Bayesian Regression Analysis Using S-PLUS and R Softwares

[1]S.P. Ahmad, [2]A.A. Khan and [1]A. Ahmed
[1]Department of Statistics, University of Kashmir, Srinagar, Kashmir
[2]Division of Agricultural Statistics and Economics, SKUAST-Kashmir, Shalimar, Srinagar, Kashmir

Abstract: This study is devoted to the Bayesian regression analysis. It contains an extended version of normal theory regression models which includes extreme-value, logistic and normal regression models as its particular cases. Methods proposed in this study are illustrated numerically, i.e., regression coefficient of pH on EC (Electrical Conductivity) of soil data is analyzed numerically in S-PLUS and R softwares.

Key words: Bayesian regression, logistic model, S-PLUS, R softwares, variable, Kashmir

INTRODUCTION

Regression analysis is a statistical technique for modeling the relationship between variables. In practice, many situations involve heterogeneous population and it is important to consider the relationship of response variable y on concomitant variable x which is explicitly recognized. One way to examine the relationship of concomitant variable (or regressor variable) to response variable y is through a regression model in which y has a distribution that depends upon the regressor variables. This involves specifying a model for the distribution of y given x where:

$$x = (x_1, x_2, ..., x_p)$$

1×p vector of regressor variables for an individual. Let the distribution of y given x is of the form:

$$f(y \mid x, \beta, \sigma) = \frac{1}{\sigma} f\left(\frac{y - x\beta}{\sigma}\right) \qquad (1)$$

where, β is a p×1 vector of regression coefficients:

$$\beta = (\beta_1, \beta_2, ..., \beta_p)^T$$

And:

$$E[y \mid x, \beta] = x\beta$$

The alternative form of Eq. 1 is:

$$y = x\beta + \sigma z \qquad (2)$$

Where:

$$z = \frac{y - x\beta}{\sigma}$$

has the standardized distribution with density function f (z). The family of models for which f (z) has a standard normal distribution is common in statistical literature (Searle, 1971; Rao, 1973; Seber, 1977; Draper and Smith, 1981; Weisberg, 1985) but models in which z has other distribution belonging to location-scale family (Eq. 2) are also important.

For example, extreme value regression models are employed in applications ranging from accelerated life testing (Lawless, 2003; Zelen, 1959) to the analysis of survival data on patients suffering from chronical diseases (Prentice, 1973; Feigl and Zelen, 1965; Krall *et al.*, 1975).

Furthermore, if data is contaminated with outliers then normal distribution can be replaced with Student's t distribution (with small degree of freedom) to have a better fit (Lange *et al.*, 1989). Model (Eq. 2) has the ability to accommodate linear as well as non-linear models for the various functional forms of $x\beta$.

None of the earlier researchers discuss Bayesian approach. Box and Tiao (1973) and Gelman *et al.* (1995) discuss this approach of regression analysis to deal with normal linear as well as non-linear non-normal models. Zellner (1971) deals with Bayesian inference in reference to econometrics. It is mostly confined to normal linear models only. The general frame work used for casual inference is presented by Rubin (1974, 1978). Bayesian approaches to analyzing regression residuals appear in Zellner (1976), Chaloner and Brant (1988) and Chaloner (1991).

Bayesian joint inference for β and σ: Suppose that associated with each individual observation is a response y_i and a regression vector:

Corresponding Author: S.P. Ahmad, Department of Statistics, University of Kashmir, Srinagar, Kashmir

