Abstract

We consider a Bayesian analysis of Binomial response data with covariates. To describe the problem under investigation, suppose we have \( n \) independent binomial observations \( Y_1, \ldots, Y_n \) where \( Y_i \sim \text{Bin}(m_i, \theta_i) \) and let \( x_i \) be \( p \)-dimensional covariate vector associated with \( Y_i \) for \( i = 1, \ldots, n \). Binomial observations can be analyzed through a generalized linear model (GLM) where we assume \( \theta_i = F(x_i^T \beta) \) for some known distribution function \( F(\cdot) \) and \( \beta \) is the vector of unknown regression coefficients. In this paper, we state necessary and sufficient conditions for propriety of the posterior distribution of \( \beta \) if an improper uniform prior is used on \( \beta \). We also consider situations where the link function is not pre-specified but belongs to a parametric family and the link function parameters are estimated along with the regression coefficients. In this case, we investigate the propriety of the joint posterior distributions of \( \beta \) and the link function parameters. There are a number of parametric family of link functions available in the literature. As a specific example, we consider Pregibon’s (1980) link function and show that our general posterior propriety results can be used to establish propriety of the posterior distributions corresponding to the Pregibon’s (1980) link. We show that Pregibon’s (1980) simple one parameter family of link function can be used to fit both positively and negatively skewed response curves. Moreover, the conditions for posterior propriety corresponding to the Pregibon’s (1980) link can be easily checked and are milder than those required by the flexible GEV link of Wang and Dey (2010). As an illustration, we analyze a data set from Ramsey and Schafer (2002) regarding the relationship between dose of Aflatoxicol and odds of liver tumor in rainbow trouts. In this example, the symmetric logit link fails to fit the data, whereas Pregibon’s (1980) skewed link yields a slightly better fit than the GEV link.

Key words and phrases. Bayesian regression, Binary, Binomial, Link function parameter, Posterior propriety, Tolerance distribution, Toxicity tests
1 Introduction

Binomial and binary response data with one or more covariates can be analyzed through a basic generalized linear model with a specified link function. Let $Y = (Y_1, \ldots, Y_n)$ be $n$ independent binomial or binary random variables with probability mass functions

\[ f(y_i|\theta_i) = \binom{m_i}{y_i} \theta_i^{y_i} (1 - \theta_i)^{m_i - y_i}; \quad y_i = 0, 1, \ldots, m_i, \]

for $m_i \geq 1$, $i = 1, 2, \ldots, n$, with

\[ \theta_i = F(x_i^T \beta), \]

where $F$ is a cumulative distribution function (cdf) ($F^{-1}$ is typically called the link function), $x_i$'s are $p$-dimensional covariates and $\beta$ is the vector of unknown regression coefficients. Some frequently used links are logit, probit and complimentary log-log links. The logit link is obtained by setting $F^{-1}(\theta_i) = \log\{\theta_i/(1 - \theta_i)\}$, whereas the function $F(\theta_i) = \Phi(\theta_i)$ yields the probit link, where $\Phi$ is the cdf of $N(0, 1)$. The complementary log-log link is specified as $F^{-1}(\theta_i) = -\log(-\log(\theta_i))$. Among these three links, both logit and probit are symmetric link functions whereas the complimentary log-log is an asymmetric link. However these popular link functions do not always provide good fit to the given data set and this can yield a substantial bias in the mean response estimates (Czado and Santner, 1992). In this case one might wish to extend the above formulation to one in which the link belongs to a parameterized family of functions, and the link function parameter is estimated along with the regression parameters.

Suppose the link function $F$ is embedded into a parametric family $\{F_\lambda(\eta), \lambda \in \Lambda\}$ where $\eta = x^T \beta$ is the linear predictor and $\Lambda \subset \mathbb{R}^l$ for some $l \geq 1$. So the joint probability mass function of $Y = (Y_1, \ldots, Y_n)^T$ becomes

\[ f_y(y|\beta, \lambda) = \prod_{i=1}^n \binom{m_i}{y_i} \{F_\lambda(\eta_i)\}^{y_i} \{1 - F_\lambda(\eta_i)\}^{m_i - y_i}, \]

where $\eta_i = x_i^T \beta$. While choosing an appropriate family of link functions $\{F^{-1}(\eta|\lambda)\}$, we need to consider the flexibility of the link functions in fitting a variety of response curves, and also the complexity of the link functions, for example, the dimension of $\Lambda$. To the best of our knowledge, consideration of parameterized link function began with Pregibon (1980), who showed that a score test for adequacy of a hypothesized link function could be obtained by embedding the hypothesized link in a parametric family of link functions. We will describe what we will call Pregibon's (1980) link function in detail in Section 3. Pregibon’s (1980) link function has only one parameter and we show that it can fit both
positively and negatively skewed response curves. Stukel (1988) proposed a class of generalized logistic models. Stukel’s (1988) two-parameter models are general and several commonly used link functions, such as the probit and complimentary log-log link models can be approximated by members of this family. Czado (1994a) studied a standardized link family called the generalized probit regression model. Chen, Dey and Shao (1999) considered a class of skewed link functions, where an underlying latent variable has a mixed-effects model structure. Kim, Chen and Dey (2008) introduced a class of generalized skewed t-link models. Some other link functions and related works include Whitemore (1983), Czado (1994b) and Czado and Raftery (2006). More recently, Wang and Dey (2010) considered a flexible skewed link function based on the generalized extreme value (GEV) distribution. The superiority of the GEV link over the logit, probit and the complimentary log-log link is due to the fact that the shape parameter of the GEV distribution purely controls its tail behavior. In fact, Wang and Dey (2010) also demonstrated that the GEV link even often dominates the Kim et al.’s (2008) skewed-t link, due to more flexibility of the GEV distribution.

Other than flexibility and complexity of the link function, another important consideration while choosing a link function is whether, in the presence of covariates, these link functions yield proper posterior distributions when noninformative improper priors are used for the regression coefficients. In the absence of prior information, often an improper uniform prior is used for the regression parameters \( \beta \), that is, it is assumed that the prior on \( \beta \), \( \pi(\beta) \propto 1; \beta \in \mathbb{R}^p \). However, use of an improper prior leads to a challenging question, whether the resulting posterior distribution is proper. It is important to verify the existence of a posterior distribution before conducting any inference based on what is taken to be a Monte Carlo sample from that distribution. For example, it can be shown that in the presence of covariates, Stukel’s (1988) link yields improper posterior distributions for many types of non-informative improper priors, including the improper uniform prior for the regression coefficients (Chen et al., 1999).

