_i)$ is the canonical form, $a(\phi)$ is the scale parameter for $\phi \in (0, \infty)$, $c(y; \theta_i)$ is the function with "y" only.

[3], [7] and [8] maintained that the approach for estimating regression coefficients of a high dimensional data via non-convex regularization or penalization to be

$$Q(\beta) = \underset{\beta \in \mathbb{R}^d}{Min}\left\{\frac{1}{n}L(\beta) + f_\eta(\beta)\right\} \tag{3}$$

$$= \underset{\beta \in \mathbb{R}^d}{Min}\left\{\frac{1}{n}\left(X\beta - y\right)^2 + f_\eta(\beta)\right\} \tag{4}$$

Where $L(\beta)$ is the loss function (the log function of equation (2)), $f_\eta(\beta)$ is the non-convex penalized or regularized function of either for SCAD or for MCP with turning parameter (otherwise known as regularized parameter "η" that must satisfy the standard optimization solution of β_η for a first-order Karush-Kuhn-Tucker (KKT) condition of

$$h = \partial\left\{L(\beta_\eta) + f_\eta(\beta_\eta)\right\} \tag{5}$$

[3], [11] and [17] defined non-convex penalty for SCAD as

$$f_\eta(\beta_k) = \begin{cases} \eta|\beta_k| & if\ \beta_k \le \eta \\ \dfrac{a\eta|\beta_k| - \frac{1}{2}(\beta_k^2 + \eta^2)}{a-1} & if\ \eta < |\beta_k| \le a\eta \\ \dfrac{\eta^2(a-1)}{2(a-1)} & if\ \beta_k > \eta \end{cases} \tag{6}$$

For "a" which is a fixed parameter > 2; $\eta > 0$

Also, [6], [15] and [16] claimed that the non-convex penalty for MCP to be

$$f_\eta(\beta_k) = \begin{cases} \eta\left(|\beta_k| - \dfrac{\beta_k^2}{2\eta b}\right) & |\beta_k| \le \eta b \\ \dfrac{\eta^2 b}{2} & |\beta_k| > \eta b \end{cases} \tag{7}$$

Otherwise, $$f_\eta(\beta_k)=\begin{cases}\dfrac{\operatorname{sgn}(\beta_k)(|\beta_k|-\eta)_+}{1-b} & if\ |\beta_k|\le\eta b\\ \beta_k & if\ |\beta_k|>\eta b\end{cases}$$

"b" Which is a fixed parameter >0

The non-convexepenalty $f_\eta(\beta_k)$ can be decomposed into sum of penalty and sum of the concave part;

$$f_\eta(\beta)=\sum_{k=1}^{d} f_\eta(\beta_k)=\eta\|\beta\|+\sum_{k=1}^{d} g_\eta(\beta_k)$$

This could be simplified by rewriting as

$$G_\eta(\beta)=\sum_{k=1}^{g} g_\eta(\beta_k)=f_\eta(\beta)-\eta\|\beta\| \qquad (8)$$

$G_\eta(\beta)$ connotes the disintegration of concave part of the non-convexepenalty $f_\eta(\beta)$.

[1], [9], [12] and [13] asserted that the regularized solution for the regression coefficients could be compressed to a coordinate proximal method via Newpton-Raphson method iteration for update via

$$\varphi_{L_m^c},\eta_m\left(\beta,\beta_m^{c+1}\right)=\beta_m^{c-1}-\frac{1}{L_m^c}H_\eta \qquad (9)$$

such that, $H_\eta=\nabla L(\beta_\eta)+\nabla G_\eta(\beta_\eta)$ for a chosen "η" and surrogate $L_m^c>0$ form of L_{η_m} , β_m^c corresponds to the c^{th} iteration within t^{th} path solution and "N" the number of path stages.

$$N=\frac{\log\left(\dfrac{\eta_0}{\eta_{nd}}\right)}{\log(\gamma^{-1})}$$

η_{nd} scales of sample size "n" and the dimension "d" and chosen γ , and via the selection of turning parameter (regularized parameter) "η" by model selecon criteria of either[2], [4], [5], or [14].

$$AIC=-2L(\beta)+2df(\eta)\ ,\ BIC=-2L(\beta)+\log(N)df(\eta)$$

$$CAIC=-2L(\beta)+\log(N+1)df(\eta)$$

$$EBIC_\eta(s)=-2\log L(\beta)+v_{(s)}\log(N)+2\gamma\log\binom{d}{k}$$

Where "df" is the degree of freedom of non-zero parameters. $0\le\gamma<1$ $s\subset\{1,\ldots\ldots,d\}$,$\theta_{(s)}$ is the parameter β whose components outside "s" being set to be zero or some pre-selected values, $\beta_{(s)}$ is the maximum likelihood estimator of $\beta_{(s)}$ while $v_{(s)}$ is the number of component(s).

