

PROPRIETY OF POSTERIOR DISTRIBUTIONS ARISING IN CATEGORICAL AND SURVIVAL MODELS UNDER GENERALIZED EXTREME VALUE DISTRIBUTION

Vivekananda Roy and Dipak K. Dey

Iowa State University and University of Connecticut

Abstract: This paper introduces a flexible skewed link function for modeling binary as well as ordinal data with covariates based on the generalized extreme value (GEV) distribution. Extreme value techniques have been widely used in many disciplines relating to risk analysis, but, applications to binary and ordinal data in a Bayesian context are sparse. There are a number of non-regular situations with the likelihood method for GEV models in which the usual asymptotic properties of MLE do not hold, suggesting Bayesian methodology for analyzing GEV models. We introduce the GEV distribution in reliability and survival models, and show that our proposed model leads to an extremely flexible hazard function. We investigate the properties of posterior distributions for binary and ordinal response models under the generalized extreme value link using a uniform prior distribution on the regression parameters. Necessary and sufficient conditions for the propriety of the posterior distribution are established. We consider similar issues for survival data models, where log survival time has a GEV distribution, and the propriety of the posterior distribution under a uniform prior on the regression coefficients is established. The flexibility of the proposed survival model is illustrated through a dataset involving a lung cancer clinical trial.

Key words and phrases: Complementary log-log link, generalized extreme value distribution, hazard function, improper prior, posterior propriety, skewness.

1. Introduction

The generalized extreme value (GEV) distribution is a family of continuous probability distributions that combines the Gumbel, Fréchet, and Weibull distributions that can be obtained as the limiting distributions of properly normalized maxima of n independent and identically distributed random variables. Extreme value analysis finds wide applications in many areas including climatology (Sang and Gelfand (2009)), environmental science (Smith (1989); Sang and Gelfand (2010); Wang, Dey, and Banerjee (2010)), financial strategy of risk management (Dahan and Mendelson (2001)) and survival analysis (Mann, Schafer, and Singpurwalla (1974); Kim and Ibrahim (2000)). In this article we show the broad applicability of the GEV distribution for analyzing binary, ordinal, and survival data.

The most popular model for binary response data is the logistic regression model based on the logit link function. Other frequently used link functions are the probit and complimentary log-log. These link functions do not always provide the best fit for a given data set. In particular, if the probability of a given binary response approaches 0 at a different rate than it approaches 1, the use of a symmetric link function such as probit or logit is inappropriate. In this case, if the link function is misspecified, there can be substantial bias in the mean response estimates (Czado and Santner (1992)). One intuitive way of guarding against link misspecification is to embed symmetric links into a wide parametric class of links. Several authors have introduced such parametric classes for binary response data. For example, Aranda-Ordaz (1981), Guerreo and Johnson (1982), Morgan (1983), and Whitemore (1983) considered different one-parameter families. Stukel (1988) extended these links by proposing a class of generalized logistic models. Stukel's (1988) models are general and such link functions as the probit and complimentary log-log can be approximated by members of this family. However, in the presence of covariates, Stukel's (1988) models yield improper posterior distributions for many types of non-informative improper priors, including the improper uniform prior for the regression coefficients (Chen, Dey, and Shao (1999)). Chen, Dey, and Shao (1999) introduced a class of skewed links that leads to proper posterior distributions for the regression coefficients under a standard improper prior. However, Chen, Dey, and Shao's (1999) model has the limitation that the intercept term is confounded with the skewness parameter. This problem was overcome in Kim, Chen, and Dey (2008) by a class of generalized skewed t-link models, though the constraint on the shape parameter δ with $0 < \delta \leq 1$ greatly reduces the possible range of skewness provided by this model.

To overcome this Wang and Dey (2010) introduced the GEV distribution as a link function. With a free shape parameter, the GEV distribution provides great flexibility in fitting a wide range of skewness in the response curve. Wang and Dey (2010) illustrated the flexibility of GEV link function with simulations and data.

The misspecification of link function can also occur for ordinal data (Wang and Dey (2011)). Many link functions for ordinal response data proposed in the literature including the probit link, Albert and Chib's (1993) family of t-links, and Chen and Dey's (2000) scale mixture of multivariate normal link functions are symmetric and may not be appropriate. Wang and Dey (2011) employed the GEV distribution for modeling ordinal response data. However, the authors did not address the issue of the propriety of the posterior distribution of the regression coefficients and of the cut points under improper uniform priors on the parameters. Here we give rigorous proofs for the propriety of the posterior distributions of the associated parameters for binomial as well as ordinal data.

We further propose survival models based on the GEV distribution, and provide sufficient conditions for the propriety of the corresponding posterior distributions when an improper uniform prior is used on the regression coefficients.

There is a close connection between categorical and survival data through the link function specification (Banerjee et al. (2007)). Banerjee et al. (2007) proposed a general class of non-proportional hazard models known as generalized odds-rate class of regression models. In a similar spirit, we develop a class of non-proportional hazard regression models using the GEV distribution. In reliability and survival analysis, the probability distribution of the time-to-failure of an equipment can be characterized by the hazard function (also known as failure rate) $\lambda(t) = f(t)/S(t)$, where $f(t)$ and $S(t)$ are failure density and survival function, respectively. Many widely used models including gamma, Weibull, and the truncated normal distribution lead to monotone hazard function. However, it has long been known that in many situations the hazard function is not monotone, it is either *upside-down shaped*, or *bathtub shaped*, or a combination of them (Lieberman (1969); Langlands et al. (1979); Bennett (1983)). A popular way of introducing non-monotone hazard function is by considering mixture distribution models (Barlow and Proschan (1975); Finkelstein (2009)). Mixtures do not always lead to a non-monotone hazard function, but mixtures of increasing failure rate can decrease, at least in some time intervals (Gurland and Sethuraman (1995)). Still, mixture modeling might not be desirable since it brings flexibility at the expense of additional parameters, consequently more parameters have to be estimated.

For a flexible hazard function, we propose the GEV distribution for $\log T$, where T denotes failure time. We show that by changing the shape parameter of the GEV distribution, we obtain a variety of shapes for the hazard function including the upside-down and bathtub shapes. The GEV distribution includes the Gumbel distribution as a special case, and if T has a Weibull distribution, then $\log T$ has a Gumbel distribution for the minimum extremes (Mann, Schafer, and Singpurwalla (1974); Kim and Ibrahim (2000)). Here the hazard (failure) rate is some power function of t , the time-to-failure, and is decreasing (increasing) if the shape parameter of the Weibull distribution is $< 1 (> 1)$ (Mann, Schafer, and Singpurwalla (1974)). However, if $\log T$ has a GEV distribution the modeling framework is much different.

We consider situations in which the distribution of failure time T depends on one or more covariates, in particular, *accelerated failure time* models that are linear models for $\log T$. Life data analysis involves analyzing times-to-failure data in order to quantify the reliability of a product. But, for products with long life time, only a few items fail during testing under normal operating conditions. The standard method is then to test under extreme operating conditions, referred

to as *accelerated life testing* (Mann, Schafer, and Singpurwalla (1974); Nelson (1990)). Accelerated failure time or log-location-scale models are also useful in other fields of applications. We introduce accelerated failure time models with GEV as error distribution. We consider a Bayesian analysis of the corresponding model under non-informative priors. Since the Jeffreys prior turns out to be extremely cumbersome in this case, we consider a uniform prior on the regression coefficients. We obtain sufficient conditions for the propriety of the corresponding posterior distribution. We demonstrate the flexibility of the proposed survival model through a lung cancer dataset.

The rest of the paper is organized as follows. Section 2 provides a short introduction to GEV distributions. Section 3 describes the GEV link models for binomial response data and provides necessary and sufficient conditions for propriety of the posterior distributions. Section 4 is devoted to the development of sufficient conditions for posterior propriety under GEV links for ordinal data. Section 5 introduces GEV distribution in reliability and accelerated failure time models. The paper concludes with a discussion in Section 6. The proofs of the theorems have been relegated to appendices.

2. Generalized Extreme Value Distribution

Suppose Y_1, Y_2, \dots is a sequence of iid random variables and let $M_n = \max\{Y_1, \dots, Y_n\}$. Extreme value theory considers the existence of $\lim_{n \rightarrow \infty} P[(M_n - b_n)/a_n \leq y] \equiv F(y)$ for two sequences of real numbers $a_n > 0$ and b_n . If $F(y)$ is a non-degenerate distribution, then it belongs to either the Gumbel, Fréchet, or the Weibull family of distributions, these can all be found in the family of GEV distributions with cumulative distribution function

$$G_{(\mu, \sigma, \xi)}(x) = \begin{cases} \exp \left[- \left\{ 1 + \xi \frac{x - \mu}{\sigma} \right\}_+^{-1/\xi} \right] & \text{if } \xi > 0 \text{ or } \xi < 0, \\ \exp \left(- \exp \left(- \frac{x - \mu}{\sigma} \right) \right) & \text{if } \xi = 0, \end{cases} \quad (2.1)$$

where $\mu \in \mathbb{R}$ is the location parameter, $\sigma \in \mathbb{R}^+$ is a scale parameter, $\xi \in \mathbb{R}$ is the shape parameter and $x_+ = \max(x, 0)$. The Gumbel, Fréchet, and the Weibull distributions are obtained from (2.1) by taking $\xi = 0$, $\xi > 0$, and $\xi < 0$, respectively. Detailed discussion of extreme value distributions can be found in Coles (2001) and Smith (1985).