Table 1: A Summary of derivatives of log-likelihoods

Derivatives	Distributions		
	Normal	Extreme-value	Logistic
l_β	$\dfrac{1}{\sigma}\sum_{i=1}^n z_i x_i^T$	$-\dfrac{1}{\sigma}\sum_{i=1}^n (1-e^{z_i})x_i^T$	$\dfrac{1}{\sigma}\sum_{i=1}^n\left(\dfrac{e^{z_i}-1}{e^{z_i}+1}\right)x_i^T$
l_σ	$\dfrac{1}{\sigma}\sum_{i=1}^n z_i^2 - \dfrac{n}{\sigma}$	$-\dfrac{1}{\sigma}\sum_{i=1}^n z_i(1-e^{z_i})-\dfrac{n}{\sigma}$	$\dfrac{1}{\sigma}\sum_{i=1}^n z_i\left(\dfrac{e^{z_i}-1}{e^{z_i}+1}\right)-\dfrac{n}{\sigma}$
$l_{\beta\sigma}$	$-\dfrac{2}{\sigma^2}\sum_{i=1}^n z_i x_i^T$	$-\dfrac{1}{\sigma^2}\sum_{i=1}^n (z_i e^{z_i}+e^{z_i}-1)x_i^T$	$-\dfrac{1}{\sigma^2}\sum_{i=1}^n\left(\dfrac{e^{2z_i}+2z_i e^{z_i}-1}{(e^{z_i}+1)^2}\right)x_i^T$
$l_{\sigma\beta}$	$-\dfrac{2}{\sigma^2}\sum_{i=1}^n z_i x_i^T$	$-\dfrac{1}{\sigma^2}\sum_{i=1}^n (z_i e^{z_i}+e^{z_i}-1)x_i$	$-\dfrac{1}{\sigma^2}\sum_{i=1}^n\left(\dfrac{e^{2z_i}+2z_i e^{z_i}-1}{(e^{z_i}+1)^2}\right)x_i$
$l_{\beta\beta}$	$-\dfrac{2}{\sigma^2}\sum x_i^T x_i$	$-\dfrac{1}{\sigma^2}\sum_{i=1}^n e^{z_i}x_i^T x_i$	$-\dfrac{2}{\sigma^2}\sum_{i=1}^n\left(\dfrac{e^{z_i}x_i^T x_i}{(e^{z_i}+1)^2}\right)$
$l_{\sigma\sigma}$	$-\dfrac{3}{\sigma^2}\sum_{i=1}^n z_i^2 + \dfrac{n}{\sigma^2}$	$-\dfrac{1}{\sigma^2}\sum_{i=1}^n (z_i^2 e^{z_i}+2z_i e^{z_i}-2z_i)+\dfrac{n}{\sigma^2}$	$-\dfrac{2}{\sigma^2}\sum_{i=1}^n z_i\left(\dfrac{e^{2z_i}+z_i e^{z_i}-1}{(e^{z_i}+1)^2}\right)+\dfrac{n}{\sigma^2}$

Where, $z_i = y_i - x_i\beta/\sigma$, $i = 1, 2,...,n$

$$x_i = (x_{i1}, x_{i2}, ..., x_{ip})$$

Thus (x_i, y_i) for $i = 1, 2,..., n$ is assumed to be a random sample from location-scale family of models in (Eq. 1) which can be represented as:

$$f(y_i | x_i, \beta, \sigma)$$

Consequently, in terms of general notation:

$$\theta^T = (\beta, \sigma)^T$$

A vector of length $(p+1)$ and likelihood is (Table 1):

$$\prod_{i=1}^n f(y_i | x_i, \beta, \sigma)$$

This implies that:

$$l(\beta,\sigma)=\log\left(\prod_{i=1}^n f(y_i | x_i, \beta, \sigma)\right) \quad (3)$$
$$= \sum_{i=1}^n \log f(z_i) - n\log\sigma$$

Where:

$$z_i = \frac{y_i - x_i\beta}{\sigma}$$

It may be noted that same notation z_i is being used for standard variate both in univariate and regression scenario. Assuming prior independence of β and σ (Jeffreys, 1961; Box and Tiao, 1973), researchers have:

$$p(\beta, \sigma) \cong p(\beta)p(\sigma) \quad (4)$$

As a result, joint posterior density of β and σ given data vector: $y^T = (y_1, y_2, ..., y_n)$ is:

$$p(\beta,\sigma | x,y) \quad \alpha \quad \prod_{i=1}^n f(y_i | x_i, \beta, \sigma)p(\beta)p(\sigma) \quad (5)$$