3. PENALIZED ESTIMATION VIA MCP AND SCAD

3.1.1 Binomial Penalized Regression Estimation

Assume $Y_i \sim Bin(n_i, p_i)$. Then the PMF; $P(p_i; y_i, n_i) = \binom{n_i}{y_i} p_i^{y_i}(1-p_i)^{n_i - y_i}$ $y_i = 1, \ldots\ldots, n_i$

$$= \exp\left\{ y_i \log \frac{p_i}{1-p_i} + n_i \log(1-p_i) - \log \binom{n_i}{y_i} \right\} \qquad (10)$$

$$b(\theta_i) = n_i(1+\exp(\theta_i)),\ \theta_i = \log \frac{p_i}{1-p_i}, c(y_i;\theta_i) = -\log \binom{n_i}{y_i}, \phi = 1,$$

$$a(\phi) = 1,\ E(Y_i) = n_i p_i,\ V(Y_i) = n_i p_i (1-p_i)$$

Where p_i is the probability of success and $1-p_i$ is the probability of failure.

3.1.2 SCAD Non-Convex Penalized Binomial Regression

From equation (10), the link function of Binomial $\mu(X\beta) = p_i = \frac{\exp(\beta' x_i)}{1+\exp(\beta' x_i)}$ called Logit function

$$Q_\eta(\beta) = -\frac{1}{n}\sum_{i=1}^{n}\left\{ (y_i \beta' x_i) - \log(1+\exp(\beta' x_i)) \right\} + G_\eta(\beta) \qquad (11)$$

$$f_{\eta(SCAD)}(\beta_k) = \begin{cases} \eta|\beta_k| & for \quad |\beta_k| \le \eta \\ \dfrac{-\beta_k^2 - 2a\eta|\beta_k| + \eta^2}{2(a-1)} & for \quad \eta < |\beta_k| \le a\eta \\ \dfrac{(a+1)\eta^2}{2} & for \quad |\beta_k| > a\eta \end{cases}$$

$$g_{\eta SCAD}(\beta_k) = \begin{cases} \dfrac{2\eta|\beta_k| - \beta_k^2}{2(a-1)} & for \quad \eta < |\beta_k| < a\eta \\ \dfrac{(a+1)\eta^2 - 2\eta|\beta_k|}{2} & for \quad |\beta_k| > a\eta \end{cases}$$

With non-convexloss function

$$\nabla_{BIN} L(\beta) = \frac{1}{n}\sum_{i=1}^{n} x_i \left\{ \frac{\exp(\beta' x_i)}{1+\exp(\beta' x_i)} - y_i \right\},$$

$$\nabla G_{\eta Bin}(\beta)_{SCAD} = \begin{cases} \eta_m \operatorname{sgn}(\beta_k) - \beta_k & for \quad \eta_m < |\beta_k| \le a\eta_m \\ -\eta_m \operatorname{sgn}(\beta_k) & for \quad |\beta_k| > a\eta_m \end{cases} \qquad \text{For } a > 2$$

Updating for optimal solution

$$\beta_{m(SCAD)}^{c+1} = \beta_m^{c-1} - \frac{1}{L_m^c}\left\{ \nabla G_{\eta Bin}(\beta^{c-1})_{SCAD} + \nabla_{BIN} L(\beta^{c-1}) \right\} V_{BIN}{}^{c-1} \qquad (12)$$

Where $V_{BIN} = diag\left\{ p_1^c(1-p_1^c), \ldots\ldots, p_n^c(1-p_n^c) \right\}$

Until convergent of the vector of regression coefficients is reached in equation (12) with starting value $\beta_m^c \neq 0$ and N[th] number of path stages.

3.1.3 MCP Non-Convex Penalized Binomial Regression

From equation (11),

$$f_{\eta MCP}(\beta_k)=\begin{cases}\left(\eta\left|\beta_k\right|-\dfrac{\beta_k^2}{2b}\right) & for \quad \left|\beta_k\right|\leq b\eta \\ \dfrac{b\eta^2}{2} & for \left|\beta_k\right|>b\eta\end{cases}$$

$$g_{\eta MCP}(\beta_k)=\begin{cases}-\dfrac{\beta_k^2}{2b} & for \quad \left|\beta_k\right|\leq b\eta \\ \left(\dfrac{b\eta^2}{2}-\eta\left|\beta_k\right|\right) & for \quad \left|\beta_k\right|>b\eta\end{cases}$$

$$\nabla_{MCP}G_{\eta}(\beta_k)=\begin{cases}-\dfrac{\beta_k}{b}\eta_m sign(\beta_k) & for \left|\beta_k\right|\leq b\eta_m \\ -\eta_m sign(\beta_k) & for \left|\beta_k\right|>b\eta_m\end{cases}$$

For $b>0$, updating for convergence,

$$\beta_{m(MCP)}^{c+1}=\beta_m^{c-1}-\frac{1}{L_m^c}\left\{\nabla_{MCP}G_{\eta Bin}(\beta^{c-1})+\nabla L(\beta^{c-1})\right\}V^{c-1} \qquad (13)$$

Until convergent of the vector of regression coefficients is reached in equation (13) with starting value $\beta_m^c \neq 0$ and N^{th} number of path stages.