The importance of GEV distribution as a link function arises from the fact that the shape parameter ξ controls the tail behavior of the distribution (Wang and Dey (2010)). The Gumbel distribution is the least positively skewed distribution in the GEV class when ξ is non-negative. Wang and Dey (2010) provide a plot of the probability distributions of the GEV family that demonstrates the flexibility of the GEV distribution.

Since the usual definition of skewness $\mu_3 = \{E(X - \mu)^3\}\{E(X - \mu)^2\}^{-3/2}$ does not work for large positive values of ξ 's for the GEV model, Wang and Dey (2010) extended Arnold and Groeneveld's (1995) skewness measure to the GEV distribution in terms of its mode. Wang and Dey (2010) showed that, based on this skewness definition, the GEV distribution is negatively skewed for $\xi < \log 2 - 1$ and positively skewed for $\xi > \log 2 - 1$.

3. Generalized Extreme Value Link for Binomial Regression Models

Suppose $\mathbf{y} = (y_1, \dots, y_n)$ is a vector of n independent binomial random variables. Also, let x_i be the $k \times 1$ vector of covariates associated with y_i , and suppose X denotes the $n \times k$ design matrix with rows x'_i . Let β be the $k \times 1$ vector of regression coefficients. Assume that $y_i \sim \text{Bin}(n_i, p_i) \quad i = 1, \dots, n$, and that

$$p_i = 1 - G_\xi(-x'_i\beta) ,$$

where $G_\xi(x)$ is the cumulative probability at x for the GEV distribution with $\mu = 0, \sigma = 1$, and an unknown shape parameter ξ . The joint pmf of \mathbf{y} is then

$$f(\mathbf{y}|\beta, \xi) = \prod_{i=1}^n \binom{n_i}{y_i} \left(1 - G_\xi(-x'_i\beta)\right)^{y_i} \left(G_\xi(-x'_i\beta)\right)^{n_i - y_i} .$$

It is possible to estimate the shape parameter ξ here by the maximum likelihood. However, the usual asymptotic properties of the maximum likelihood estimator may not hold. Smith (1985) studied maximum likelihood estimation for the three-parameter GEV distribution and found that when $\xi < -0.5$ the standard asymptotic likelihood results do not follow. Since it does not depend on such regularity assumptions Bayesian inference provides a viable alternative for analyzing the GEV link model.

In the next two subsections we consider the uniform and the Jeffreys priors on (β, ξ) and study the property of the corresponding posterior distributions.

3.1. Uniform prior

We consider an improper uniform prior on β , $\pi(\beta) \propto 1, \beta \in \mathbb{R}^k$, and a proper prior on ξ , $\pi(\xi) = 0.5I_{[-1,1]}(\xi)$. The joint posterior density is

$$\pi(\beta, \xi|\mathbf{y}) \propto \prod_{i=1}^n \binom{n_i}{y_i} \left(1 - G_\xi(-x'_i\beta)\right)^{y_i} \left(G_\xi(-x'_i\beta)\right)^{n_i - y_i} \frac{1}{2}I_{[-1,1]}(\xi).$$

We provide sufficient conditions for propriety of the posterior density, $\pi(\beta, \xi|\mathbf{y})$. It is proper if and only if

$$c(\mathbf{y}) := \int_{-1}^1 \int_{\mathbb{R}^k} f(\mathbf{y}|\beta, \xi) d\beta d\xi < \infty.$$

We denote the pmf of y_i by

$$f(y_i|\beta, \xi) = \binom{n_i}{y_i} (1 - G_\xi(-x'_i\beta))^{y_i} (G_\xi(-x'_i\beta))^{n_i - y_i},$$

and have

$$f(y_i|\beta, \xi) \leq \begin{cases} G_\xi(-x'_i\beta) & \text{if } y_i = 0, \\ 1 - G_\xi(-x'_i\beta) & \text{if } y_i = n_i, \\ \binom{n_i}{y_i} (1 - G_\xi(-x'_i\beta)) (G_\xi(-x'_i\beta)) & \text{if } 1 \leq y_i \leq n_i - 1. \end{cases}$$

Let $N_n = \{1, \dots, n\}$. We partition N_n as $N_n = I_1 \uplus I_2 \uplus I_3$, where $I_1 = \{i \in N_n : y_i = 0\}$, $I_2 = \{i \in N_n : y_i = n_i\}$ and $I_3 = \{i \in N_n : 1 \leq y_i \leq n_i - 1\}$ (Chen, Ibrahim, and Shao (2004b)), so

$$\begin{aligned} f(\mathbf{y}|\beta, \xi) &= \prod_{i=1}^n f(y_i|\beta, \xi) \\ &\leq \prod_{i \in I_1} G_\xi(-x'_i\beta) \prod_{i \in I_2} (1 - G_\xi(-x'_i\beta)) \prod_{i \in I_3} \binom{n_i}{y_i} (1 - G_\xi(-x'_i\beta)) (G_\xi(-x'_i\beta)). \end{aligned}$$

Let $q = \#(I_3)$ be the cardinality of I_3 . Let the $q \times k$ matrix with rows $x'_i, i \in I_3$ be denoted by \tilde{X} and let \tilde{X} be the $(n + q) \times k$ matrix

$$\begin{pmatrix} X \\ \tilde{X} \end{pmatrix}.$$

Define $\tau_1, \tau_2, \dots, \tau_{n+q}$ where $\tau_i = -1$ if $i \leq n$ and $i \in I_1 \cup I_3$, $\tau_i = 1$ if $i \leq n$ and $i \in I_2$, and $\tau_{n+i} = 1$ for $i = 1, \dots, q$. Let X^* be the $(n + q) \times k$ matrix with i th row $-\tau_i \tilde{x}'_i$, where \tilde{x}'_i is the i th row of \tilde{X} , and let $X_{\ell, m}^*$ be the $(m - \ell) \times k$ matrix with rows $-\tau_i \tilde{x}'_i, \ell < i \leq m$.

Theorem 1. *Suppose that there exist integers p, m_0, \dots, m_p such that $p > k$, $0 = m_0 < \dots < m_p \leq n + q$, and that $X_{m_{\ell-1}, m_\ell}^*$ is of full rank with positive vectors a_1, a_2, \dots, a_p such that $a'_\ell X_{m_{\ell-1}, m_\ell}^* = 0$ for $\ell = 1, \dots, p$. Then $c(\mathbf{y}) < \infty$.*

The proof of the Theorem 1 is given in Appendix A.

Notice that binary regression models can be obtained as a special case of binomial regression models by taking $n_i = 1$ for $i = 1, \dots, n$. In this case $I_3 = \emptyset, q = 0$, and X^* is an $n \times k$ matrix with i th row $x'_i I_{\{0\}}(y_i) - x'_i I_{\{1\}}(y_i)$. In order to gain intuition behind the conditions in Theorem 1, consider the special case of binary regression models. If X is of full rank, the existence of a positive vector a with $a'X^* = 0$ implies that there is no point $\beta_0 \in \mathbb{R}^k \setminus \{0\}$

such that $x_i^* \beta_0 \leq 0$ for all $i = 1, 2, \dots, n$ (Roy and Hobert (2007, p.261)). Thus every point in $\mathbb{R}^k \setminus \{0\}$ lies on the positive side of some of the n hyperplanes $x_i^* \beta_0 = 0$ and on the negative side of the rest. The existence of a positive vector a satisfying $a'X^* = 0$ also implies that the data set is *overlapped* (Albert and Anderson (1984)). Since the GEV distribution need not have higher order moments (for example, it does not have finite second moment for $\xi \geq 1/2$), we need to impose stronger conditions than the mere existence of a positive vector a satisfying $a'X^* = 0$ (Chen and Shao (2000)). If we assume that $\xi < 1/k$, then the GEV distribution has finite k th moment, and the existence of $a > 0$ with $a'X^* = 0$ implies that $c(\mathbf{y}) < \infty$. Roy and Hobert (2007) provide a simple way to check the existence of a positive vector a with $a'X^* = 0$ that involves maximizing $\mathbf{1}'g$ subject to $g'X^* = 0, (J - I)g \leq \mathbf{1}$ (element wise), and $g_i \geq 0$ for $i = 1, \dots, n + q$, where $\mathbf{1}$ and J denote a column vector and the matrix of 1s, respectively. This can be easily implemented in many statistical software languages. For example, the “simplex” function in the “boot” library of R (R Development Core Team (2011)) can be used.