Where:

$$x = (x_1, x_2, ..., x_n)^T$$

is a n×p matrix of covariates (or regressors) corresponding to response vector y. Now Joint inference for β and σ can be made from posterior (Eq. 5). Posterior mode $(\hat{\beta}, \hat{\sigma})^T$ of $p(\beta, \sigma | x, y)$ serves as a point estimate of β and σ. Its calculations requires partial derivatives of log posterior:

$$l^*(\beta, \sigma) = l(\beta, \sigma)+\log p(\beta)+\log p(\sigma) \quad (6)$$

Defining partial derivatives:

$$l_\beta^* = \frac{\partial l^*}{\partial\beta}, \text{ a vector of } (p\times1) \text{ partial derivatives}$$

$$l_\sigma^* = \frac{\partial l^*}{\partial\sigma}, \text{ a scalar}$$

$$l_{\beta\sigma}^* = \frac{\partial^2 l^*}{\partial\beta\partial\sigma}, \text{ a } (p\times1) \text{ vector}$$

$$1^*_{\sigma\beta} = \frac{\partial^2 1^*}{\partial\sigma\partial\beta}, \ \text{a } (1\times p) \text{ vector}$$

$$1^*_{\beta\beta} = \frac{\partial^2 1^*}{\partial\beta\partial\beta^T}, \ \text{a } (p\times p) \text{ matrix}$$

And:

$$1^*_{\sigma\sigma} = \frac{\partial^2 1^*}{\partial\sigma\partial\sigma^T}$$

These derivatives can be defined more explicitly as:

$$1^*_\beta = 1_\beta + \frac{p'(\beta)}{p(\beta)}$$

$$1^*_\sigma = 1_\sigma + \frac{p'(\sigma)}{p(\sigma)}$$

$$1^*_{\beta\sigma} = 1_{\beta\sigma}$$

$$1^*_{\sigma\beta} = 1_{\sigma\beta}$$

$$1^*_{\beta\beta} = 1_{\beta\beta} + \left[\frac{p'(\beta)}{p(\beta)}\right]'$$

And:

$$1^*_{\sigma\sigma} = 1_{\sigma\sigma} + \left[\frac{p'(\sigma)}{p(\sigma)}\right]'$$

Consequently, score vector U (β, σ) and Hessian matrix H (β, σ) are:

$$U(\beta,\sigma) = \begin{bmatrix} 1^*_\beta \\ 1^*_\sigma \end{bmatrix}, \ \text{a } (p+1) \text{ vector}$$

And:

$$H(\beta,\sigma) = \begin{bmatrix} 1^*_{\beta\beta} & 1^*_{\beta\sigma} \\ 1^*_{\sigma\beta} & 1^*_{\sigma\sigma} \end{bmatrix}, \ \text{a } (p+1)\times(p+1) \text{ matrix}$$

Therefore, making use of Newton-Raphson iteration scheme there can obtain posterior mode vector $(\hat\beta, \hat\sigma)^T$ as:

$$\begin{bmatrix} \hat\beta \\ \hat\sigma \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \sigma_0 \end{bmatrix} - H^{-1}(\beta_0,\sigma_0) \begin{bmatrix} 1^*_\beta \\ 1^*_\sigma \end{bmatrix} \tag{7}$$

The asymptotic posterior covariance matrix of (Eq. 5) can be obtained as:

$$I^{-1}(\hat\beta,\hat\sigma) = -H^{-1}(\hat\beta,\hat\sigma)$$

$$= \sum(\hat\beta,\hat\sigma)$$

More clearly, posterior density:

$$p(\beta,\sigma\,|\,x,y) \cong N_{p+1}\left((\hat\beta,\hat\sigma)^T, I^{-1}(\hat\beta,\hat\sigma)\right) \left(1 + O(n^{-\frac{1}{2}})\right) \tag{8}$$

where, N_r (a, b) is the r-variate normal distribution with mean vector a and a covariance matrix b. This is a first order approximation of the posterior density (Berger, 1985). An equivalent version of this approximation is the Chi-square approximation, i.e.:

$$W(\beta,\sigma) = -2[1^*(\beta,\sigma) - 1^*(\hat\beta,\hat\sigma)] \approx \chi^2_{p+1}$$