3.2 Poisson Penalized Regression Estimation

Assume Y_i □ $Poisson(\lambda_i)$. Then the PMF;

$$P(\lambda_i;y_i)=\exp(\lambda_i)\frac{\lambda_i^{y_i}}{y_i!} \quad y_i=0,1,\ldots\ldots$$

$$=\exp\left\{y_i\log(\lambda_i)-\lambda_i-\log(y_i!)\right\}$$

$$\theta_i=\log(\lambda_i),\ b(\theta_i)=\exp(\theta_i),\ \phi=1,\ a(\phi)=1,\ c(y_i;\theta_i)=-\log(y_i!)$$

$$E(Y_i)=V(Y_i)=\lambda_i$$

Where λ_i is the parameter of interest

The link function of Poisson $\mu(X'\beta)=\exp(\beta' x_i)=\log(\lambda_i)$ called Probit function. Then the criterion for the SCAD penalized Poisson regression

$$Q_\eta(\beta)=\frac{1}{n}\sum_{i=1}^{n}\left\{\exp(\beta' x_i)-(y_i\beta' x_i)\right\}+G_\eta(\beta) \qquad (14)$$

$$g_{\eta SCAD}(\beta_k)=\begin{cases}\dfrac{2\eta\left|\beta_k\right|-\beta_k^2}{2(a-1)} & for \quad \eta<\left|\beta_k\right|<a\eta \\ \dfrac{(a+1)\eta^2-2\eta\left|\beta_k\right|}{2} & for \quad \left|\beta_k\right|>a\eta\end{cases}$$

$$\nabla_{POI}L(\beta)=\frac{1}{n}\sum_{i=1}^{n}x_i\left\{\exp(\beta' x_i)-y_i\right\}$$

$$\nabla G_{\eta Poi}(\beta)_{SCAD} = \begin{cases} \eta_m \operatorname{sgn}(\beta_k) - \beta_k & for \quad \eta_m < |\beta_k| \leq a\eta_m \\ -\eta_m \operatorname{sgn}(\beta_k) & for \quad |\beta_k| > a\eta_m \end{cases}$$

Updating for convergence, $\beta^{c+1}_{m(SCAD)} = \beta^{c-1}_m - \frac{1}{L^c_m}\left\{\frac{1}{n}\sum_{i=1}^{n} x_i \left\{\exp(\beta^{'c-1}x_i) - y_i\right\} + \nabla G_{\eta POI}(\beta^{c-1})_{SCAD}\right\} V^{c-1}_{POI}$

(15)

Where, $V_{POI} = diag\left\{\exp(\beta^{'c}x_1), \ldots\ldots, \exp(\beta^{'c}x_n)\right\}$

3.2.2 MCP Non-Convex Penalized Poisson Regression

From equation (14)

$$g_{\eta MCP}(\beta_k) = \begin{cases} -\frac{\beta_k^2}{2b} & for \quad |\beta_k| \leq b\eta \\ \left(\frac{b\eta^2}{2} - \eta|\beta_k|\right) & for \quad |\beta_k| > b\eta \end{cases} \quad \text{For } b > 0$$

$$\nabla_{POI} L(\beta) = \frac{1}{n}\sum_{i=1}^{n} x_i \left\{\exp(\beta' x_i) - y_i\right\}$$

$$\nabla_{MCP} G_{\eta POI}(\beta_k) = \begin{cases} -\frac{\beta_k}{b}\eta_m sign(\beta_k) & for\ |\beta_k| \leq b\eta_m \\ -\eta_m sign(\beta_k) & for\ |\beta_k| > b\eta_m \end{cases}$$

For $b > 0$

$$\beta^{c+1}_{m(MCP)} = \beta^{c-1}_m - \frac{1}{L^c_m}\left\{\frac{1}{n}\sum_{i=1}^{n} x_i \left\{\exp(\beta^{'c-1}x_i) - y_i\right\} + \nabla_{MCP} G_{\eta POI}(\beta^{c-1})\right\} V^{c-1}_{POI} \quad (16).$$

until convergence of the vector of regression coefficients is reached in equation (16) with starting value $\beta^c_m \neq 0$ and N[th] number of path stages.