Theorem 2. *For binary regression models, for $c(\mathbf{y}) < \infty$ it is necessary that the design matrix X is of full rank, and there exists a positive vector $a = (a_1, a_2, \dots, a_n)' \in \mathbb{R}^n$ such that $a'X^* = 0$.*

The proof of the Theorem 2 is given in Appendix B. Thus, if it is assumed that $\xi < 1/k$, these conditions are necessary and sufficient for $c(\mathbf{y}) < \infty$.

3.2. Jeffreys prior

Consider the prior on (β, ξ) given by

$$\pi_1(\beta, \xi) = \pi(\beta|\xi)\pi(\xi),$$

where $\pi(\beta|\xi) \propto |I(\beta|\xi)|^{1/2}$, with $I(\beta|\xi)$ the Fisher information matrix for the Binomial distribution with the GEV link and $\pi(\xi) = 0.5I_{[-1,1]}(\xi)$. The posterior density is then

$$\pi_1(\beta, \xi|\mathbf{y}) \propto f(\mathbf{y}|\beta, \xi)|I(\beta|\xi)|^{1/2}I_{[-1,1]}(\xi).$$

Theorem 3. *The posterior density $\pi_1(\beta, \xi|\mathbf{y})$ is proper.*

The proof of the Theorem 3 is given in Appendix C.

4. Generalized Extreme Value Link for Independent Ordinal Regression Models

Suppose we have n observations y_1, y_2, \dots, y_n , where y_i takes value in $\{j : j = 1, \dots, J\}$. A common way to model ordinal data is to consider underlying continuous latent variables $w_i, i = 1, \dots, n$ and assume that we observe

$$y_i = j \quad \text{if} \quad \gamma_{j-1} < w_i \leq \gamma_j,$$

where $-\infty = \gamma_0 < \gamma_1 < \gamma_2 < \dots < \gamma_{J-1} < \gamma_J = \infty$ are cut point parameters that determine the discretization of the data into J ordered categories (Albert and Chib (1993)). Here we assume that

$$w_i = x'_i \beta + \epsilon_i ,$$

$i = 1, \dots, n$, where the x'_i 's are k -dimensional vectors of covariates, β is the vector of regression parameters, and $\epsilon_i \sim GEV(\mu = 0, \sigma = 1, \xi)$ (Wang and Dey (2011)). Since

$$\begin{aligned} P(y_i = j) &= P(\gamma_{j-1} < w_i \leq \gamma_j) \\ &= P(\gamma_{j-1} - x'_i \beta < \epsilon_i \leq \gamma_j - x'_i \beta) \\ &= G_\xi(\gamma_j - x'_i \beta) - G_\xi(\gamma_{j-1} - x'_i \beta) \end{aligned}$$

the likelihood function for the above model is

$$L(\beta, \gamma, \xi | \mathbf{y}) = \prod_{i=1}^n \left[G_\xi(\gamma_{y_i} - x'_i \beta) - G_\xi(\gamma_{y_i-1} - x'_i \beta) \right].$$

Consider the priors on the parameters β , ξ , and $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_{J-1})$ given by

$$\begin{aligned} \Pi(\beta) &\propto 1, \quad \beta \in \mathbb{R}^p, \\ \Pi(\xi) &= \frac{1}{2} I_{[-1,1]}(\xi), \\ \Pi(\gamma) &= 1 I_{[\gamma_1 < \gamma_2 < \dots < \gamma_{J-1}]}(\gamma) . \end{aligned}$$

The posterior density of β , γ , and ξ is then

$$\begin{aligned} \Pi(\beta, \gamma, \xi | \mathbf{y}) &\propto L(\beta, \gamma, \xi | \mathbf{y}) \Pi(\beta) \Pi(\gamma) \Pi(\xi) \\ &\propto \prod_{i=1}^n \left[G_\xi(\gamma_{y_i} - x'_i \beta) - G_\xi(\gamma_{y_i-1} - x'_i \beta) \right] I_{[\gamma_1 < \gamma_2 < \dots < \gamma_{J-1}]}(\gamma) I_{[-1,1]}(\xi) . \end{aligned}$$

We provide sufficient conditions for the posterior density $\Pi(\beta, \gamma, \xi | \mathbf{y})$ to be proper. In order to state them, we introduce some notations. Partition the set $N_n = \{1, \dots, n\}$ into $N_n = U \uplus L \uplus M$ where

$$\begin{aligned} U &= \{i \in N_n : y_i = J\}, \\ L &= \{i \in N_n : y_i = 1\}, \\ M &= \{i \in N_n : 1 < y_i < J\}. \end{aligned}$$

Let X be the $n \times k$ design matrix with rows x'_i and take $x_i^* = (1, x'_i)'$ for $i = 1, \dots, n$.

Theorem 4. *Assume that*

(A1) *there exists $p > k + J - 1$ such that we can partition*

$$U = \biguplus_{\ell=1}^p U_\ell, \quad L = \biguplus_1^p L_\ell, \quad \text{and} \quad M = \biguplus_1^p M_\ell,$$

where U_ℓ , and L_ℓ are non-empty for $\ell = 1, \dots, p$. Define

$$\begin{aligned} X_{1\ell} &= \{x_i^*, i \in U_\ell, -x_j^*, j \in L_\ell \cup M_\ell\}', \\ X_{2\ell} &= \{x_j^*, j \in L_\ell, -x_i^*, i \in U_\ell \cup M_\ell\}', \ell = 1, \dots, p. \end{aligned}$$

Then the posterior is proper if one of the following two conditions is satisfied.

(A2)' $X_{1\ell}$ is of full column rank and $\exists b_\ell > 0$ such that $b'_\ell X_{1\ell} = 0$ for $\ell = 1, \dots, p$.

(A2)'' $X_{2\ell}$ is of full column rank and $\exists b_\ell > 0$ such that $b'_\ell X_{2\ell} = 0$ for $\ell = 1, \dots, p$.

The proof of Theorem 4 is given in Appendix D.

Here is another set of sufficient conditions for a proper posterior. Following Chen and Shao (1999), we define

$$\begin{aligned} T_{\ell,1} &= \{(i, j) : i \in U, j \in L, x_{i\ell} - x_{j\ell} > 0\}, \\ T_{\ell,-1} &= \{(i, j) : i \in U, j \in L, x_{i\ell} - x_{j\ell} < 0\}. \end{aligned}$$

For $\eta = (\eta_1, \eta_2, \dots, \eta_k)$ where $\eta_\ell = \pm 1$, let

$$T(\eta) = \bigcap_{\ell=1}^k T_{\ell, \eta_\ell}.$$

Suppose there exist $U(\eta) \subset U$ and $L(\eta) \subset L$ such that $U(\eta) \times L(\eta) \subset T(\eta)$. Let $M^* = \min_{\eta} \min(\#U(\eta), \#L(\eta))$, where as before $\#A$ is the cardinality of the set A .

Theorem 5. *If $M^* > k + J - 1$, then the posterior $\Pi(\beta, \gamma, \xi | \mathbf{y})$ is proper.*

The proof of Theorem 5 is given in Appendix E.

Note that if either L or U is empty, then there is no information available to estimate γ_1 or γ_{J-1} , so we need at least the sets U and L to be non-empty for a proper posterior. On the other hand, the posterior can still be proper even if the set M is empty. Also, the full rank condition of the design matrix is a necessary condition for the posterior to be proper (Chen and Shao (1999)).

5. Generalized Extreme Value Distribution in Survival Analysis

5.1. Shape of the hazard function

Suppose T denotes time to failure. We assume that $\log T \sim GEV(\mu = 0, \sigma = 1, \xi)$, so the pdf of T is

$$f_{\xi}(t) = \begin{cases} \frac{\exp[-(1+\xi \log t)^{-1/\xi}]}{t(1+\xi \log t)^{1/\xi+1}} & t > \exp(-\frac{1}{\xi}) & \text{if } \xi > 0; \\ & t < \exp(-\frac{1}{\xi}) & \text{if } \xi < 0, \\ \exp(-\frac{1}{t}) \cdot \frac{1}{t^2} & 0 < t < \infty & \text{if } \xi = 0. \end{cases}$$

The survival function $S_{\xi}(t) = P(T \geq t)$ here is

$$S_{\xi}(t) = \begin{cases} 1 - \exp\left(- (1 + \xi \log t)_+^{-\frac{1}{\xi}}\right) & \text{if } \xi \neq 0, \\ 1 - \exp(-\frac{1}{t}) & \text{if } \xi = 0, \end{cases}$$

and the hazard function $\lambda_{\xi}(t) = f_{\xi}(t)/S_{\xi}(t)$ is

$$\lambda_{\xi}(t) = \begin{cases} \frac{1}{t(1+\xi \log t)_+^{1/\xi+1} \left\{ \exp\left((1+\xi \log t)_+^{-1/\xi}\right) - 1 \right\}} & \text{if } \xi \neq 0, \\ \frac{1}{t^2(\exp(1/t)-1)} & \text{if } \xi = 0. \end{cases}$$

Figure 1 shows the plot of the GEV hazard function for $\xi = 0.3, 0, -0.3, -0.5$, and -1.5 . It shows that a GEV model is extremely flexible in modeling survival data. Another advantage is that by varying only the shape parameter, ξ , we obtain different shapes for the hazard function, whereas mixture models provide flexibility at the expense of many extra parameters.