Our objective in this article is to present necessary and sufficient conditions for propriety of the joint posterior distribution of the regression coefficients \( \beta \) and the link function parameters \( \lambda \) that can be easily verified in applications. These conditions suppose an improper uniform prior on \( \beta \), and depend on both the covariates \( x_i \)’s and the particular link functions \( F(\eta|\lambda) \). As a specific example, we apply our general results to Pregibon’s (1980) link function for which verification of the needed conditions is particularly straightforward. This is in contrast to what is required in using the GEV link of Wang and Dey (2010) in practice. In achieving our objective we also generalize conditions for posterior propriety
under fixed link functions for binary data to the aggregated setting of binomial responses. In addition, we rediscover the flexibility of very simple family of link functions.

There have been some works on establishing propriety of the posterior distribution for binary and binomial regression models with pre-specified (fixed) link functions. Ibrahim and Laud (1991) gave necessary and sufficient conditions for propriety of the posterior distribution of $\beta$ for a generalized linear model with Jeffreys prior on $\beta$. Recently Chen, Ibrahim and Kim (2008) studied theoretical and computational properties of Jeffreys prior for inference in binomial regression models with fixed link functions. For binary response models with fixed link functions, Chen and Shao (2000) provided necessary and sufficient conditions for posterior propriety under improper uniform priors on $\beta$. Natarajan and McCulloch (1995) gave necessary and sufficient conditions for propriety of the posterior distribution of the variance components in a class of mixed models for binary responses assuming proper prior on $\beta$. More recently, Chen, Ibrahim and Shao (2004) characterized the propriety of the posterior distribution for some general classes of regression models including generalized linear models with covariates missing at random.

The rest of the article is organized as follows. In Section 2 we present necessary and sufficient conditions for posterior propriety for binomial response models. In Section 3 we prove the existence of posterior distributions when Pregibon’s (1980) link function is used. As an illustration, in section 4 we analyze a data set from Ramsey and Schafer (2002) which is used to describe the relationship between dose of Aflatoxicol and odds of liver tumor in rainbow trout. Some concluding remarks are offered in Section 5.

2 Binomial response models

Recall from the Introduction that a standard binomial regression model assumes that $\theta_i = F(\eta_i)$, where $F$ is a fixed cdf. In this case, if an improper uniform prior is used on the regression coefficients $\beta$, that is, if the prior $\pi(\beta) \propto 1; \beta \in \mathbb{R}^p$, then the posterior density $\pi(\beta|y)$ is proportional to the likelihood function $f(y|\beta) = \prod_{i=1}^n \left( \frac{m_i}{y_i} \right) \left( F(\eta_i) \right)^{y_i} \left( 1 - F(\eta_i) \right)^{m_i - y_i}$. The posterior density $\pi(\beta|y)$ is proper if and only if

$$\int_{\mathbb{R}^p} f(y|\beta) \, d\beta < \infty. \quad (4)$$

Conditions satisfying (4) can be gleaned from Chen and Shao (2000) and Chen et al. (2004). For the sake of completeness we here provide necessary and sufficient conditions for (4) to hold.
In order to state the conditions for posterior propriety, we introduce some notation. Let \( N_n = \{1, 2, \cdots, n\} \) and as in Chen et al. (2004) partition the index set \( N_n \) as \( N_n = I_1 \uplus I_2 \uplus I_3 \), where we define \( I_1 = \{i \in N_n : y_i = 0\} \), \( I_2 = \{i \in N_n : y_i = m_i\} \) and \( I_3 = \{i \in N_n : 1 \leq y_i \leq m_i - 1\} \).

Let \( X \) be the \( n \times p \) design matrix, with \( i \)-th row \( x_i^T \) and let \( k = |I_3| \) be the cardinality of \( I_3 \). Denote the \( k \times p \) matrix, with rows \( x_i^T, i \in I_3 \) by \( \tilde{X} \) and let the \((n + k) \times p \) matrix \( X^* \) be

\[
X^* = \begin{pmatrix} X \\ \tilde{X} \end{pmatrix}.
\]

We define \( \tau_1, \tau_2, \cdots, \tau_{n+k} \) where \( \tau_i = 1 \) if \( i \in I_1 \cup I_3 \), \( \tau_i = -1 \) if \( i \in I_2 \) and \( \tau_{n+j} = -1 \) for \( j = 1, 2, \cdots, k \). Let \( W \) be the \((n + k) \times p \) matrix with \( i \)-th row being \( \tau_i x_i^* \) where \( x_i^* \) is the \( i \)-th row of \( X^* \).

We show that under certain conditions on \( W \), \( \int_{\mathbb{R}^p} f(y|\beta) d\beta < C \int |u|^p F(u) du \), for some constant \( C \).

So if \( F \) has finite \( p \)th moment then (4) holds. We state this in the following theorem.

**Theorem 1.** Assume the following conditions are satisfied

A1 The design matrix \( X \) is full rank, that is, \( \text{rank}(X) = p \).

A2 There exists a vector \( g = (g_1, g_2, \ldots, g_{n+k})^T \in \mathbb{R}^{n+k} \) with \( g_i > 0 \) for all \( i = 1, 2, \ldots, n + k \), such that

\[ W^T g = 0. \]

A3 The cdf \( F \) has finite \( p \)th moment, that is \( \int |u|^p F(u) du < \infty \).

Then (4) holds.

The proof of Theorem 1 is given in Appendix A.

In the following corollary, we provide sufficient conditions for existence of posterior moments.

**Corollary 1.** Assume that conditions A1 and A2 of Theorem 1 are satisfied. For any \( r > 0 \), if \( \int |u|^{p+r} F(u) du < \infty \) then \( \int_{\mathbb{R}^p} ||\beta||^r f(y|\beta) d\beta < \infty \).

**Proof.** The proof follows using similar arguments as in the proof of Theorem 1. \( \Box \)

The following corollary states that the conditions A1 and A2 in Theorem 1 can be replaced by the condition that \((x_i, i \in I_3)^T \) is full rank.

**Corollary 2.** If \((x_i, i \in I_3)^T \) is full rank and \( \int |u|^p F(u) du < \infty \), then (4) holds.
The proof of Corollary 2 is given in Appendix A.

Remark 1. If there exists a positive vector $\tilde{g}$ such that $\tilde{W}^T \tilde{g} = 0$, where $\tilde{W} = (\tau_i x_i, i \in I_1 \cup I_2)^T$, then (A2) holds.