A more accurate approximation, Laplace's approximation (Tierney and Kadane, 1986; Reid, 1988) can be also used, i.e.:

$$p(\beta,\sigma\,|\,x,y) \cong (2\pi)^{-\frac{p+1}{2}} |I(\hat\beta,\hat\sigma)|^{\frac{1}{2}} \exp\left[-\frac{1}{2}W(\hat\beta,\hat\sigma)\right]\left(1 + O(n^{-1})\right) \tag{9}$$

Any of the approximations can be used both for hypothesis testing and construction of credible regions.

The marginal Iinference for β and σ: In regression analysis there are often interested with the regression coefficient vector β (Weisberg, 1985; Zellner, 1971) and σ is treated as nuisance parameter (Box and Tiao, 1973; Gianola and Fernando, 1986; Gelman *et al.*, 1995) (Table 2).

Therefore, nuisance parameter must be integrated out from the joint posterior density (Eq. 5) to obtain marginal posterior density of β, denoted by p ($\beta|$x, y).

However, if the interest is concerned with the scale parameter σ, its posterior p ($\sigma|$x, y) can also be obtained after integrating out the joint posterior with respect to β. To be precise, marginal densities for β and σ are:

Table 2: A Summary of prior densities for location parameter β

Name of density	$p(\mu)$	$\dfrac{p'(\mu)}{p(\mu)}$	$\left[\dfrac{p'(\mu)}{p(\mu)}\right]'$
Non-informative	Constant	Zero	Zero
Normal	$c\exp\left(-\dfrac{1}{2}D^{T}D\right)$	$-\dfrac{D}{\sigma_0}$	$-\dfrac{1}{\sigma_0^2}I$

$$p(\beta\,|\,x,y)=\int p(\beta,\sigma\,|\,x,y)\,d\sigma \qquad (10)$$

Similarly, marginal posterior of σ can be obtained, i.e:

$$p(\sigma\,|\,x,y)=\int p(\beta,\sigma\,|\,x,y)\,d\beta \qquad (11)$$

Bayesian analysis is to be based on these two posteriors. For normal model, $p(\beta|x, y)$ and $p(\sigma|x, y)$ can be obtained in closed form (Zellner, 1971). However, for non-normal members of location-scale family these marginals can be obtained through numerical integration only (Naylor and Smith, 1982). Alternative approach is to deal with asymptotic theory approach (Tierney *et al.*, 1989; Leonard *et al.*, 1989). Normal and Laplace's approximations can be written directly for posterior densities $p(\beta|x, y)$ and $p(\sigma|x, y)$ as under.

Normal approximation: Marginal posterior density of β can be approximated by normal distribution as (Table 3):

$$p(\beta\,|\,x,y)=N_p\,(\hat{\beta},I_{11}^{-1}) \qquad (12)$$

Where:
$\hat{\beta}$ = The posterior mode
I_{11}^{-1} = A p×p matrix

Defined as:

$$I^{-1}(\hat{\beta},\hat{\sigma})=\begin{bmatrix} I_{11}^{-1} & I_{12}^{-1} \\ I_{21}^{-1} & I_{22}^{-1} \end{bmatrix}$$

where, suffixes 1 and 2 to I stand for $\hat{\beta}$ and $\hat{\sigma}$, respectively. This approximation is equivalent to the Chi square approximation defined as:

$$(\beta-\hat{\beta})^{T}I_{11}(\beta-\hat{\beta})\approx\chi_p^2$$

Corresponding approximations for $p(\sigma|x, y)$ can be written as:

$$p(\sigma\,|\,x,y)=N_1\,(\hat{\sigma},I_{22}^{-1}) \qquad (13)$$

This is equivalent to the Chi square approximation, i.e:

$$(\sigma-\hat{\sigma})^{T}I_{22}(\sigma-\hat{\sigma})\approx x_1^2$$

Table 3: Regression coefficient of pH on E. C for various models

Regression model	$\hat{\beta}$ (Intercept) $\hat{\beta}_0$	$\hat{\beta}_1$	Posterior Std. $\hat{\beta}_0$	$\hat{\beta}_1$
Extreme-value	6.71	2.32	0.0447	0.2881
Logistic	6.29	2.41	0.0365	0.2332
Normal	6.33	2.00	0.0335	0.1890