We have a result regarding the hazard function of $\lambda_0(t)$.

Theorem 6. *The hazard function, $\lambda_0(t)$ is an upside down function.*

The proof of Theorem 5 is given in Appendix F.

5.2. Generalized extreme value regression models

Here we consider GEV as the error distribution in accelerated failure time models. Let T_i denote the failure times, and assume that

$$\log T_i \sim GEV(x_i' \beta, \sigma, \xi),$$

for $i = 1, \dots, n$, where the x_i 's are the k dimensional covariates, β is the vector of regression coefficients, and σ is the scale parameter. A version of the extreme value distribution is widely used in survival data analysis; for example,

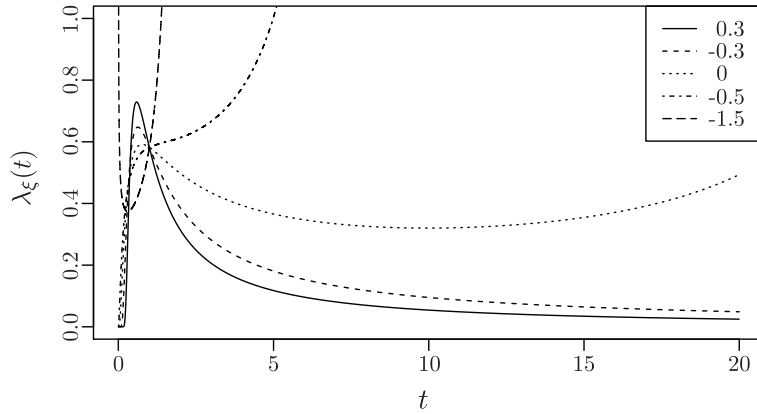


Figure 1. Hazard functions of the generalized extreme value distribution for different values of ξ .

Kim and Ibrahim (2000) consider the extreme value regression model when T_i has a Weibull distribution.

Let $\{(t_i, \nu_i); i = 1, \dots, n\}$ be the observed data where $t_i, i = 1, \dots, n$ denotes the observed failure or right censored time and ν_i is an indicator variable taking value 1 if t_i is an observed failure time and 0 if t_i is censored. Let $\mathbf{t} = (t_1, t_2, \dots, t_n)$ and $\boldsymbol{\nu} = (\nu_1, \nu_2, \dots, \nu_n)$. The likelihood function (assuming right censoring) is given by

$$L(\beta, \sigma, \xi | \mathbf{t}, \boldsymbol{\nu}) = \prod_{i=1}^n \left\{ \frac{1}{\sigma t_i \left(1 + \xi \frac{y_i - x'_i \beta}{\sigma}\right)_+^{1/\xi+1}} \exp \left[- \left(1 + \xi \frac{y_i - x'_i \beta}{\sigma}\right)_+^{-1/\xi} \right] \right\}^{\nu_i} \times \left\{ 1 - \exp \left[- \left(1 + \xi \frac{y_i - x'_i \beta}{\sigma}\right)_+^{-1/\xi} \right] \right\}^{1-\nu_i}, \tag{5.1}$$

where $y_i = \log t_i$ for $i = 1, \dots, n$.

The GEV distribution is irregular as its support depends on the parameters (Smith (1985)). When σ and ξ are known, the Jeffreys prior for β , $\pi(\beta | \sigma, \xi)$, is proportional to the square root of the determinant of the Fisher information matrix, $\pi(\beta | \sigma, \xi) \propto |I(\beta | \sigma, \xi)|^{1/2}$. It can be shown that

$$I(\beta | \sigma, \xi) := E \left(- \frac{\partial^2}{\partial \beta_i \partial \beta_j} \log L(\beta, \sigma, \xi | \mathbf{t}, \boldsymbol{\nu}) \Big| \beta, \sigma, \xi \right)_{1 \leq i, j \leq k} = E(X^T W X | \beta, \sigma, \xi),$$

where X is the $n \times k$ covariate matrix, and W is an $n \times n$ diagonal matrix. The i th diagonal element of W is a very complicated function of $d_i = 1 + \xi(y_i - x'_i \beta)/\sigma$. So we use a uniform prior on β . Kim and Ibrahim (2000) made similar comments regarding the Jeffreys prior for their extreme value regression model.

Consider a prior on (β, σ, ξ) ,

$$\pi(\beta, \sigma, \xi) \propto \pi(\sigma)\pi(\xi); \beta \in \mathbb{R}^k, \sigma \in \mathbb{R}^+,$$

where $\pi(\sigma)$ is a (proper or improper) density on \mathbb{R}^+ and $\pi(\xi) = 0.5I_{[-1,1]}(\xi)$. The posterior density is

$$\pi(\beta, \sigma, \xi | \mathbf{t}, \nu) \propto L(\beta, \sigma, \xi | \mathbf{t}, \nu) \pi(\sigma) \pi(\xi). \quad (5.2)$$

Theorem 7. *Let \check{X} be an $n \times k$ matrix with rows $\nu_i x'_i, i = 1, \dots, n$. Assume that $r(\check{X}) = k$ and $\int_0^\infty (1/\sigma^{m-k}) \pi(\sigma) d\sigma < \infty$, where $m = \#\{i : \nu_i = 1\}$ is the number of uncensored observations. Then the posterior density in (5.2) is proper.*

The proof of Theorem 7 is given in Appendix G.

Remark 1. [Kim and Ibrahim (2000)] considered conditions for posterior propriety in the special case that T_i has a Weibull distribution. One of their conditions was that the likelihood function based on any $n - k$ observations be bounded. But we have propriety without such a restriction.

5.3. An illustrative example

We consider the survival data on 40 advanced lung cancer patients as in Lawless (2003, p. 7). The dataset has three covariates: performance status (PS) at diagnosis (a measure of general medical condition on a scale of 10 to 90, with lower numbers indicating poorer conditions), age of the patient at diagnosis in years (age), and the number of months from diagnosis of cancer to entry into the study (diag). Three of the 40 observations are censored. This dataset has been previously analyzed by Kim and Ibrahim (2000) who assumed that the survival time follows a Weibull distribution. The shape parameter of the Weibull distribution was estimated to be 0.949 which implies a monotone decreasing hazard rate. Figure 2 shows the plots of the estimated baseline hazard function using the nonparametric kernel methods described in Müller and Wang (1994). We used the “muhaz” package in R (R Development Core Team (2011)) to make these plots. The plot in the left panel was obtained using the global bandwidth selection algorithms of Müller and Wang (1994) and the maximum time was taken to be the time at which ten patients remain at risk (default choice in the “muhaz” function). The plot in the right panel was based on the local bandwidth choices as prescribed in Müller and Wang (1994), and the time domain was stretched to the maximum observed survival time (999 days). The plots in Figure 2 suggest that the true hazard rate may be U-shaped or modified bathtub shaped so a Weibull model for the survival time may not be appropriate here. We used the

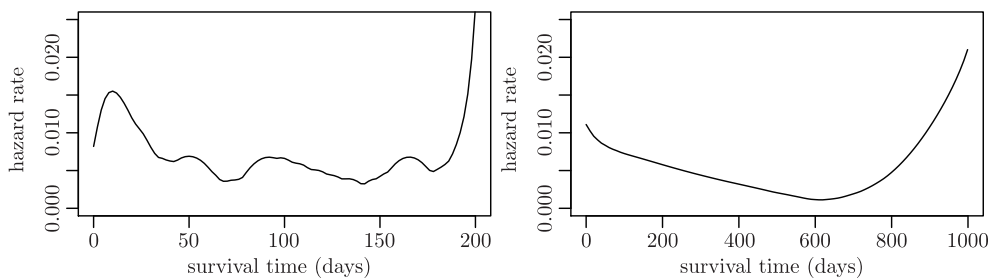


Figure 2. Estimated hazard functions for lung cancer patients.

GEV accelerated failure time model proposed in Section 5.2 to analyze this data set.

We considered the improper uniform prior on the regression coefficients, and the inverse gamma $IG(1,1)$ prior on σ . The posterior estimates reported here are fairly robust with respect to the hyperparameter values of the IG prior. Since $r(\check{X}) = 4$, from Theorem 7 we know that the posterior density, $\pi(\beta, \sigma, \xi | \mathbf{t}, \nu)$ in (5.2) is proper. We used the Metropolis-Hastings (with normal and truncated normal kernels) within Gibbs sampling algorithm for MCMC sampling. We standardized the covariate values to improve convergence of the MCMC algorithms. The R codes implementing the MCMC sampling scheme is available in the supplementary materials for the paper. The convergence of all the results was examined by visual trace plots, autocorrelation plots, Geweke's (1992) test statistic, and the Gelman-Rubin scale reduction factor (Brooks and Gelman (1998)) based on multiple sequences with widely dispersed starting values. The posterior means and 95% central credible intervals for ξ and σ were $-0.34(-0.62, -0.04)$, and $1.26(0.98, 1.67)$, respectively. The baseline hazard function corresponding to $\xi = -0.34$ and $\sigma = 1.26$ is modified bathtub shaped (as with $\xi = -0.3$ in Figure 1). The posterior means and 95% central credible intervals for the intercept parameter and the regression coefficients corresponding to the three variables PS, age, and diag were $3.72(3.25, 4.16)$, $1.16(0.74, 1.6)$, $0.07(-0.38, 0.52)$, and $0.04(-0.32, 0.42)$, respectively. As noted by Lawless (2003), the variable PS is important whereas the other two variables are not significant.