3.3 Negative-Binomial Penalized Regression Estimation

Assume $Y_i \sim NB(k, p_i)$ .Then the PMF;

$$P(r, p_i; y_i) = \binom{y-1}{r-1} p_i^r \ (1-p_i)^{y_i} \quad y = r, r+1, r+2\ldots\ldots$$

$$= \exp\left\{ y_i \log(1-p_i) + r\log\frac{p_i}{1-p_i} - \log\binom{y-1}{r-1}\right\}$$

$$b(\theta) = -r\log\frac{1-\exp(\theta)}{\exp(\theta)},\ c(y_i;\theta_i) = -\log\binom{y-1}{r-1},\theta = \log(1-p_i),\ a(\phi) = 1,$$

$$E(Y_i) = \frac{r}{p} = \frac{r}{1-\exp(\theta)},\ V(Y_i) = \frac{r(1-p)}{p^2} = \frac{r\exp(\theta)}{(1-\exp(\theta))^2}$$

"r" Is the shape parameter for measuring the degree of clumping or aggregation (dispersion).

3.3.1 SCAD Non-Convex Penalized Negative-Binomial Regression

The link function of Negative Binomial $\mu(X^{'}\beta)=\eta=\dfrac{1}{1+\exp(\beta^{'}x_i)}=p_i$ called log of the log function.

Then the criterion for the SCAD penalized Negative-Binomial regression.

$$Q_\eta(\beta)=-\frac{1}{n}\sum_{i=1}^{n}\left\{(y_i\beta^{'}x_i)-\log\left\{\frac{1-\exp(\beta^{'}x_i)}{\exp(\beta^{'}x_i)}\right\}\right\}+G_\eta(\beta) \qquad (17)$$

$$g_{\eta SCAD}(\beta_k)=\begin{cases}\dfrac{2\eta|\beta_k|-\beta_k^2}{2(a-1)} & for \quad \eta<|\beta_k|<a\eta\\[2ex] \dfrac{(a+1)\eta^2-2\eta|\beta_k|}{2} & for \quad |\beta_k|>a\eta\end{cases}$$

Then the non-convexloss function

$$\nabla_{NB}L(\beta)=\frac{1}{n}\sum_{i=1}^{n}x_i\left\{\log\left\{\frac{1}{1+\exp(\beta^{'}x_i)}\right\}-y_i\right\}$$

$$\nabla G_{\eta NB}(\beta)_{SCAD}=\begin{cases}\eta_m\operatorname{sgn}(\beta_k)-\beta_k & for \quad \eta_m<|\beta_k|\le a\eta_m\\ -\eta_m\operatorname{sgn}(\beta_k) & for \quad |\beta_k|>a\eta_m\end{cases} \quad \text{For } a>2$$

Updating for optimal regression coefficient

$$\beta_{m(SCAD)}^{c+1}=\beta_m^{c-1}-\frac{1}{L_m^c}\left\{\nabla_{SCAD}G_{\eta NB}(\beta^{c-1})+\nabla L(\beta^{c-1})\right\}V_{NB}^{\ c-1} \qquad (18)$$

For $V_{NB}=diag\left\{\dfrac{r\exp(\beta^{'c}x_1)}{(\exp(\beta^{'c}x_1))^2},\ldots\ldots,\dfrac{r\exp(\beta^{'c}x_n)}{(\exp(\beta^{'c}x_n))^2}\right\}$

Until convergence of the vector of regression coefficients is reached in equation (18) with starting value $\beta_m^c\neq 0$ and N^{th} number of path stages.

3.3.2 MCP Non-Convex Penalized Negative-Binomial Regression

From equation (17),

$$g_{\eta MCP}(\beta_k)=\begin{cases}-\dfrac{\beta_k^2}{2b} & for \quad |\beta_k|\le b\eta\\[2ex] \left(\dfrac{b\eta^2}{2}-\eta|\beta_k|\right) & for \quad |\beta_k|>b\eta\end{cases} \quad \text{For } b>0$$

$$\nabla_{NB}L(\beta)=\frac{1}{n}\sum_{i=1}^{n}x_i\left\{\log\left\{\frac{1}{1+\exp(\beta^{'}x_i)}\right\}-y_i\right\}$$

$$\nabla_{MCP}G_{\eta NB}(\beta_k)=\begin{cases}-\dfrac{\beta_k}{b}\eta_m sign(\beta_k) & for \ |\beta_k|\le b\eta_m\\ -\eta_m sign(\beta_k) & for \ |\beta_k|>b\eta_m\end{cases}$$

Updating for optimal regression coefficient

$$\beta_{m(MCP)}^{c+1}=\beta_m^{c-1}-\frac{1}{L_m^c}\left\{\nabla_{MCP}G_{\eta NB}(\beta^{c-1})+\nabla L(\beta^{c-1})\right\}V_{NB}^{\ c-1} \qquad (19)$$

Until convergence of the vector of regression coefficients is reached in equation (19) with starting value $\beta_m^c \neq 0$ and N[th] number of path stages.