Laplace's approximation: Laplace's approximation can also be used to approximate marginal density of β, i.e:

$$p(\beta\,|\,x,y)\cong\left[\frac{|I(\hat{\beta},\hat{\sigma})|}{2\pi\,|I(\beta,\hat{\sigma}(\beta))|}\right]^{\frac{1}{2}} \qquad (14)$$
$$\exp[l^{*}(\hat{\beta},\hat{\sigma}(\beta))-l^{*}(\hat{\beta},\hat{\sigma})]$$

where, $\hat{\sigma}(\beta)$ is the posterior mode of σ for a fixed β. Corresponding approximation for $p(\sigma|x, y)$ can also be written as:

$$p(\sigma\,|\,x,y)\cong(2\pi)^{-\frac{p}{2}}\left[\frac{|I(\hat{\beta},\hat{\sigma})|}{|I(\hat{\beta}(\sigma),\sigma)|}\right]^{\frac{1}{2}} \qquad (15)$$
$$\exp[l^{*}(\hat{\beta}(\sigma),\sigma)-l^{*}(\hat{\beta},\hat{\sigma})]$$

BAYESIAN REGRESSION ANALYSIS OF LOGISTIC MODEL

Let y be the response vector and x_i be the vector for the ith observation. Assume that:

$$z_i=\frac{y_i-x_i^{T}\beta}{\sigma}\sim f \qquad (16)$$

For some f (logistic distribution). Consequently, in terms of general notation:

$$\theta=(\beta,\sigma)^{T}$$

A vector of length $(\beta+1)$ and likelihood is given by:

$$\prod_{i=1}^{n}f(y_i\,|\,x_i,\beta,\sigma)$$

This implies that:

$$l(\beta,\sigma)=\log\prod_{i=1}^{n}f(y_i\,|\,x_i,\beta,\sigma) \qquad (17)$$
$$=\sum_{i=1}^{n}\log f(z_i)-n\log\sigma$$

where, z_i is defined in (Eq. 16). Researchers take partial derivatives with respect to μ and σ:

$$l_\beta = \frac{\partial l}{\partial \beta}$$
$$= \frac{1}{\sigma}\sum_{i=1}^{n}\left(\frac{e^{z_i}-1}{e^{z_i}+1}\right)x_i^T$$

$$l_\sigma = \frac{\partial l}{\partial \sigma}$$
$$= \frac{1}{\sigma}\sum_{i=1}^{n}z_i\left(\frac{e^{z_i}-1}{e^{z_i}+1}\right)-\frac{n}{\sigma}$$

$$l_{\beta\sigma} = \frac{\partial^2 l}{\partial \beta \partial \sigma}$$
$$= -\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(\frac{e^{2z_i}+2z_i e^{z_i}-1}{(e^{z_i}+1)^2}\right)x_i^T$$

$$l_{\sigma\beta} = \frac{\partial^2 l}{\partial \sigma \partial \beta}$$
$$= -\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(\frac{e^{2z_i}+2z_i e^{z_i}-1}{(e^{z_i}+1)^2}\right)x_i^T$$

$$l_{\beta\beta} = \frac{\partial^2 l}{\partial \beta \partial \beta^T}$$
$$= -\frac{2}{\sigma^2}\sum_{i=1}^{n}\frac{e^{z_i}x_i x_i^T}{(e^{z_i}+1)^2}$$

$$l_{\sigma\sigma} = \frac{\partial^2 l}{\partial \sigma \partial \sigma^T}$$
$$= -\frac{2}{\sigma^2}\sum_{i=1}^{n}z_i\left(\frac{e^{2z_i}+z_i e^{z_i}-1}{(e^{z_i}+1)^2}\right)+\frac{n}{\sigma^2}$$