Remark 2. Roy and Hobert (2007) provide a simple way to check condition (A2) which involves maximizing $1^T g$ subject to $W^T g = 0$, $(J-I)g \leq 1$ (element wise) and $g_i \geq 0$ for $i = 1, 2, \ldots, n + k$ where $1$ and $J$ denote a column vector and the matrix of $1$s respectively. This can be easily implemented in most statistical software languages.

Remark 3. Binary regression models can be obtained as a special case of binomial regression models by taking $m_i = 1$ for $i = 1, 2, \ldots, n$. In this case $I_3 = \phi$, a null set, $k = 0$, $X^* = X$ and $W$ is an $n \times p$ matrix with $i$th row $x_i^T I\{0\}(y_i) - x_i^T I\{1\}(y_i)$. Conditions A1 and A2 of Theorem 1 are then precisely conditions (C1) and (C2) of Chen and Shao (2000) for propriety of the posterior distribution of $\beta$ for binary response variables.

The next theorem provides a necessary condition for (4).

**Theorem 2.** If $0 < F(0-) \leq F(0) < 1$, then A1 and A2 are necessary conditions for (4).

The proof of Theorem 2 is given in Appendix A.

We now consider situations where the link function is not fixed and it belongs to a parametric family $\{F(\eta|\lambda), \lambda \in \Lambda\}$. Since we are assuming a uniform prior on $\beta$, the joint posterior density of $\beta$ and $\lambda$, if it exists, will be given by

$$
\pi(\beta, \lambda|y) \propto f(y|\beta, \lambda) \pi(\lambda), \quad \beta \in \mathbb{R}^p, \lambda \in \Lambda,
$$

where the likelihood function $f(y|\beta, \lambda)$ is defined in (3) and $\pi(\lambda)$ is the prior density for the link function parameter, $\lambda$. We assume that $\pi(\lambda)$ is a proper prior with the support $\Lambda$ being a compact set in $\mathbb{R}^l$. For example, if $\lambda$ is a scalar then in the absence of any prior information, a uniform prior can be used on $\lambda$, that is $\pi(\lambda) = \frac{1}{b-a}$ for $a \leq \lambda \leq b$. In practice, $a$ and $b$ can be chosen so that $\{F(\eta|\lambda), \lambda \in [a, b]\}$ is sufficiently flexible. For example, Wang and Dey (2010) argue that the Uniform $[-1, 1]$ prior on the shape parameter $\xi$ for their generalized extreme value (GEV) link provides flexible range of skewness in the link function. The posterior (5) will exist (i.e., the posterior is proper) if and only if

$$
c(y) = \int_{I^l} \int_{\mathbb{R}^p} f(y|\beta, \lambda) \pi(\lambda) \, d\beta \, d\lambda < \infty,
$$
and we now give necessary and sufficient conditions for this to be true.

Let, $\Pi(\lambda)$ be the probability measure corresponding to the prior pdf $\pi(\lambda)$ and write $c(y) = \int_\Lambda \int_{\mathbb{R}^p} f(y|\beta, \lambda) d\beta \Pi(d\lambda)$. It might be very difficult to directly show that $f(y|\beta, \lambda)$ is jointly integrable with respect to $d\beta$ and $\Pi(d\lambda)$. On the other hand, in order to show that $c(y) < \infty$, we may find an upper bound of $\int_{\mathbb{R}^p} f(y|\beta, \lambda) d\beta$ for any fixed $\lambda$ such that this upper bound is integrable with respect to $\Pi(d\lambda)$. For a fixed $\lambda$, we can use Theorem 1 to show that $\int_{\mathbb{R}^p} f(y|\beta, \lambda) d\beta \leq C m(\lambda)$, for some constant $C$, where $m(\lambda) = \int |u|^p dF_\lambda(u)$. So $c(y) \leq C \int_\Lambda m(\lambda) \Pi(d\lambda)$ and if $m(\lambda)$ is continuous in $\Lambda$, the posterior distribution $\pi(\beta, \lambda | y)$ is proper. We state it as the following theorem.

**Theorem 3.** If both $A1$ and $A2$ of Theorem 1 hold and if

$$A4 \quad \text{The function } m(\lambda) \text{ is continuous in } \Lambda,$$

then $c(y) < \infty$.

Notice that when the link function belongs to a parametric family, mere finiteness of the $p$th moment of $F_\lambda$ is not sufficient for posterior propriety. Similar to Corollary 1, the following corollary gives conditions for finiteness of posterior moments with respect to $\pi(\beta, \lambda | y)$.

**Corollary 3.** Assume that $A1$ and $A2$ are satisfied. For any $r > 0$, if $\int |u|^{p+r} F_\lambda(u) du$ is continuous on $[a, b]$ then $\int_\Lambda \int_{\mathbb{R}^p} ||(\beta, \lambda)||^r \pi(\beta, \lambda | y) d\beta d\lambda < \infty$.

We now give necessary conditions for $c(y) < \infty$.

**Theorem 4.** If there exists $\Lambda' \subseteq \Lambda$, such that for some $\delta > 0$,

$$0 < \inf_{\lambda \in \Lambda'} F_\lambda(-\delta) \leq \sup_{\lambda \in \Lambda'} F_\lambda(\delta) < 1,$$

then $A1$ and $A2$ are necessary conditions for $c(y) < \infty$.

The proof of Theorem 4 is given in Appendix A.

In the next section, we describe Pregibon’s (1980) link function. Applying our general results developed in this section, we show that the posterior distribution corresponding to Pregibon’s (1980) link is proper as long as the design matrix $X$ satisfies the conditions $A1$ and $A2$. Although we discuss Pregibon’s (1980) link in details, we can apply our general results to other families of link functions as well. For example, we have verified that the conditions developed in this section can be used to prove propriety of posterior densities corresponding to the families of link functions proposed in Aranda-Ordaz (1981) and Whitemore (1983).
3 Pregibon’s link

Pregibon’s (1980) link is a simple, one-parameter family of link functions and it is obtained by assuming \( \theta_i = H_\lambda(\eta_i) \) in (2), where

\[
H_\lambda(\eta) = \begin{cases} 
1 - \frac{1}{\{\lambda \exp(\eta) + 1\}^{1/\lambda}} & \text{if } \lambda \neq 0 \\
1 - \exp\{-\exp(\eta)\} & \text{if } \lambda = 0,
\end{cases}
\]