4. ANALYSIS

A study case of colorectal cancer was identified from the Nnamdi Azikwe teaching hospital, Anambra state, Nigeria from 2014 to 2017. Colorectal cancer is a pelvic swollen turmoil in male patients .The data comprises of six (6) covariates $d = 6$ as against the sample size of five (5) $n = 5$. The covariates are influences or factors that determined whether the the swollen pelvic will be subjected to operation or uses of drugs in order to stop/suppress the growing turmoil. The count covariates are age in years, status = level of progression, nodes=size of the turmoil (small or large cell cancer), e-type = type of the cells, rx = level of completion of cancer therapy, and obstruct = chronic or acute level of the unwanted growth. From the stated date, only five cases were recorded.

Table 1: Poisson coefficients of model criteria, penalty, selected variables, residual deviance and penalized residual.

Measures	MCP	SCAD	GLM
BIC	14.4463	14.4463	20.5098
AIC	15.2274	15.2274	21.2270
L1 (Loss function)	-11.22741	-11.22741	--
Log-likelihood	-5.6137	-5.6137	-5.6137
Deviance	0.8345	0.8345	8.3457e-01
Residual deviance	1.7492	1.7492	2.5596e-21
SelectedVariable(s)	4	4	4
PV	0.0009	0.0014	--
PR	2.3798	2.4056	--
RP	0.02449 a=3	0.02439 b=3.7	--

Keys: PV= Penalized Value; PR= Penalized Residual; RP= Regularized Parameter

Table 2: Binomial coefficients of model criteria, penalty, selected variables, residual deviance and penalized residual.

Measures	MCP	SCAD	GLM
BIC	5.5705	4.5171	9.2030
AIC	6.3516	3.7360	10.000
L1(Loss function)	-2.3516	-0.5171	----
Log-likelihood	-7.1758	-7.2586	-7.0261
Deviance	0.3012	0.30112	6.7301e+00
Residual deviance	0.55167	0.317173	2.1434e-10
SelectedVariable(s)	4	4	4
PV	0.0009	0.00141	--
PR	0.28033	0.16563	--
RP	0.02449 a=3	0.02449 b=3.7	

Keys: PV= Penalized Value; PR= Penalized Residual; RP= Regularized Parameter

Table 3: Negative-Binomial coefficients of model criteria, penalty, selected variables, residual deviance and penalized residual.

Measures	MCP	SCAD	GLM
BIC	14.4675	14.4675	18.4901
AIC	15.2486	15.2486	19.0686
L_1 (Loss function)	-11.2486	-11.2486	--
Log-likelihood	-5.6137	-5.6243	-5.6243
Deviance	0.8395	0.8394	0.9686
Residual deviance	1.7492	1.7492	1.9631
Selected Variable(s)	4	4	4
PV	0.0009	0.00139	--
PR	2.3798	2.4056	--
RP	0.02439 a=3	0.02439 b=3.7	--
Dispersion	0.0252	0.0251	--

Keys: PV= Penalized Value; PR= Penalized Residual; RP= Regularized Parameter

From the three distributional tables, two covariates “status” and “nodes” were penalized while four covariates “extent”, “age”, “etype” and “rx” are the significant and contributing factors in the three different distribution responses considered (selected variables). It is obvious that Binomial non-convex penalized estimation has the minimum error model selection criteria of BICs of (5.5705, 4.5171) and AICs of (6.3516, 3.7360) for MCP and SCAD respectively as against a higher BICs and AICs of (14.4463, 14.4463) (14.4675, 14.4675) and (15.2274, 15.2274) (15.2486, 15.2486) for Poisson and Negative Binomial MCPs and SCADs respectively. In collaboration with the model selectio n criteria, the log-likelihood of the Binomial MCP and SCAD happened to be smallest of all the log-likelihoods of the distributions. This affirmed the assertion that the smaller the log-likelihood the ideal and robust the model. The penalized values for the MCPs and SCADs of the probability distributions coincide but their residuals vary. The penalized errors of Binomial function for MCP and SCAD are (0.28033 & 0.28033) respectively compared to Poisson and Negative Binomial MCPs and SCADs of same value at (2.3798 & 2.4056). It is to be noted that the GLM approach indirectly penalized covariates by ignoring irrelevant covariates without necessary provision for penalization.