6. Concluding Remarks

Extending our results to multivariate categorical response and discrete choice models is quite challenging (Chen, Dey, and Ibrahim (2004a)). For the life testing and survival analysis models, further study can be done on fitting regression models for ordinal response and a proportional hazards model with a frailty distribution. The methodology proposed here can be extended to left censored or interval censored data.

Acknowledgements

The authors thank an editor and two anonymous reviewers for helpful comments and valuable suggestions that led to several improvements in the manuscript.

Appendices

A. Proof of Theorem 1

Proof of Theorem 1. Take u, u_1, \dots, u_{n+q} to be iid random variables with common distribution function $G_\xi(\cdot)$. Let $u^* = (\tau_1 u_1, \tau_2 u_2, \dots, \tau_{n+q} u_{n+q})'$, where the τ_i 's are as defined in Section 3.1. Then by Fubini's Theorem,

$$\begin{aligned} & \int_{-1}^1 \int_{\mathbb{R}^k} f(\mathbf{y}|\beta, \xi) d\beta d\xi \\ & \leq \left[\prod_{i \in I_3} \binom{n_i}{y_i} \right] E \left\{ \int_{-1}^1 \int_{\mathbb{R}^k} I(\tau_i u_i \geq -\tau_i x'_i \beta, 1 \leq i \leq n+q) d\beta d\xi \right\} \\ & \leq \left[\prod_{i \in I_3} \binom{n_i}{y_i} \right] \int_{-1}^1 \int_{\mathbb{R}^{n+q}} \int_{\mathbb{R}^k} I(\|\beta\| \leq c \min_{\ell} \max_{m_{\ell-1} < i \leq m_{\ell}} |u_i|) d\beta d\mathbf{G}_\xi d\xi \\ & = c^* \int_{-1}^1 \int_{\mathbb{R}^{n+q}} \left(\min_{\ell} \max_{m_{\ell-1} < i \leq m_{\ell}} |u_i| \right)^k d\mathbf{G}_\xi d\xi \\ & \leq c^* \int_{-1}^1 \prod_{\ell=1}^p \left(\sum_{m_{\ell-1} < i \leq m_{\ell}} E_\xi |u_i|^{k/p} \right) d\xi, \end{aligned} \tag{A.1}$$

where c and c^* are two constants and $d\mathbf{G}_\xi = dG_\xi(u_1) \dots dG_\xi(u_{n+q})$.

Note that if we can show that $E_\xi(|u|^a)$ is a continuous function of ξ when $0 < a < 1$, then it will follow that (A.1) is finite. Since $u \sim G_\xi(\cdot)$, the pdf of u is

$$g_\xi(u) = \begin{cases} \exp \left[- \{1 + \xi u\}^{-1/\xi} \right] \frac{1}{\{1 + \xi u\}^{1/\xi + 1}} & u > -\frac{1}{\xi} \text{ if } \xi > 0 ; u < -\frac{1}{\xi} \text{ if } \xi < 0, \\ \exp(-\exp(-u)) \exp(-u) & -\infty < u < \infty \text{ if } \xi = 0. \end{cases}$$

For $\xi \neq 0$, taking the transformation $t = (1 + \xi u)^{-1/\xi}$, it follows that

$$E_\xi(|u|^a) = \int_0^\infty \left| \frac{1}{\xi} (t^{-\xi} - 1) \right|^a e^{-t} dt .$$

Similarly when $\xi = 0$, taking the transformation $t = e^{-u}$, it follows that

$$E_\xi(|u|^a) = \int_0^\infty |-\log t|^a e^{-t} dt .$$

For fixed $t > 0$ let

$$h_t(\xi) = \begin{cases} \frac{t^{-\xi} - 1}{\xi} & \text{if } \xi \neq 0, \\ -\log t & \text{if } \xi = 0. \end{cases}$$

For $\xi \neq 0$, $h'_t(\xi) = \{-(\xi \log t + 1)t^{-\xi} + 1\}/\xi^2$. Since $x - x \log x \leq 1$ for $x > 0$, $h'_t(\xi) \geq 0$. So for fixed $t > 0$, we have

$$-(t - 1) \leq h_t(\xi) \leq \frac{1}{t} - 1 \text{ for } -1 \leq \xi \leq 1.$$

Then, if $0 < |\xi| \leq 1$, we have

$$\begin{aligned} \left| \frac{t^{-\xi} - 1}{\xi} \right| &\leq \max\left(\frac{1}{t} + 1, t + 1\right) \text{ for } t > 0, \\ \left| \log t \right| &\leq \max\left(\frac{1}{t} + 1, t + 1\right) \text{ for } t > 0. \end{aligned}$$

Let

$$h(t) = \begin{cases} \left(\frac{1}{t} + 1\right)^a & \text{if } 0 < t < 1, \\ (t + 1)^a & \text{if } t \geq 1. \end{cases}$$

So that for $0 < |\xi| \leq 1$,

$$\begin{aligned} \left| \frac{1}{\xi} (t^{-\xi} - 1) \right|^a &\leq h(t) \text{ for } t > 0, \\ \left| \log t \right|^a &\leq h(t) \text{ for } t > 0. \end{aligned}$$

As $0 < a < 1$,

$$\int_0^\infty h(t)e^{-t} dt < \infty.$$

Since for any fixed $t > 0$, $|\frac{1}{\xi}(t^{-\xi} - 1)|^a$ is a continuous function of ξ , by the Dominated Convergence Theorem it follows that $E_\xi|u|^a$ is a continuous function of ξ , which completes the proof of Theorem 1.

B. Proof of Theorem 2

Proposition B.1. *The family of distribution functions $\{G_\xi(\cdot)\}$ is stochastically increasing.*

Proof of Proposition B.1. We are to show that for given x , $G_\xi(x)$ is a strictly decreasing function in ξ . When $\xi \neq 0$, $G_\xi(x) = \exp[-\{1 + \xi x\}_+^{-1/\xi}]$. Take $\tilde{G}_x(\xi) := -\log(G_\xi(x)) = \{1 + \xi x\}_+^{-1/\xi}$, so $\log \tilde{G}_x(\xi) = -\log(1 + \xi x)/\xi$ when $1 + \xi x > 0$. For fixed x we denote $d\tilde{G}_x(\xi)/d\xi$ by $\tilde{G}'_x(\xi)$, so

$$\begin{aligned} \frac{\tilde{G}'_x(\xi)}{\tilde{G}_x(\xi)} &= -\frac{\xi \frac{x}{1+\xi x} - \log(1 + \xi x)}{\xi^2} \\ &= \frac{1 - (1 + \xi x) + (1 + \xi x) \log(1 + \xi x)}{\xi^2(1 + \xi x)}. \end{aligned}$$

Then, since $\tilde{G}_x(\xi) = \{1 + \xi x\}_+^{-1/\xi} > 0$, from the inequality $y - y \log y \leq 1$ for nonnegative y , it follows that for fixed x , $\tilde{G}'_x(\xi) > 0$ for $\xi \neq 0$. Then by the Mean Value Theorem, it follows that $\tilde{G}_\xi(x)$ is an increasing function in ξ on the entire real line. Hence for fixed x , $G_\xi(x)$ is a decreasing function in ξ .

Proof of Theorem 2. If X is not a full rank matrix then $\int_{\mathbb{R}^k} f(\mathbf{y}|\beta, \xi) d\beta = \infty$ for all ξ . Note that $0 < G_\xi(x) < 1$ for all $x \in [-1/2, 1/2]$ if $\xi \in [-1, 1]$. In particular, for $\delta \in (0, 1/2)$, $0 < G_\xi(-\delta) \leq G_\xi(\delta) < 1$ for all $\xi \in [-1/2, 1/2]$. Since the y_i 's are binary random variables, if there does not exist any positive vector $a \in \mathbb{R}^n$ with $a'X^* = 0$, by doing similar calculations as in the proof of Chen and Shao's (2000) Theorem 2.2, we have

$$\begin{aligned} & \int_{-1}^1 \int_{\mathbb{R}^k} f(\mathbf{y}|\beta, \xi) d\beta d\xi \\ & \geq \int_{-1/2}^{1/2} \int_{\mathbb{R}^k} f(\mathbf{y}|\beta, \xi) d\beta d\xi \\ & \geq c_1 \prod_{i:y_i=0} \left\{ 1 - \sup_{\xi \in [-\frac{1}{2}, \frac{1}{2}]} G_\xi(\delta) \right\} \prod_{i:y_i=1} \left\{ \inf_{\xi \in [-\frac{1}{2}, \frac{1}{2}]} G_\xi(-\delta) \right\} \int_{s_1 \geq 0, |s_j| \leq \eta, 2 \leq j \leq k} ds \\ & = c_1 \{1 - G_{-0.5}(\delta)\}^{p_1} \{G_{0.5}(-\delta)\}^{n-p_1} \int_{s_1 \geq 0, |s_j| \leq \eta, 2 \leq j \leq k} ds \\ & = \infty, \end{aligned}$$

where c_1 is a nonzero constant, $\eta > 0$ is chosen such that $k\eta \max_{1 \leq i \leq n} \|x_i\| \leq \delta$, $ds = ds_1 \dots ds_k$, and $p_1 = \#\{i : y_i = 0\}$; the first equality follows from Proposition B.1.