(6)

where \( x_+ = \max(x, 0) \). The inverse function of \( H_\lambda(\cdot) \) is given by

\[
g(\theta_i|\lambda) = \log \left[ \frac{(1 - \theta_i)^{-\lambda} - 1}{\lambda} \right] = \eta_i.
\]

(7)

Note that, for \( \lambda = 0 \), (7) reduces to the asymmetric complementary log-log link, whereas the logit link is obtained for \( \lambda = 1 \). So, these two models can be compared through the parameter \( \lambda \) if we use the above link function. Since we can’t analytically calculate the skewness of Pregibon’s (1980) link using the usual definition \( \mu_3/\mu_2^{3/2} \), as in Wang and Dey (2010), we use Arnold and Groeneveld’s (1995) skewness measure which is defined in terms of the mode of a distribution. For a continuous random variable \( X \) defined on an interval with cdf \( F \) and if the pdf of \( X \) is differentiable and \( M_x \) is the unique mode of \( X \), then Arnold and Groeneveld (1995) defined the skewness of \( X \) as \( 1 - 2F(M_x) \). Based on this skewness definition, the skewness of Pregibon’s (1980) link is given by \( \gamma_M = 1 - 2H_\lambda(0) = 2(1 + \lambda)^{-1/\lambda} - 1 \).

Note that \( \gamma_M > (\sim)0 \) if \( (1 + \lambda)^{1/\lambda} < (\sim)2 \). Since \( (1 + \lambda)^{1/\lambda} \) is a decreasing function, Pregibon’s (1980) link is positively skewed for \( \lambda > 1 \) and it is negatively skewed for \( \lambda < 1 \). On the other hand, Pregibon’s (1980) link with \( \lambda = 1 \), which is the logit link is symmetric. Note that \( \gamma_M \downarrow -1 \) as \( \lambda \downarrow -1 \) and \( \gamma_M \uparrow 1 \) as \( \lambda \uparrow \infty \). In particular, if \( \lambda \in [-1, 30] \) then \( \gamma_M \in [-1, 0.7837] \). On the other hand, the skewness of the Wang and Dey’s (2010) GEV link lies in \([-0.7291, 1]\) when the shape parameter of the GEV distribution \( \xi \in [-1, 1] \) (Wang and Dey (2010) used \( \pi(\xi) = 0.5, -1 \leq \xi < 1 \) as the prior on \( \xi \) for their posterior propriety results.). Figure 1 shows the plots of probability density functions and response curves corresponding to Pregibon’s (1980) link function with different values of \( \lambda \). From this figure, we see that Pregibon’s (1980) link can provide a wide range of skewness in fitted response curves.

As previously mentioned, the link function \( g(\theta_i|\lambda) \) first appeared in Pregibon (1980) and has been used by Aranda-Ordaz (1981) and others in analysis of binomial response data. In the context of acute toxicity tests, where Pregibon’s (1980) link function has been used, the link function parameter \( \lambda \) determines the shape of the tolerance distributions, while functions of the regression parameters determine location and scale of these distributions. The link function (7) is flexible enough to represent skew tolerance distribution in acute toxicity tests. Although even more flexible families of link functions are
The prior distributions on $\beta$ and $\lambda$ are $\pi(\beta) \propto 1; \beta \in \mathbb{R}^p$ and $\pi(\lambda) = \frac{1}{b-a}$ with $a \leq \lambda \leq b$, and the posterior distribution is then given by

$$\tilde{\pi}(\beta, \lambda | y) \propto \prod_{i=1}^{n} \left( \frac{m_i}{y_i} \right) (H_\lambda(\eta_i))^{y_i} (1 - H_\lambda(\eta_i))^{m_i-y_i} \pi(\lambda).$$

From Theorem 3 we know that $\tilde{\pi}$ is proper as long as $m_H(\lambda) := \int |u|^p dH_\lambda(u)$ is a continuous function on $[a, b]$ and $A1, A2$ hold. The following theorem states that $\tilde{\pi}$ is proper if $A1$ and $A2$ hold.

**Theorem 5.** If $A1$ and $A2$ of Theorem 1 are satisfied then the posterior distribution $\tilde{\pi}$ is proper.

The proof of Theorem 5 is given in Appendix A.

From Corollary 3 and the proof of Theorem 5 it follows that if $A1$ and $A2$ hold then $\tilde{\pi}$ has posterior moments of all orders. The following theorem shows that if $b > -1$, which is going to the case in any practical situation, then $A1$ and $A2$ are in fact necessary for $\tilde{\pi}$ to be proper.

**Theorem 6.** If $b > -1$ then $A1$ and $A2$ are necessary for $\tilde{\pi}$ to be proper.
The proof of Theorem 6 is given in Appendix A.

As mentioned in Remark 2, the conditions A1 and A2 can be easily checked using standard statistical packages. For example, the “simplex” function in the “boot” library of R (R Development Core Team, 2011) can be used to check the condition A2. The conditions for posterior propriety corresponding to the Wang and Dey’s (2010) GEV link with an improper uniform prior on \( \beta \) and a proper uniform \([-1, 1]\) prior on the shape parameter \( \xi \) of the GEV link, are much stronger than A1 and A2. Since the GEV distribution does not have higher order moments (for example, it does not have finite second moment for \( \xi \geq 1/2 \)), Wang and Dey (2010) needed to impose stronger conditions on the matrix \( X^* \) (defined in Section 2), than simply the existence of a positive vector \( g \) satisfying \( W^T g = 0 \) (A2). It is easy to construct examples where A2 holds but Wang and Dey’s (2010) sufficient conditions for posterior propriety fail to hold, e.g., consider the case \( n = 3, p = 2, m_i = 1 \) for \( i = 1, 2, 3 \), \( y = (0, 1, 0) \), \( x_1' = (1/2, 1/3) \), \( x_2' = (1, 1) \) and \( x_3' = (1/2, 2/3) \).

In the next section, we analyze a data set from Ramsey and Schafer (2002).

4 An example: Aflatoxical and Liver Tumors in Trout

An experiment at the Marine/Freshwater Biomedical Sciences Center at Oregon State University investigated the carcinogenic effects of Aflatoxicol, a metabolite of Aflatoxin B1, which is a toxic by-product produced by a mold that infects cottonseed meal, peanuts, and grains. Twenty tanks of rainbow trout embryos were exposed to one of five doses of Aflatoxicol for one hour. The complete data set is available in (Ramsey and Schafer, 2002, p. 641) that present the number of fish in each tank and the number of these that had liver tumors after one year. In Table 1 below we present the data set with counts summed over the four tanks. We disregard any tank effects that may be present in the data, and analyze the data set using Pregibon’s (1980) link function to describe the relationship between dose of Aflatoxicol and odds of liver tumor. We also compare the fit obtained by Pregibon’s (1980) link function with that of the logistic, complementary log-log and Wang and Dey’s (2010) GEV model.