Researchers follow the standard approach of Box and Tiao (1973) and Gelman *et al.* (1995) assuming the prior:

$$p(\beta,\sigma)\cong p(\beta)p(\sigma) \qquad (18)$$

where, $p(\beta)$ and $p(\sigma)$ are priors for β and σ. By using Bayes theorem there obtain the posterior density $p(\beta,\sigma|y)$ as:

$$p(\beta,\sigma|x,y) \quad \alpha \quad \prod_{i=1}^{n}f(y_i|x_i,\beta,\sigma)p(\beta,\sigma) \qquad (19)$$

The log-posterior is given by:

$$\log p(\beta,\sigma|x,y)=\log\prod_{i=1}^{n}p(y_i|x_i,\beta,\sigma)+$$
$$\log p(\beta)+\log p(\sigma)$$

Or:

$$l^*(\beta,\sigma)=l(\beta,\sigma)+\log p(\beta)+\log p(\sigma) \qquad (20)$$

For a prior:
$$p(\beta,\sigma)\cong p(\beta)p(\sigma)=1$$

Researchers have:
$$1^*_\beta = l_\beta$$

$$1^*_\sigma = l_\sigma$$

$$1^*_{\beta\sigma} = l_{\beta\sigma}$$

$$1^*_{\sigma\beta} = l_{\sigma\beta}$$

$$1^*_{\beta\beta} = l_{\beta\beta}$$

And:
$$1^*_{\sigma\sigma} = l_{\sigma\sigma}$$

The posterior mode is obtained by maximizing (Eq. 20) with respect to β and σ. The score vector of log posterior is given by:

$$U(\beta,\sigma) = (1^*_\beta, 1^*_\sigma)^T$$

And Hessian matrix of log posterior is:

$$H(\beta,\sigma)=\begin{bmatrix} 1^*_{\beta\beta} & 1^*_{\beta\sigma} \\ 1^*_{\sigma\beta} & 1^*_{\sigma\sigma} \end{bmatrix}$$

Posterior mode $(\hat{\beta},\hat{\sigma})$ can be obtained from iteration scheme:

$$\begin{bmatrix} \hat{\beta} \\ \hat{\sigma} \end{bmatrix}=\begin{bmatrix} \beta_0 \\ \sigma_0 \end{bmatrix}-H^{-1}(\beta_0,\sigma_0)\begin{bmatrix} 1^*_\beta \\ 1^*_\sigma \end{bmatrix} \qquad (21)$$

Consequently, modal variance Σ can be obtained as:

$$I^{-1}(\hat{\beta},\hat{\sigma}) = -H^{-1}(\hat{\beta},\hat{\sigma})$$

Using normal approximation there can write directly a bivariate normal approximation of $p(\beta,\sigma|x,y)$ as:

$$p(\beta,\sigma\,|\,y)\cong N_{p+1}\left(\begin{array}{c}(\hat{\beta},\hat{\sigma})^T, I^{-1}(\hat{\beta},\hat{\sigma})\\[2mm]\left(1+O(n^{-\frac{1}{2}})\right)\end{array}\right) \qquad (22)$$

Similarly there can write Bayesian analog of likelihood ratio criterion as:

$$W(\beta,\sigma)=-2[l^*(\mu,\sigma)-l^*(\hat{\mu},\hat{\sigma})]\approx\chi^2_{p+1} \qquad (23)$$

Using Laplace's approximation p (β, $\sigma\,|$x, y) can write as:

$$p(\beta,\sigma\,|\,x,y)\cong(2\pi)^{-\frac{p-1}{2}}\,|\,I(\hat{\beta},\hat{\sigma})\,|^{\frac{1}{2}}$$
$$\exp[l^*(\mu,\sigma)-l^*(\hat{\mu},\hat{\sigma})]\left(1+O(n^{-1})\right) \qquad (24)$$