C. Proof of Theorem 3

Proof of Theorem 3. Since the likelihood function $f(\mathbf{y}|\beta, \xi)$ is bounded, it is enough to show that the prior $\pi_1(\beta, \xi)$ is proper. We know that the Fisher information matrix $I(\beta|\xi)$ can be written as $I(\beta|\xi) = X'\Omega(\beta|\xi)X$ where $\Omega(\beta|\xi)$ is an $n \times n$ diagonal matrix with i th diagonal element $\omega_i = n_i v_i \delta_i^2$, $v_i = v(x'_i \beta) = d^2 b(\theta_i) / d\theta_i^2$, and $\delta_i = \delta(x'_i \beta) = d\theta_i / d\eta_i$ is the so-called "link adjustments" ($\eta_i = x'_i \beta$). Here we use the standard notation θ_i to denote the canonical parameter for the binomial, and $b(\theta_i) = \log(1 + e^{\theta_i})$. Then following Ibrahim and Laud (1991), we have

$$\int_{-1}^1 \int_{\mathbb{R}^k} \pi_1(\beta, \xi) d\beta d\xi \leq \sum_T (c(x_{i_1}, x_{i_2}, \dots, x_{i_k}))^{1/2} \int_{-1}^1 \int_{\mathbb{R}^k} \prod_{j=1}^k n_{i_j}^{1/2} v_{i_j}^{1/2} \delta_{i_j} d\beta d\xi, \tag{C.1}$$

where $T = \{(i_1, i_2, \dots, i_k) : 1 \leq i_1 < \dots < i_k \leq n\}$, x'_{i_j} is the i_j th row of X , $c(x_{i_1}, x_{i_2}, \dots, x_{i_k}) = |X_*|^2$, and X_* is a $k \times k$ matrix with j th column x_{i_j} . Now,

without loss of generality, we can assume that X_* is non-singular since otherwise $c(x_{i_1}, x_{i_2}, \dots, x_{i_k}) = 0$. Then, as in Ibrahim and Laud (1991), considering the transformation $u = X_*\beta$ and letting $r_{i_j} = \theta(u_{i_j})$, it follows that a non-zero term in the expression on the right hand side of (C.1) is proportional to

$$\prod_{j=1}^k n_{i_j}^{1/2} \int_{-1}^1 \int_{-\infty}^{\infty} \left(\frac{d^2b(r_{i_j})}{dr_{i_j}^2} \right)^{1/2} dr_{i_j} d\xi.$$

The proof follows from the fact that $\int_{-1}^1 \int_{-\infty}^{\infty} (d^2b(r_{i_j})/dr_{i_j}^2)^{1/2} dr_{i_j} d\xi = 2\pi$.

D. Proof of Theorem 4

Proof of Theorem 4. Let r_1, r_2, \dots, r_n be iid random variables with common distribution $G_\xi(\cdot)$. So

$$\begin{aligned} & G_\xi(\gamma_{y_i} - x'_i\beta) - G_\xi(\gamma_{y_{i-1}} - x'_i\beta) \\ &= \int I(\gamma_{y_{i-1}} - x'_i\beta < r_i \leq \gamma_{y_i} - x'_i\beta) dG_\xi(r_i). \end{aligned}$$

By Fubini's Theorem,

$$\begin{aligned} c_1(\mathbf{y}) &:= \int_{-1}^1 \int_{\gamma_1 < \dots < \gamma_{J-1}} \cdots \int_{\mathbb{R}^k} \left\{ \prod_{i=1}^n [G_\xi(\gamma_{y_i} - x'_i\beta) - G_\xi(\gamma_{y_{i-1}} - x'_i\beta)] \right\} d\beta d\gamma d\xi \\ &= \int_{-1}^1 \int_{\mathbb{R}^n} \int_{\gamma_1 < \dots < \gamma_{J-1}} \cdots \int_{\mathbb{R}^k} I\{\gamma_{y_{i-1}} - x'_i\beta < r_i < \gamma_{y_i} - x'_i\beta; 1 \leq i \leq n\} \\ & \hspace{20em} d\beta d\gamma dG_\xi(\tilde{r}) d\xi, \end{aligned}$$

where $dG_\xi(\tilde{r}) = dG_\xi(r_1) \cdots dG_\xi(r_n)$. With

$$\begin{aligned} h(\tilde{r}) &= \int_{\gamma_1 < \dots < \gamma_{J-1}} \cdots \int_{\mathbb{R}^k} I\{\gamma_{y_{i-1}} - x'_i\beta < r_i < \gamma_{y_i} - x'_i\beta; 1 \leq i \leq n\} d\beta d\gamma, \\ c_1(\mathbf{y}) &= \int_{-1}^1 \int_{\mathbb{R}^n} h(\tilde{r}) dG_\xi(\tilde{r}) d\xi. \end{aligned}$$

We show that

$$h(\tilde{r}) \leq C \min_{1 \leq i \leq p} \max_{j \in Q_i} |r_j|^{k+J-1}, \tag{D.1}$$

where $Q_i = U_i \cup L_i \cup M_i$, and C is a constant depending on X and y only. Then,

$$\begin{aligned} c_1(\mathbf{y}) &\leq C \int_{-1}^1 \int_{\mathbb{R}^n} \min_{1 \leq i \leq p} \max_{j \in Q_i} |r_j|^{k+J-1} dG_\xi(\tilde{r}) d\xi \\ &\leq C \int_{-1}^1 \prod_{i=1}^p \left(\sum_{j \in Q_i} E_\xi |r_j|^{(k+J-1)/p} \right) d\xi. \end{aligned}$$

Since from the proof of Theorem 1 we know that $E_\xi|r_j|^{(k+J-1)/p}$ is a continuous function of ξ , it follows that $c_1(\mathbf{y}) < \infty$. Now, we show that (D.1) holds.

Consider the transformation $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_{J-1}) \rightarrow \theta = (\theta_1, \theta_2, \dots, \theta_{J-1})$ with $\theta_1 = \gamma_1$, $\theta_i = \gamma_i - \gamma_{i-1}$ for $2 \leq i \leq J-1$. The Jacobian can be shown to be 1. With $\tilde{\theta} = (\theta_2, \theta_3, \dots, \theta_{J-1})$,

$$\begin{aligned}
h(\tilde{r}) &= \int_{(\mathbb{R}^+)^{J-2}} \int_{\mathbb{R}^k} \int_{-\infty}^{\infty} I(\theta_1 \geq r_i + x'_i \beta; i \in L) \times I\left(\sum_{\ell=1}^{J-1} \theta_\ell < r_i + x'_i \beta, i \in U\right) \\
&\quad \times I\left(\sum_{\ell=1}^{y_i-1} \theta_\ell < r_i + x'_i \beta, i \in M\right) \times I\left(\sum_{\ell=1}^{y_i} \theta_\ell \geq r_i + x'_i \beta, i \in M\right) d\theta_1 d\beta d\tilde{\theta} \\
&= \int_{(\mathbb{R}^+)^{J-2}} \int_{\mathbb{R}^k} \left(\min_{1 \leq i \leq p} \left\{ \min_{j \in U_i \cup M_i} \left[r_j + x'_j \beta - \sum_{\ell=2}^{y_j-1} \theta_\ell \right] \right\} \right. \\
&\quad \left. - \max_{1 \leq i \leq p} \left\{ \max_{j \in L_i \cup M_i} \left[r_j + x'_j \beta - \sum_{\ell=2}^{y_j} \theta_\ell \right] \right\} \right) \\
&\quad \times I\left(\min_{1 \leq i \leq p} \left\{ \min_{j \in U_i \cup M_i} \left[r_j + x'_j \beta - \sum_{\ell=2}^{y_j-1} \theta_\ell \right] \right\} \right. \\
&\quad \left. \geq \max_{1 \leq i \leq p} \left\{ \max_{j \in L_i \cup M_i} \left[r_j + x'_j \beta - \sum_{\ell=2}^{y_j} \theta_\ell \right] \right\} \right) d\beta d\tilde{\theta} \\
&\leq \int_{(\mathbb{R}^+)^{J-2}} \int_{\mathbb{R}^k} \min_{1 \leq i \leq p} \{f(i) - g(i)\} I\left(\min_{1 \leq i \leq p} \{f(i) - g(i)\} \geq 0\right) d\beta d\tilde{\theta}, \quad (\text{D.2})
\end{aligned}$$

where

$$\begin{aligned}
f(i) &= \min_{j \in U_i \cup M_i} \left(r_j + x'_j \beta - \sum_{\ell=2}^{y_j-1} \theta_\ell \right), \\
g(i) &= \max_{j \in L_i \cup M_i} \left(r_j + x'_j \beta - \sum_{\ell=2}^{y_j} \theta_\ell \right).
\end{aligned}$$