In this example, \( p = 2 \) and \( \beta = (\beta_0, \beta_1) \) where \( \beta_0 \) and \( \beta_1 \) are intercept and slope parameters, respectively. The natural logarithm of dose of Aflatoxicol are used as covariates. We use an improper uniform prior on \( \beta \), that is, \( \pi(\beta) \propto 1 \) for all link models. For the shape parameter \( \lambda \) of Pregibon’s link function, we consider the Uniform\([-1, b]\) prior with \( b = 30 \). Other values of \( b \geq 10 \) do not change the posterior estimates substantially. We use the simple linear program mentioned in Remark 2 to verify that
<table>
<thead>
<tr>
<th>Dose (ppm)</th>
<th>Number of trout with liver tumors (y)</th>
<th>Number in tank (m)</th>
<th>y/m</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.010</td>
<td>25</td>
<td>347</td>
<td>0.072</td>
</tr>
<tr>
<td>0.025</td>
<td>132</td>
<td>346</td>
<td>0.382</td>
</tr>
<tr>
<td>0.050</td>
<td>226</td>
<td>353</td>
<td>0.640</td>
</tr>
<tr>
<td>0.100</td>
<td>281</td>
<td>355</td>
<td>0.792</td>
</tr>
<tr>
<td>0.250</td>
<td>286</td>
<td>338</td>
<td>0.846</td>
</tr>
</tbody>
</table>

Table 1: Dose of Aflatoxicol and odds of liver tumor in rainbow trout.

the condition A2 is satisfied. In this example $n = k = 5$. It is easy to see that rank $(X) = 2$. Since the maximizer $g^*$ (obtained by the “simplex” function in the “boot” library of R) maximizing $1^T g$ subject to $W^T g = 0, (J - I)g \leq 1$, and $g_i \geq 0$ for $i = 1, 2, \ldots, 10$, is such that $g_i^* > 0$ for all $i = 1, 2, \ldots, 10$, it follows from Roy and Hobert (2007) that the posterior density $\hat{\pi}(\beta, \lambda|y)$ given in (8) is proper. For the GEV link, following Wang and Dey (2010), we use the Uniform $[-1, 1]$ prior on $\xi$.

MCMC algorithms are used for exploring the posterior densities. For the logistic and the complementary log-log models, we employ a Metropolis Hastings algorithm with bivariate normal jumping kernels for MCMC sampling. We used $c(X^T X)^{-1}$ as the covariance matrix of the bivariate normal kernels where the constant $c$ is chosen such that 30% - 40% acceptance rate is obtained. For Pregibon’s model and the GEV model we use Metropolis Hastings within Gibbs sampling algorithm. Wang and Dey (2010) also used a similar algorithm for the GEV model. We standardize the covariate values to improve convergence of the MCMC algorithms. The convergence of all results was examined by visual trace plots and autocorrelations plots. We also calculated the Gelman-Rubin scale reduction factor (Brooks and Gelman, 1998) computed from three widely dispersed sets of starting values and the MCMC sampling is continued until the scale reduction factor is below 1.002. Geweke’s (1992) Z scores are calculated for testing equality of means between the first (10 %) and last (50 %) parts of the chains corresponding to each variables. For example, the values of Geweke’s (1992) test statistic corresponding to the $\beta_0, \beta_1$ and $\lambda$ variables of the chain for the posterior density $\hat{\pi}(\beta, \lambda|y)$ given in (8) are $-1.08, -1.04$ and $-1.25$ respectively.

In many families of parameterized link functions, such as that of Pregibon considered here, direct comparison of regression parameters is difficult among models with different values of the link parameter. For example, the logit link is a member of Pregibon’s family with $\lambda = 1$. For the data analyzed in this section, the estimated value of $\lambda$ was 5.54. The regression parameters for the logistic model were
estimated to be $\beta_0 = 0.27$ and $\beta_1 = 1.61$, while those for the model with estimated link were $\beta_0 = 3.68$ and $\beta_1 = 4.67$. The effect of such different values on the estimated response function cannot be immediately identified on the basis of direct comparison (e.g., 1.61 versus 4.67). For this reason, we focus on comparison of estimated response curves themselves rather than the parameters that control them. A plot of observed proportions and response curves corresponding to the different link models is presented in Figure 2. From the response curves plot, we see that the symmetric logit link fits the data poorly. The positively skewed complementary log-log link fits the data better than the symmetric logit link although there is still discrepancy between the observed values and response curve in the tails. This is because the skewness present in the data is more than what is available in the complementary log-log link, which provides only a fixed amount of skewness. On the other hand, both the GEV link and Pregibon’s link is better able to pick up the tail behavior than the logit or the complementary log-log link, albeit at the cost of an additional parameter in the model. We also plot the estimated tolerance densities corresponding to the different link models.

In order to examine the fit of the models to the data, we consider the following checks based on the posterior predictive distributions. We simulate $N = 30,000$ replicated data sets from the posterior predictive distributions corresponding to each of the four models. Our criteria are based on the number of fish with liver tumor at the lowest and highest doses of Aflatoxicol. For a given model $M$, let $y^*_{l,M,j}(y^*_{h,M,j})$ denote the response at the lowest (highest) dose observed in the $j$th sample from the posterior predictive distribution for $j = 1, \ldots, N$, and let $y_l(y_h)$ denote that same quantity in the actual observed data. The criteria for assessing performance of the models are

$$Q_{l,M} = \frac{1}{N} \sum_{j=1}^{N} I(y^*_{l,M,j} \leq y_l), \quad Q_{h,M} = \frac{1}{N} \sum_{j=1}^{N} I(y^*_{h,M,j} \geq y_h),$$

(9)

where $I(A)$ is the indicator function that assumes a value of 1 if the statement $A$ is true and 0 otherwise. These measures assess the ability of fitted models to reflect the observed data behavior at the lowest and highest doses used in the study. This is related to the flexibility of the estimated response functions in being able to curve rapidly enough at both lower and upper ends of the dose range to capture the extreme observed proportions. This, in turn, is related to tail behavior in tolerance distributions that correspond to those response curves. It is crucial that these measures be interpreted as a pair. A model that fits well will return values of these measures that are both reasonably large. A small value for one or the other indicates that estimation of one tail of the tolerance distribution is being compromised due to inflexibility of the response curve (i.e., link function).
The values of Table 2 indicate that the logistic and Cloglog models are not able to simultaneously fit both the small (0.072) and large (0.846) responses at the lowest and highest doses, respectively. That the values of $Q_{l,M}$ are more variable across models than the values of $Q_{h,M}$ implies that differences in estimated tolerance distributions are more pronounced in the left tail than the right tail, which is in concert with the visual impression of Figure 2.