The marginal Bayesian inference about β and σ is to be based on marginal posterior densities of these parameters. Marginal posterior for β can be obtained after integrating out p (β, $\sigma\,|$x, y) with respect to σ, i.e:

$$p(\beta\,|\,x,y)=\int p(\beta,\sigma\,|\,x,y)d\sigma \qquad (25)$$

Similarly, marginal posterior of σ can be obtained, i.e:

$$p(\sigma\,|\,x,y)=\int p(\beta,\sigma\,|\,x,y)d\beta \qquad (26)$$

The Normal and Laplace's approximation's of p ($\beta\,|$x, y) and p ($\sigma\,|$x, y) are given as. There can write normal approximation of marginal posterior p ($\beta\,|$x, y) as:

$$p(\beta\,|\,x,y)=N_p(\hat{\beta},I_{11}^{-1}) \qquad (27)$$

where, $\hat{\beta}$ is the posterior mode and I_{11}^{-1} is a (p×p) matrix defined as:

$$I^{-1}(\hat{\beta},\hat{\sigma})=\begin{bmatrix}I_{11}^{-1} & I_{12}^{-1}\\ I_{21}^{-1} & I_{22}^{-1}\end{bmatrix}$$

Bayesian analog of likelihood ratio criterion can also be defined as a test criterion as:

$$(\beta-\hat{\beta})^T I_{11}(\beta-\hat{\beta})\approx\chi^2_p \qquad (28)$$

Laplace's approximation of marginal posterior density p ($\beta\,|$x, y) can be given by:

Table 4: Approximate Normal posterior quantiles for regression coefficient of various models

Models	Posterior	Posterior quantile					
		0.025	0.25	0.50	0.75	0.95	0.975
Extreme-value	Normal	6.63	6.68	6.72	6.75	6.79	6.80
Logistic	Normal	6.16	6.20	6.23	6.25	6.29	6.30
Normal	Normal	6.26	6.31	6.33	6.35	6.39	6.40

$$p(\beta\,|\,x,y)\cong\left[\frac{|\,I(\hat{\beta},\hat{\sigma})\,|}{2\pi\,|\,I(\hat{\beta},\hat{\sigma}(\beta))\,|}\right]^{\frac{1}{2}} \qquad (29)$$
$$\exp[l^*(\hat{\beta},\hat{\sigma}(\beta))-l^*(\hat{\beta},\hat{\sigma})]$$

Similarly, p ($\sigma\,|$x, y) can be approximated and results corresponding to normal and Laplace's approximation can be written as:

$$p(\sigma\,|\,x,y)=N_1(\hat{\sigma},I_{22}^{-1}) \qquad (30)$$

Or equivalently:

$$(\sigma-\hat{\sigma})^T I_{22}(\sigma-\hat{\sigma})\approx\chi^2_1 \qquad (31)$$

$$p(\sigma\,|\,x,y)\cong(2\pi)^{-\frac{p}{2}}\left[\frac{|\,I(\hat{\beta},\hat{\sigma})\,|}{|\,I(\hat{\beta}(\sigma),\sigma)\,|}\right]^{\frac{1}{2}} \qquad (32)$$
$$\exp[l^*(\hat{\beta}(\sigma),\sigma)-l^*(\hat{\beta},\hat{\sigma})]$$

Numerical illustrations: Numerical illustrations are implemented in S-PLUS software for Bayesian regression analysis. These illustrations are meant for the purpose of showing strength of Bayesian methods in practical situations. There have used survReg and censorReg functions for Bayesian analysis of various regression models with non-informative prior. S-PLUS has a function censorReg for regression analysis. This has a very substantial overlap with survReg but is more general in that it allows truncation as well as censoring (Venables and Ripley, 2002) (Table 4). The usage of survReg and censorReg are given below:

survReg (formula, data, dist)
censorReg (formula, data, dist)

Where:

formula = A formula expression as for other regression models

data = Optional data frame in which to interpret the variable occurring in the formula

dist = Assumed distribution for y variable

$D^T = (D_1, D_2, ..., D_p)$ a (p×1) vector:

$$D_i = \frac{\beta_i - \beta_{0i}}{\sigma_{0i}}$$

for i = 1, 2, ..., p; I stands for identity (p×p) matrix and c is the normalizing constant.