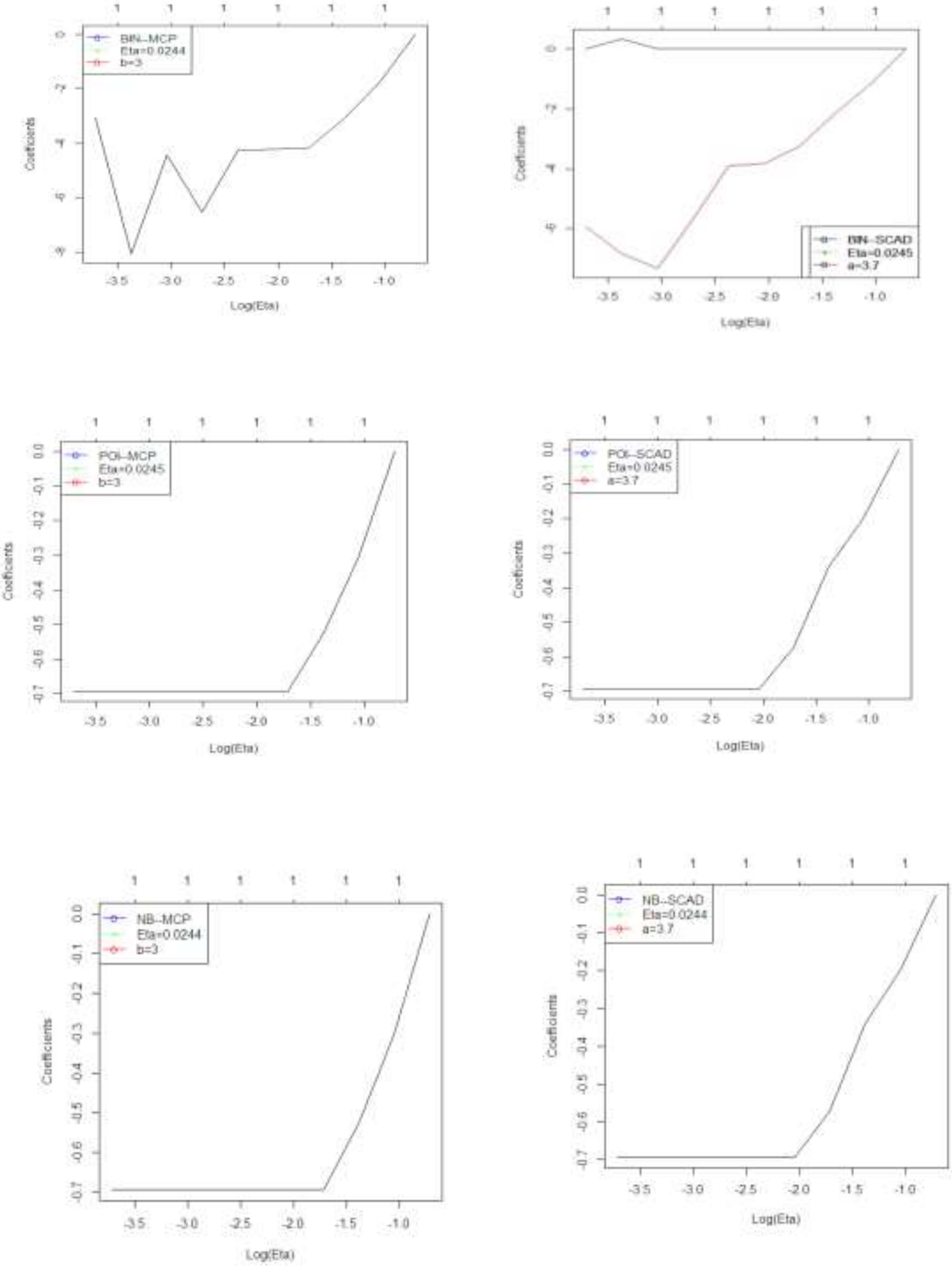

Figure 1: Coefficient Plot of the Coefficient Paths for the Fitted MCPs and SCADs.

Figure 1. shows the path fitted of the regression coefficients by each of the MCPs and SCADs of the considered probability mass distributions. Having said that each of the non-convex penalized estimation selected four covariates, Binomial non-convex penalizations of estimates of βs showed a more non-homogeneity in nature for conformity as against a parallel estimates by Poisson and Negative Binomial non-convex penalizations. Figure 1 buttresses the aberrant of the actual responses in the MCPs, SCADs of Poisson, and Negative Binomial non-convex penalization to their predicted responses. Moreover, figure 1. elucidates on the high magnitude of values suppressed by Poisson and Negative Binomial distributions in the process of penalizing based on the considered case study. In addition, Negative Binomial

distribution estimated a smaller value for the frequency distribution scattering from an average, that is, the Negative-Binomial responses revealed an over-dispersion of 2.5% by the covariates, which is of negligible consequence as to affecting the estimated parameters. Furthermore, The GLM approach was out-performed by the measurement indexes of the non-convex of the penalizations.