<table>
<thead>
<tr>
<th>Link</th>
<th>$Q_{l,M}$</th>
<th>$Q_{h,M}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pregibon</td>
<td>0.48</td>
<td>0.91</td>
</tr>
<tr>
<td>GEV</td>
<td>0.47</td>
<td>0.90</td>
</tr>
<tr>
<td>Logit</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Cloglog</td>
<td>0.16</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Table 2: Values of $Q_{l,M}$ and $Q_{h,M}$ from $M = 30,000$ posterior predictive data sets.

We also used the Deviance Information Criterion (DIC) of Spiegelhalter, Best, Carlin and Linde (2002) to assess models under different link function, although some limitation of this measure has recently been identified Ando (2007). The deviance is defined as negative of twice the log-likelihood, that is $D(y, (\beta, \lambda)) = -2 \log f(y|\beta, \lambda)$. The (posterior) mean deviance $\hat{D}(y) = \sum_{i=1}^{N} D(y, (\beta_i, \lambda_i))/N$ serves as a Bayesian measure of fit or “adequacy” of a model. The DIC measure is defined as $\text{DIC} = \hat{D}(y) + p_D$, where $p_D$ is the effective number of parameters of a Bayesian model. The quantity $p_D$ is calculated as $p_D = \hat{D}(y) - D(y, (\hat{\beta}, \hat{\lambda}))$, where $\hat{\beta}$ and $\hat{\lambda}$ are posterior means of $\beta$ and $\lambda$. The smaller the value of DIC, the better the model fits the data. The values of DIC for the logit, complementary log-log, GEV, and the Pregibon’s link functions are 75.95, 46.82, 41.74 and 38.04 respectively. Prior impressions of fit are confirmed by these DIC values. The symmetric logit link is the poorest fit, a complimentary log-log link offers substantial improvement, the GEV link of Wang and Dey (2010) gives some additional improvement, and Pregibon (1980) link provides the best fit.

It is worth pointing out that the flexibility gained through use of a family of parameterized link functions is distinct from effects that might result from the use of different prior distributions on regression parameters with a fixed link. We previously have mentioned the analysis of a model with fixed link and Jeffreys’s prior as developed by Chen et al. (2008). This analysis is attractive in application because it avoids the need for MCMC while providing a transformation invariant prior. To illustrate that such an analysis may not attain the flexibility provided through the use of a parameterized link function, we fit
models with fixed links using the Jeffreys’s prior on $\beta$ and following the estimation procedure of Chen et al. (2008). We also produced samples from the associated posterior predictive distribution through the use of importance resampling with the weights obtained from that procedure. Models with both logit and complementary log log links were fit. The results were essentially identical to those reported previously for logistic and Cloglog models with uniform priors on $\beta$. Posterior means for the logistic model with Jeffreys’s prior were 0.28 and 1.61 for $\beta_0$ and $\beta_1$, respectively, while these values were 0.27 and 1.61 with a uniform prior. For the Cloglog model estimates using Jeffreys’s prior were 0.67 and 1.19, which are identical to two decimal places to those obtained using a uniform prior. Values of the posterior predictive test statistics $Q_{l,M}$ and $Q_{h,M}$ were also identical to those given previously in Table 2 for both logistic and Cloglog models. The DIC values for models with Jeffreys’s priors were 75.91 for the logistic model and 46.58 for the Cloglog model, nearly the same as the values of 75.95 and 46.82 reported in the previous paragraph for the models with uniform priors.

5 Concluding Remarks

In this article we have considered analysis of binomial response variables with a fixed or a class of parameterized link functions. We have provided necessary and sufficient conditions for posterior propriety if improper priors are placed on the regression parameters. Although we have verified our conditions for a number of families of link functions (see Section 2), here we considered Pregibon’s (1980) link function in details. We showed that when improper uniform priors are specified on the regression parameters, Pregibon’s (1980) link function leads to a proper posterior under certain easily verifiable conditions on the covariates, which is not the case for the GEV link of Wang and Dey (2010). The link function considered here is a fairly simple one-parameter family. The conditions for posterior propriety mentioned in this paper are general and can be used to verify propriety of posterior distributions when other families of link functions are used.

As an illustration, we analyze a data set from Ramsey and Schafer (2002) which shows the usefulness of estimating (rather than assuming a fixed link) link function parameters along with other parameters in the model. The class of problems that may be dealt with by models for binary or binomial responses with estimated link functions is, however, much larger than this one example. Physical systems in which failure of a component is related to one or more stressors, social systems in which the probability of a behavior is related to certain stimuli, and economic systems in which the occurrence of
Figure 2: Estimated response curves (upper) and tolerance densities (lower) for the trout data using models based on the Pregibon’s link, GEV link, Cloglog link and the Logit link.

A phenomenon is related to the level of a given indicator of activity are all areas in which these same models might prove useful. In all of these cases, levels of a covariate that produce a given event can
be represented as a distribution across a heterogeneous population. There is a correspondence between these distributions and the link function used in a binary or binomial generalized linear model. Because the shape of these distributions provides valuable information about the problem under investigation, it is beneficial to estimate them rather than assume a particular form as part of model formulation.

Appendices

A Proofs of main results

Some of the proofs in this appendix make use of results stated in Appendix B.