REFERENCES

Berger, J.O., 1985. Statistical Decision Theory and Bayesian Analysis. 2nd Edn., Springer-Verlag, New York.

Box, G.E.P. and G. Tiao, 1973. Bayesian Inference in Statistical Analysis. Adison and Wesley, UK., pp: 608.

Chaloner, K. and R. Brant, 1988. A bayesian approach to outlier detection and residual analysis. Biometrika, 75: 651-659.

Chaloner, K., 1991. Bayesian Residual Analysis in the Presence of Censoring. Biometrika, 78: 637-644.

Draper, N.R. and H. Smith, 1981. Applied Regression Analysis. John Wiley and Sons Inc., New York.

Feigl, P. and M. Zelen, 1965. Estimation of exponential survival probabilities with concomitant information. Biometrics, 21: 826-838.

Gelman, A., J.B. Carlin, H.S. Stern and D.B. Rubin, 1995. Bayesian Data Analysis. 1st Edn., Chapman and Hall, London.

Gianola, D. and R.L. Fernando, 1986. Bayesian methods in animal breeding theory. J. Anim. Sci., 63: 217-244.

Jeffreys, H., 1961. Theory of Probability. 3rd Edn., Oxford University Press, London, pp: 470.

Krall, J.M., V.A. Uthoff and J.B. Harley, 1975. A step-up procedure for selecting variables associated with survival. Biometrics, 31: 49-57.

Lange, K.L., R.J.A. Little and J.M.G. Taylor, 1989. Robust statistical modeling using the t distribution. J. Am. Stat. Assoc., 84: 881-896.

Lawless, J.F., 2003. Statistical Models and Methods for Lifetime Data. Wiley-Interscience, UK.

Leonard, T., J.S.J. Hsu and K.W. Tsui, 1989. Bayesian marginal inference. J. Am. Stat. Assoc., 84: 1051-1058.

Naylor, J.C. and A.F.M. Smith, 1982. Applications of a method for efficient computation of posterior distributions. Applied Stat., 31: 214-225.

Prentice, R.L., 1973. Exponential survivals with censoring and explanatory variables. Biometrika, 60: 279-288.

Rao, C.R., 1973. Linear Statistical Inference and its Applications. 2nd Edn., John Wiley and Sons, New York, ISBN: 0-471-21875-8.

Reid, N., 1988. Saddlepoint methods and statistical inference. Stat. Sci., 3: 213-227.

Rubin, D.B., 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. J. Educ. Psychol., 66: 688-701.

Rubin, D.B., 1978. Bayesian inference for causal effects: The role of randomization. Ann. Stat., 6: 34-58.

Searle, S.R., 1971. Linear Models. John Wiley and Sons, Inc., New York.

Seber, G.A.F., 1977. Linear Regression Analysis. John Wiley and Sons, New York.

Tierney, L. and J.B. Kadane, 1986. Accurate approximations for posterior moments and marginal densities. J. Am. Stat. Assoc., 81: 82-86.

Tierney, L., R.E. Kass and J.B. Kadane, 1989. Fully exponential laplace approximations to expectations and variances of nonpositive functions. J. Am. Stat. Assoc., 84: 710-716.

Venables, W.N. and B.D. Ripley, 2002. S Programming. Springer-Verlag, New York.

Weisberg, S., 1985. Applied Linear Regression. 2nd Edn., John Wiley and Sons, New York.

Zelen, M., 1959. Factorial experiments in life testing. Technometrics, 1: 269-288.

Zellner, A., 1971. An Introduction to Bayesian Inference in Econometrics. Wiley-Interscience, New York.

Zellner, A., 1976. Bayesian and non-bayesian analysis of the regression model with multivariate student-t error terms. J. Am. Stat. Assoc., 71: 400-405.