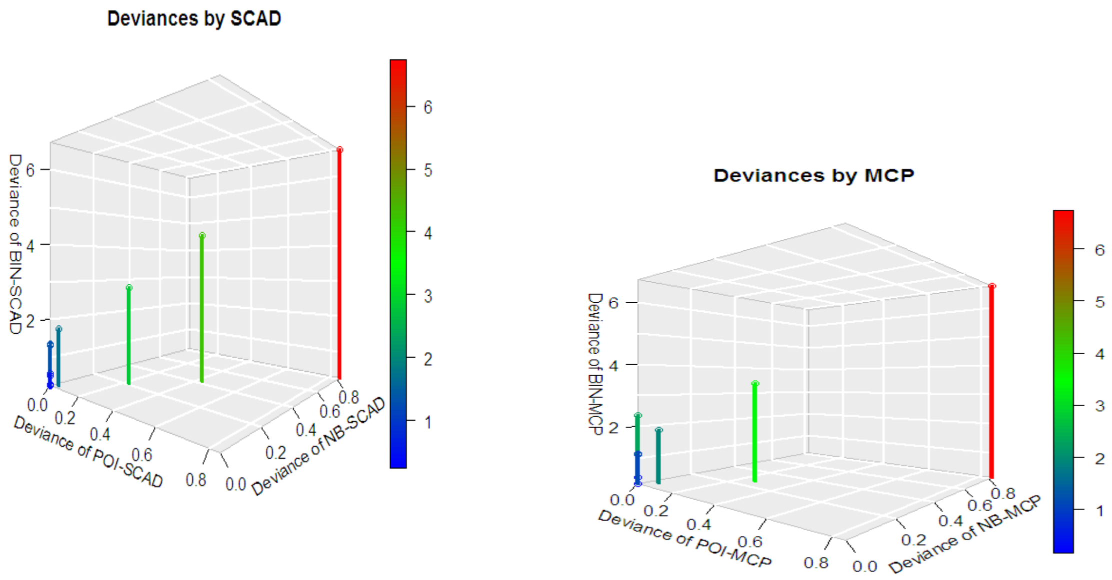

Figure 2: Deviance plot of the MCPs and SCADs via Binomial, Poisson Negative Binomial distributions.

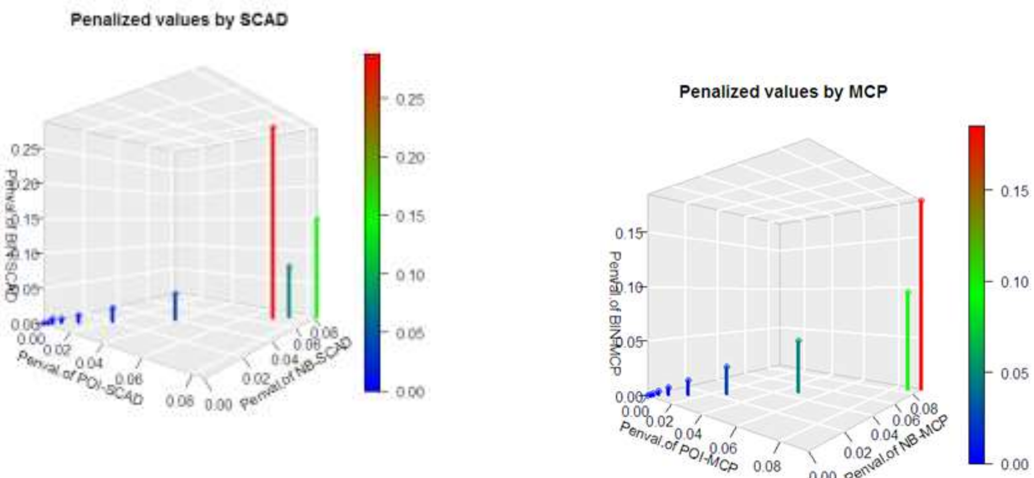

Figure 3: Penalized values by each of the MCPs and SCADs in Binomial, Poisson Negative Binomial distributions.

5. CONCLUSION

Given the previous analysis, it is safe to state that the Binomial related responses of non-convex penalization via MCP and SCAD was pre-eminent in the regulation of factors contributing to a surgical or drug use approach in curbing colorectal cancer via penalization of covariates. Apart from the Poisson, Binomial and Negative Binomial responses non-convex penalization considered, it should be noted that responses with different PMFs or PDFs could also be constrained to the non-convex penalization via GLM provided they belong to the exponential family. However, the Ordinary Least Square (OLS) could also be considered as an alternative approach for the GLM when circumventing the problem of number of covariates exceeding the sample size via MCP and SCAD.