Proof of Theorem 1:

Proof. Let

\[ f(y_i | \beta) = \binom{m_i}{y_i} (F(\eta_i))^{y_i} (1 - F(\eta_i))^{m_i - y_i}. \]

It is easy to verify that,

\[ f(y_i | \beta) \leq \begin{cases} 1 - F(\eta_i) & \text{if } i \in I_1 \\ F(\eta_i) & \text{if } i \in I_2 \\ \binom{m_i}{y_i} (F(\eta_i))(1 - F(\eta_i)) & \text{if } i \in I_3 \end{cases}. \]

Then the following is an upper bound for \( f(y|\beta) \),

\[
  f(y|\beta) \leq \prod_{i \in I_1} (1 - F(\eta_i)) \prod_{i \in I_2} F(\eta_i) \prod_{i \in I_3} \binom{m_i}{y_i} (F(\eta_i))(1 - F(\eta_i)).
\]  

(10)

Let \( U \) be a random variable with distribution function \( F(\cdot) \). Then for \( -\infty < x < \infty \),

\[
  1 - F(x) = \mathbb{E}(I(U > x)) \quad \text{and} \quad F(x) = \mathbb{E}(I(U \geq -x)).
\]

Now define \( U_1, U_2, \ldots, U_{n+k} \) to be independent random variables with common distribution function \( F(\cdot) \). Then by Fubini’s theorem it follows from (10) that

\[
  \int_{\mathbb{R}^p} f(y|\beta) d\beta \leq \left[ \prod_{i \in I_3} \binom{m_i}{y_i} \right] \mathbb{E} \left( \int_{\mathbb{R}^p} I(\tau_i x_i \beta^T \leq \tau_i U_i; 1 \leq i \leq n + k) d\beta \right) = \left[ \prod_{i \in I_3} \binom{m_i}{y_i} \right] \mathbb{E} \left( \int_{\mathbb{R}^p} I(W \beta \leq \tilde{U}) d\beta \right),
\]

16
where $\tilde{U} = (\tau_1 U_1, \tau_2 U_2, \ldots, \tau_{n+k} U_{n+k})^T$. Now using Lemma 1 from Appendix B, we have that

$$\int_{\mathbb{R}^p} f(y|\beta) d\beta \leq \left[ \prod_{i \in I_3} \left( \frac{m_i}{y_i} \right) \right] E\left( \int_{\mathbb{R}^p} I(\|\beta\| \leq C \|\tilde{U}\|) d\beta \right)$$

$$= C^* E\left( \|\tilde{U}\|^p \right)$$

$$\leq C^{**} \int |u|^p dF(u) < \infty,$$

where $C^*$ and $C^{**}$ are two appropriate constants depending only on $y$ and $X$. Since the condition $A3$ is in force, (4) holds. \square

**Proof of Corollary 2:**

**Proof.** Since

$$f(y|\beta) \leq \prod_{i \in I_3} \left( \frac{m_i}{y_i} \right) (F(\eta_i))(1 - F(\eta_i)),$$

we have

$$\int_{\mathbb{R}^p} f(y|\beta) d\beta \leq \left[ \prod_{i \in I_3} \left( \frac{m_i}{y_i} \right) \right] E\left( \int_{\mathbb{R}^p} I(W_1 \beta \leq \tilde{U}_1) d\beta \right),$$

where

$$W_1 = \begin{pmatrix} x_i^T, i \in I_3 \\ -x_i^T, i \in I_3 \end{pmatrix}.$$

and $\tilde{U}_1 = (U_1, U_2, \ldots, U_k, -U_{k+1}, \ldots, -U_{2k})$ are $2k$ iid random variables with common distribution function $F(\cdot)$. Recall that $k = |I_3|$ is the cardinality of $I_3$. Now the proof follows since for any positive vector $g^T = (g_1, g_2, \ldots, g_k) > 0$, we have $(g^T, g^T)W_1 = 0$. \square
Proof of Theorem 2:

Proof. If $X$ is not a full rank matrix, then $W$ is not full rank either. In this case it easily follows that
\[
\int_{\mathbb{R}^p} f(y|\beta) d\beta = \infty.
\]
Since $0 < F(0-)$, there exists a $\delta > 0$, such that $0 < F(-\delta) \leq F(\delta) < 1$. As in Chen and Shao (2000), we know that \{\(W^T g : g > 0, g \in \mathbb{R}^{n+k}\)\} is a convex cone in \(\mathbb{R}^p\). So if A2 doesn’t hold, by corollary 11.7.3 of Rockafellar (1970) there exists some nonzero vector $b \in \mathbb{R}^p$ such that $b^T W^T g \leq 0$ for all $g \geq 0$. In particular, we have
\[
\tau_i b^T x_i^+ \leq 0 \quad \text{for} \quad i = 1, 2, \ldots, n + k,
\]
that is $b^T x_i \leq 0$ for $i \in I_1$, $b^T x_i \geq 0$ for $i \in I_2$ and $b^T x_i = 0$ for $i \in I_3$. Consider the transformation $\beta \to s = (s_1, s_2, \ldots, s_p)$ where $\beta = s_1 b + (0, s_2, \ldots, s_p)$. Here without loss of generality we assume that $b_1 \neq 0$. Then
\[
\int_{\mathbb{R}^p} f(y|\beta) d\beta = \int_{\mathbb{R}^p} \prod_{i=1}^n \left( \frac{m_i}{y_i} \right) (F(\eta_i))^y_i (1 - F(\eta_i))^{m_i - y_i} d\beta
\]
\[
= |b_1| \prod_{i \in I_3} \left( \frac{m_i}{y_i} \right) \int_{\mathbb{R}^p} \prod_{i \in I_1} \left( 1 - F(s_1 x_i^T b + x_i^T (0, s_2, \ldots, s_p)) \right)^{m_i}
\]
\[
\times \prod_{i \in I_2} \left( 1 - F(0, s_2, \ldots, s_p) \right)^{y_i} (1 - F(x_i^T (0, s_2, \ldots, s_p)))^{m_i - y_i} ds
\]
\[
\geq |b_1| \prod_{i \in I_3} \left( \frac{m_i}{y_i} \right) \int_{s_1 \geq 0, |s_2| \leq \eta, 2 \leq j \leq p} \prod_{i \in I_1 \cup I_3} \left( 1 - F(p\eta ||x_i||) \right)^{m_i - y_i} \prod_{i \in I_2 \cup I_3} (F(-p\eta ||x_i||))^{y_i} ds
\]
\[
\geq |b_1| \prod_{i \in I_3} \left( \frac{m_i}{y_i} \right) \prod_{i \in I_1 \cup I_3} (1 - F(\delta))^{m_i - y_i} \prod_{i \in I_2 \cup I_3} (F(-\delta))^{y_i} \int_{s_1 \geq 0, |s_j| \leq \eta, 2 \leq j \leq p} ds
\]
\[
= \infty,
\]
where $\eta > 0$ is chosen such that $p\eta \max_{1 \leq i \leq n} ||x_i|| \leq \delta$. 