Then with a similar calculation as in Chen and Shao (1999), we get

$$\begin{aligned}
f(i) - g(i) &= \min_{j \in U_i \cup M_i} \left(r_j + x'_j \beta - \sum_{\ell=2}^{y_j-1} \theta_\ell \right) - \max_{j \in L_i \cup M_i} \left(r_j + x'_j \beta - \sum_{\ell=2}^{y_j} \theta_\ell \right) \\
&= \min_{j \in L_i \cup M_i} \left(-r_j - x'_j \beta + \sum_{\ell=2}^{y_j} \theta_\ell \right) - \max_{j \in U_i \cup M_i} \left(-r_j - x'_j \beta + \sum_{\ell=2}^{y_j-1} \theta_\ell \right) \\
&\leq 2 \max_{j \in Q_i} |r_j| - \tilde{M} \left\{ \max_{j \in U_i \cup M_i} \left(\sum_{\ell=2}^{y_j-1} \frac{\theta_\ell}{\tilde{M}} - \sum_{\ell=1}^k x_{j\ell} \frac{\beta_\ell}{\tilde{M}} \right) \right\}
\end{aligned}$$

$$- \min_{j \in L_i \cup M_i} \left(\sum_{\ell=2}^{y_j} \frac{\theta_\ell}{\tilde{M}} - \sum_{\ell=1}^k x_{j\ell} \frac{\beta_\ell}{\tilde{M}} \right) \Big\},$$

where $\tilde{M} = \max(|\beta_\ell|, \theta_\ell) > 0$. Take

$$d_i = \inf_{\substack{0 \leq a_\ell \leq 1; 2 \leq \ell \leq J-1 \\ -1 \leq b_r \leq 1, 1 \leq r \leq k}} \left\{ \max_{j \in U_i \cup M_i} \left(\sum_{\ell=2}^{y_j-1} a_\ell + \sum_{r=1}^k x_{jr} b_r \right) - \min_{j \in L_i \cup M_i} \left(\sum_{\ell=2}^{y_j} a_\ell + \sum_{r=1}^k x_{jr} b_r \right) \right\},$$

and $d = \min_{1 \leq i \leq p} d_i$. So,

$$f(i) - g(i) \leq 2 \max_{j \in Q_i} |r_j| - \tilde{M}d.$$

From (D.2), we need only consider the case

$$0 \leq \min_{1 \leq i \leq p} (f(i) - g(i)).$$

Thus if $d > 0$ then

$$\tilde{M} \leq \frac{2}{d} \min_{1 \leq i \leq p} \max_{j \in Q_i} |r_j|,$$

which implies that

$$\begin{aligned} h(\tilde{r}) &\leq 2 \iint_{\{\max(|\beta_\ell|, \theta_\ell) \leq \frac{2}{d} \min_{1 \leq i \leq p} \max_{j \in Q_i} |r_j|; \theta_\ell \geq 0\}} \min_{1 \leq i \leq p} \max_{j \in Q_i} |r_j| d\beta d\tilde{\theta} \\ &\leq C \min_{1 \leq i \leq p} \max_{j \in Q_i} |r_j|^{k+J-1}, \end{aligned}$$

where C is a constant. Thus (D.1) is proved if we can show that $d > 0$.

If $d_i > 0$ for all i , then $d > 0$. We show that (A2)' implies that $d_i > 0$ for all $i = 1, \dots, p$. With calculations as in Chen and Shao (1999), we can show that (A2)' implies that, $\forall 0 \leq a_v \leq 1, 2 \leq v \leq J-1$, and $\forall -1 \leq b_r \leq 1, 1 \leq r \leq k, \Sigma |b_r| > 0$,

$$\min_{j \in L_i \cup M_i} \left(\sum_{v=2}^{y_j} a_v + \sum_{r=1}^k x_{jr} b_r \right) \leq \max_{j \in U_i \cup M_i} \left(\sum_{v=2}^{y_j-1} a_v + \sum_{r=1}^k x_{jr} b_r \right), \quad (D.3)$$

and the equality in (D.3) holds only if

$$\sum_{r=1}^k x_{jr} b_r = c,$$

for some constant c and for all $j \in Q_i$. That is, the equality in (D.3) holds only if

$$X_{1i} \begin{pmatrix} -c \\ \tilde{b} \end{pmatrix} = 0,$$

where $\tilde{b} = (b_1, \dots, b_k)'$. But this contradicts the fact that X_{1i} is assumed to be of full column rank. Since the a_v 's and b_r 's are defined on compact intervals, it follows that $d_i > 0$ for all $i = 1, \dots, p$, which completes the proof.

E. Proof of Theorem 5

Proof of Theorem 5. Doing similar calculations as in Chen and Shao (1999), we can show that

$$\begin{aligned}
 c_1(\mathbf{y}) &\leq c_1 \sum_{\substack{\eta_\ell = \pm 1 \\ 1 \leq \ell \leq k}} \int_{-1}^1 E \left[\min_{(i,j) \in T(\eta)} (|r_i| + |r_j|)^{k+J-1} \right] d\xi \\
 &\leq 2^{k+J-1} c_2 \sum_{\substack{\eta_\ell = \pm 1 \\ 1 \leq \ell \leq k}} \int_{-1}^1 \left[\prod_{i \in U(\eta)} E_\xi \left(|r_i|^{(k+J-1)/\#U(\eta)} \right) \right. \\
 &\quad \left. + \prod_{j \in L(\eta)} E_\xi \left(|r_j|^{(k+J-1)/\#L(\eta)} \right) \right] d\xi, \tag{E.1}
 \end{aligned}$$

where c_1 and c_2 are two finite constants. Since $M^* > k + J - 1$, from the proof of Theorem 1 it follows that the integrand in (E.1) is a continuous function of ξ , and hence $c_1(\mathbf{y}) < \infty$.

F. Proof of Theorem 6

Proof of Theorem 6. Since $f_0(t) = \exp\{-1/t\}/t^2$,

$$f'_0(t) = \frac{e^{-1/t}(1 - 2t)}{t^4}.$$

So we obtain

$$\eta(t) := -\frac{f'_0(t)}{f_0(t)} = \frac{2t - 1}{t^2}, \eta'(t) = \frac{2(1 - t)}{t^3}.$$

Hence from Glaser (1980) it follows that $\lambda_0(t)$ is either upside-down or a decreasing function of t . Then the proof follows from the fact that $\lim_{t \rightarrow 0} \lambda_0(t) = 0$.

G. Proof of Theorem 7

Proof of Theorem 7. In (5.1), note that if $\nu_i = 0$, then

$$\left\{ 1 - \exp \left[- \left(1 + \xi \frac{y_i - x'_i \beta}{\sigma} \right)_+^{-1/\xi} \right] \right\}^{1-\nu_i} \leq 1. \tag{G.1}$$

On the other hand when $\nu_i = 1$, we show that there exists a finite constant M such that

$$\left\{ \frac{1}{\sigma t_i \left(1 + \xi (y_i - x'_i \beta) / \sigma \right)_+^{1/\xi+1}} \exp \left[- \left(1 + \xi \frac{y_i - x'_i \beta}{\sigma} \right)_+^{-1/\xi} \right] \right\}^{\nu_i} \leq \frac{M}{\sigma t_i}. \tag{G.2}$$

For a fixed $\xi \geq -1$, let $f_\xi(v) = v^{\xi+1}e^{-v}, v > 0$. It can be shown that $f_\xi(v) \leq (\xi + 1)^{\xi+1}e^{-(\xi+1)}$ for all $v > 0$. Let $M := \sup_{\xi \in [-1,1]} (\xi + 1)^{\xi+1}e^{-(\xi+1)}$. Then (G.2) follows since

$$\begin{aligned} & \frac{1}{\left(1 + \xi(y_i - x'_i\beta)/\sigma\right)_+^{1/\xi+1}} \exp \left[- \left(1 + \xi \frac{y_i - x'_i\beta}{\sigma}\right)_+^{-1/\xi} \right] \\ &= \left\{ \left(1 + \xi \frac{y_i - x'_i\beta}{\sigma}\right)_+^{-1/\xi} \right\}^{\xi+1} \exp \left[- \left(1 + \xi \frac{y_i - x'_i\beta}{\sigma}\right)_+^{-1/\xi} \right]. \end{aligned}$$

As \check{X} is of full rank, there must exist k linearly independent covariate vectors x_{i_1}, \dots, x_{i_k} such that $\nu_{i_1} = \dots = \nu_{i_k} = 1$. Without loss of generality, we assume that $i_1 = 1, \dots, i_k = k$.