6. ACKNOWLEDGMENT

We would like to acknowledge and thank the core management of Nnamdi Azikwe teaching hospital, Anambra state, Nigeria for providing datasets on of colorectal cancer for this research.

7. REFERENCES

[1] Agarwal, A., Negahban, S. and Wainwright, M. J, "Fast global convergence of gradient methods for high-dimensional statistical recovery" Annals of Statistics, vol. 40, no. 5, pp 2452–2482, 2012. doi:10.1214/12-AOS1032.
[2] Chen, J. and Chen, Z., "Extended Bayesian information criteria for model selection with large model spaces" Biometrika, vol. 95, no. 3, pp. 759–771, 2008.doi:10.1093/biomet/asn034.
[3] Fan, J. and Li, R., "Variable selection via non-concave penalized likelihood and its oracle properties" Journal of the American Statistical Association, vol. 96, no. 456, pp. 1348–1360, 2001. doi: 10.1198/016214501753882273.
[4] Hirose, K and Yamamoto, M., "Penalized likelihood factor analysis via non-convex penalties" 5th International Conference of the ERCIM WG on Computing & Statistics (ERCIM 2012), Conference Centre, Oviedo, Spain, pp.1-33, December 2012.
[5] Kim, Y., Kwon, S. and Choi, H, "Consistent model selection criteria on high dimensions" Journal of Machine Learning Research, vol. 13, no. 1, pp. 1037-1057, 2012.
[6] Liu, H.Y. Yao, T. and Li, R, "Global solutions to folded concave penalized non-convex learning" The Annals of Statistics. Vol. 44, no. 2, pp. 629–659, 2016. doi:10.1214/15-AOS1380. © Institute of Mathematical Statistics.
[7] Loh, P-L. and Wainwright, M. J, "Regularized M-estimators with non-convexity: Statistical and algorithmic theory for local optima" Journal of Machine Learning Research, vol. 16, pp. 559–616, 2015.
[8] Loh, P-L, "Local optima of non-convex regularized M-estimators" Electrical Engineering and Computer Sciences University of California at Berkeley. Technical Report No. UCB/EECS-2013-54, 2013. http://www.eecs.berkeley.edu/Pubs/TechRpts/2013/EECS-2013-54.html
[9] Nesterov, Y, "Gradient methods for minimizing composite functions" Mathematical Programming, vol.140, no. 1, pp. 125–161, 2013. Doi:10.1007/s10107-012-0629-5.
[10] Wang, Z., Han, H. and Zhang, T, "Optimal computational and statistical rates of convergence for sparse non-convex learning problems" The Annals of Statistics, vol. 42, no. 6, pp. 2164–2201, 2014. doi: 10.1214/14-AOS1238.
[11] Wang, L., Kim, Y. and Li, R, "Calibrating non-convex penalized regression in ultra-high dimension" Annals of Statistics, vol. 41, no. 5, pp. 2505–2536, 2013. doi: 10.1214/13-AOS1159 © Institute of Mathematical Statistics.
[12] Xiao, L., Zhang, T., "A proximal-gradient homotopy method for the sparse least-squares problem" SIAM Journal on Optimization, vol. 23, no. 2, pp. 1062–1091, 2013.
[13] Yang, N. and Han, L., "A general theory of hypothesis tests and confidence regions for sparse high dimensional models" The Annals of Statistics, vol. 45, no. 1, pp. 158-195, 2017. doi:10.1214/16AOS1448.
[14] Yu, Y. and Feng, Y., "APPLE: approximate path for penalized likelihood estimators" Journal of Statistical Computing, vol. 24, pp. 803–819, 2014. doi: 10.1007/s11222-013-9403-7.
[15] Zhang, C.H., "Nearly unbiased variable selection under minimax concave penalty" Annals of Statistics, vol. 38, no. 2, pp. 894–942, 2010. doi: 10.1214/09-aos729.
[16] Zhang, C.-H and Zhang, T., "A general theory of concave regularization for high dimensional sparse estimation problems" Statistical Science, vol. 27, no. 4, pp. 576-593, 2012.
[17] Zhang, K., Yin, F. and Xiong, F., "Comparisons of penalized least squares methods by simulations" 1-16, 2014. arXiv:1405.1796v1 [stat. Co].
[18] Zhong, L.W. and Kwok, J.F., "Gradient descent with proximal average for non-convex and composite regularization, Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, pp. 1 – 7, 2014.