The posterior density $\pi(\beta, \sigma, \xi | \mathbf{t}, \nu)$ in (5.2) is proper if

$$\int_{-1}^1 \int_0^\infty \int_{\mathbb{R}^k} L(\beta, \sigma, \xi | \mathbf{t}, \nu) \pi(\sigma) \pi(\xi) d\beta d\sigma d\xi < \infty.$$

As before let $N_k = \{1, \dots, k\}$. From (G.1) and (G.2) we have

$$\begin{aligned} & \int_{-1}^1 \int_0^\infty \int_{\mathbb{R}^k} L(\beta, \sigma, \xi | \mathbf{t}, \nu) \pi(\sigma) \pi(\xi) d\beta d\sigma d\xi \\ & \leq \int_{-1}^1 \int_0^\infty \int_{\mathbb{R}^k} \left(\prod_{i:\nu_i=0} 1 \right) \left(\prod_{i:\nu_i=1, i \notin N_k} \frac{M}{\sigma t_i} \right) \\ & \quad \times \left(\prod_{i=1}^k \frac{1}{\sigma t_i \left(1 + \xi \frac{y_i - x'_i\beta}{\sigma}\right)_+^{1/\xi+1}} \exp \left[- \left(1 + \xi \frac{y_i - x'_i\beta}{\sigma}\right)_+^{-1/\xi} \right] \right) \\ & \quad \pi(\sigma) \pi(\xi) d\beta d\sigma d\xi. \tag{G.3} \end{aligned}$$

Consider the one-to-one, linear transformation $w_i = x'_i\beta, i = 1, 2, \dots, k$. The right hand side of (G.3) is proportional to

$$\begin{aligned} & \int_{-1}^1 \int_0^\infty \frac{1}{\sigma^{m-k}} \left(\prod_{i=1}^k \int_{\mathbb{R}} \frac{1}{\sigma \left(1 + \xi(y_i - w_i)/\sigma\right)_+^{1/\xi+1}} \exp \left[- \left(1 + \xi \frac{y_i - w_i}{\sigma}\right)_+^{-1/\xi} \right] dw_i \right) \\ & \quad \pi(\sigma) \pi(\xi) d\sigma d\xi \\ &= \int_{-1}^1 \pi(\xi) d\xi \int_0^\infty \frac{1}{\sigma^{m-k}} \pi(\sigma) d\sigma < \infty, \end{aligned}$$

completing the proof.

References

- Albert, A. and Anderson, J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika* **71**, 1-10.
- Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *J. Amer. Statist. Assoc.* **88**, 669-679.
- Aranda-Ordaz, F. L. (1981). On two families of transformations to additivity for binary response data. *Biometrika* **68**, 357-363.
- Arnold, B. and Groeneveld, R. (1995). Measuring skewness with respect to the mode. *Amer. Statist.* **49**, 34-38.
- Banerjee, T., Chen, M.-H., Dey, D. K. and Kim, S. (2007). Bayesian analysis of generalized odds-rate hazards models for survival data. *Lifetime Data Anal.* **13**, 241-260.
- Barlow, R. and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing*. Holt, Rinehart, and Winston, New York.
- Bennett, S. (1983). Logistic regression model for survival data. *Appl. Statist.* **32**, 165-171.
- Brooks, S. P. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *J. Comput. Graph. Statist.* **7**, 434-455.
- Chen, M.-H. and Dey, D. K. (2000). Bayesian analysis for correlated ordinal data models. In *Generalized Linear Model: A Bayesian perspective* (Edited by D. Dey, S. Ghosh and B. Mallick), 133-157. Marcel-Dekker, New York.
- Chen, M.-H., Dey, D. K. and Ibrahim, J. G. (2004a). Bayesian criterion based assessment for categorical data. *Biometrika* **91**, 40-63.
- Chen, M.-H., Dey, D. K. and Shao, Q.-M. (1999). A new skewed link model for dichotomous quantal response data. *J. Amer. Statist. Assoc.* **94**, 1172-86.
- Chen, M.-H., Ibrahim, J. G. and Shao, Q.-M. (2004b). Propriety of the posterior distribution and existence of the MLE for regression models with covariates missing at random. *J. Amer. Statist. Assoc.* **99**, 421-438.
- Chen, M.-H. and Shao, Q.-M. (1999). Existence of Bayesian estimates for the polychotomous quantal response models. *Ann. Inst. Statist. Math.* **51**, 637-656.
- Chen, M.-H. and Shao, Q.-M. (2000). Propriety of posterior distribution for dichotomous quantal response models. *Proc. Amer. Math. Soc.* **129**, 293-302.
- Coles, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer Verlag, New York.
- Czado, C. and Santner, T. (1992). The effect of link misspecification on binary regression inference. *J. Statist. Plann. Inference* **33**, 213-231.
- Dahan, E. and Mendelson, H. (2001). An extreme value model of concept testing. *Management Sci.* **47**, 102-116.
- Finkelstein, M. (2009). Understanding the shape of the mixture failure rate (with engineering and demographic applications). *Appl. Stochastic Models in Business and Industry* **25**, 643-663.
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In *Bayesian Statistics 4* (Edited by J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), 169-193. Clarendon Press, Oxford, UK.
- Glaser, R. E. (1980). Bathtub and related failure rate characterizations. *J. Amer. Statist. Assoc.* **75**, 667-672.

- Guerreo, V. M. and Johnson, R. A. (1982). Use of the box-cox transformation with binary response models. *Biometrika* **69**, 309-14.
- Gurland, J. and Sethuraman, J. (1995). How pooling failure data may reverse interesting failures rate. *J. Amer. Statist. Assoc.* **90**, 1416-1423.
- Ibrahim, J. and Laud, P. (1991). On Bayesian analysis of generalized linear models using Jeffreys's prior. *J. Amer. Statist. Assoc.* **86**, 981-986.
- Kim, S., Chen, M. H. and Dey, D. K. (2008). Flexible generalized t-link models for binary response data. *Biometrika* **95**, 93-106.
- Kim, S. and Ibrahim, J. (2000). On Bayesian inference for proportional hazards models using noninformative priors. *Lifetime Data Anal.* **6**, 331-341.
- Langlands, A. O., Pocock, S. J., Kerr, G. R. and Gore, S. M. (1979). Long term survival of patients with breast cancer: A study of the curability of the disease. *British Medical J.* **17**, 1247-1251.
- Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*. 2nd edition. Wiley, New Jersey.
- Lieberman, G. J. (1969). The status and impact of reliability methodology. *Naval Research Logistics Quarterly* **14**, 17-35.
- Mann, N. R., Schafer, R. E. and Singpurwalla, N. D. (1974). *Methods for Statistical Analysis of Reliability and Life Data*. Wiley, New York.
- Morgan, B. J. T. (1983). Observations on quantitative analysis. *Biometrics* **39** 879-86.
- Müller, H.-G. and Wang, J.-L. (1994). Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics* **50**, 61-76.
- Nelson, W. (1990). *Accelerated Testing: Statistical Models, Test Plans and Data Analyses*. John Wiley, New York.
- R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Roy, V. and Hobert, J. P. (2007). Convergence rates and asymptotic standard errors for MCMC algorithms for Bayesian probit regression. *J. Roy. Statist. Soc. Ser. B* **69**, 607- 623.
- Sang, H. and Gelfand, A. E. (2009). Hierarchical modeling for extreme values observed over space and time. *Environmental and Ecological Statist.* **16**, 407-426.
- Sang, H. and Gelfand, A. E. (2010). Continuous spatial process models for extreme values. *J. Agricultural, Biological and Environmental Statist.* **15**, 49-65.
- Smith, R. L. (1985). Maximum likelihood estimation in a class of non-regular cases. *Biometrika* **72**, 67-90.
- Smith, R. L. (1989). Extreme value analyses of environmental time series: An application to trend direction in ground-level ozone (with discussion). *Statist. Sci.* **4**, 367-393.
- Stukel, T. (1988). Generalized logistic models. *J. Amer. Statist. Assoc.* **83**, 426-431.
- Wang, X. and Dey, D. K. (2010). Generalized extreme value regression for binary response data: An application to B2B electronic payments system adoption. *Ann. Appl. Statist.* **4**, 2000-2023.
- Wang, X. and Dey, D. K. (2011). Generalized extreme value regression for ordinal response data. *Environmental and Ecological Statist.* **18**, 619-634.

- Wang, X., Dey, D. K. and Banerjee, S. (2010). Non-gaussian hierarchical generalized linear geostatistical model selection. In *Frontiers of Statistical Decision Making and Bayesian Analysis* (Edited by M. H. Chen, D. K. Dey, P. Müller, D. Sun and K. Ye), 484-496. Springer.
- Whitmore, A. S. (1983). Transformations to linearity in binary regression. *SIAM J. Appl. Math.* **43**, 703-710.

Department of Statistics, Iowa State University, 3415 Snedecor Hall, Ames, IA 50011, USA.

E-mail: vroy@iastate.edu

Department of Statistics, University of Connecticut, 215 Glenbrook Road, Unit 4098, Storrs, CT 06269-4098, USA.

E-mail: dipak.dey@uconn.edu

(Received January 2011; accepted March 2013)