\[\square\]
Proof of Theorem 4:

Proof. Since \( \int_{\Lambda} \int_{\mathbb{R}^p} f(y|\beta, \lambda) d\beta \Pi(d\lambda) \geq \int_{\Lambda'} \int_{\mathbb{R}^p} f(y|\beta, \lambda) d\beta \Pi(d\lambda) \), from the proof of Theorem 2 it follows that

\[
\int_{\Lambda} \int_{\mathbb{R}^p} f(y|\beta, \lambda) d\beta \Pi(d\lambda) \geq |b_1| \left[ \prod_{i \in I_3} \left( m_i \right) \right] \prod_{i \in I_2} \left( 1 - \sup_{\lambda \in \Lambda'} F_\lambda(\delta) \right)^{m_i} \prod_{i \in I_1} \left( \inf_{\lambda \in \Lambda'} F_\lambda(-\delta) \right)^{m_i} \prod_{i \in I_3} \left( \inf_{\lambda \in \Lambda'} F_\lambda(-\delta) \right)^{m_i} \prod_{i \in I_2} \left( \sup_{\lambda \in \Lambda'} F_\lambda(\delta) \right)^{m_i} \prod_{i \in I_1} \left( 1 - \sup_{\lambda \in \Lambda'} F_\lambda(\delta) \right)^{m_i} \int_{\Lambda} \Pi(d\lambda) \int_{\Lambda'} s_1 \geq 0, |s_j| \leq \eta, 2 \leq j \leq p \]

where \( \eta \) is chosen as in the proof of Theorem 2.

\[ \Box \]

Proof of Theorem 5:

Proof. We need to show that \( \int_a^b m_H(\lambda) \Pi(d\lambda) < \infty \). Suppose, \( Z \sim H(\lambda) \). Then the pdf of \( Z \) is given by

\[
h_\lambda(z) = \begin{cases} 
\frac{e^z}{\{\lambda e^{z+1} +1\}^{1+\frac{1}{p}}} & -\infty < z < \infty \quad \text{if } \lambda > 0 \\
\frac{e^z}{\{\lambda e^{z+1} +1\}^{1+\frac{1}{p}}} & z < -\log(-\lambda) \quad \text{if } \lambda < 0 \\
e^{-e}e^z & -\infty < z < \infty \quad \text{if } \lambda = 0 
\end{cases}
\]

So

\[
E[Z|^p] = \begin{cases} 
\int_{-\infty}^{\infty} |z|^p \frac{e^z}{\{\lambda e^{z+1} +1\}^{1+\frac{1}{p}}} dz & \text{if } \lambda > 0 \\
\int_{-\infty}^{-\log(-\lambda)} |z|^p \frac{e^z}{\{\lambda e^{z+1} +1\}^{1+\frac{1}{p}}} dz & \text{if } \lambda < 0 \\
\int_{-\infty}^{\infty} |z|^p e^{-e}e^z dz & \text{if } \lambda = 0
\end{cases}
\]

Using the transformations \( v = (\lambda e^z +1)^{-\frac{1}{p}} \) when \( \lambda \neq 0 \) and \( v = e^{-e} \) when \( \lambda = 0 \), we have

\[
m_H(\lambda) = E[Z|^p] = \begin{cases} 
\int_0^1 \left| \log \left( \frac{v^{\lambda-1}}{\lambda} \right) \right|^p dv & \text{if } \lambda \neq 0 \\
\int_0^1 \left| \log(-\log(v)) \right|^p dv & \text{if } \lambda = 0
\end{cases}
\]

We first consider the case when \( p \) is even. If we define

\[
q(v, \lambda) = \begin{cases} 
\left( \log \left( \frac{v^{\lambda-1}}{\lambda} \right) \right)^p & \text{if } \lambda \neq 0 \\
\left( \log(-\log(v)) \right)^p & \text{if } \lambda = 0
\end{cases}
\]

then \( m_H(\lambda) = \int_0^1 q(v, \lambda) dv \). Suppose \( a < c < b \) and \( \lambda_n \in [a, b] \) is such that \( \lambda_n \to c \). From Proposition 1 in Appendix B
we know that
\[ q(v, \lambda_n) \leq q(v, a) + q(v, b) \]
for all \( \lambda_n \in [a, b] \). From Lemma 2 in Appendix B we know that the moment generating functions of \( H_\lambda(\cdot) \) exist in a neighborhood of 0, which implies that \( m_H(\lambda) \) is finite for all \( \lambda \). In particular, it shows that both \( m(a) = \int_0^1 q(v, a) dv \) and \( m(b) = \int_0^1 q(v, b) dv \) are finite. Then since \( q(v, \lambda_n) \to q(v, c) \) it follows by the Lebesgue’s dominated convergence theorem that \( m_H(\lambda) \) is a continuous function when \( p \) is an even number.

Now consider the case when \( p \) is an odd number. By Liapounov inequality we know that
\[
\int |u|^p dH_\lambda(u) \leq \left( \int |u|^{p+1} dH_\lambda(u) \right)^{\frac{p}{p+1}}.
\]
We will be done if we show that
\[
\int \left( \int |u|^{p+1} dF_\lambda(u) \right)^{\frac{p}{p+1}} d\Pi(\lambda) < \infty,
\]
but this is true since \( p+1 \) is an even number and using the above reasoning we know that \( \left( \int |u|^{p+1} dH_\lambda(u) \right)^{\frac{p}{p+1}} \) is a continuous function in \( \lambda \).

**Proof of theorem 6:**

**Proof.** We need to show that if \( b > -1 \), then we can find \( c, d \) and \( \delta > 0 \) with \( a \leq c < d \leq b \) such that
\[
\inf_{\lambda \in [c, d]} H_\lambda(-\delta) > 0 \quad \text{and} \quad \sup_{\lambda \in [c, d]} H_\lambda(\delta) < 1.
\]
From Lemma 3 in Appendix B we know that \( \{H_\lambda(\cdot)\} \) is a stochastically increasing family of distributions. So \( \inf_{\lambda \in [c, d]} H_\lambda(-\delta) = H_d(-\delta) \) and \( \sup_{\lambda \in [c, d]} H_\lambda(\delta) = H_c(\delta) \). Since \( b > -1 \), the proof is complete. \( \square \)

**B Preliminary results**

The results of this appendix are used in the proofs of the main results.

**Lemma 1** (Chen and Shao (2000)). Assume that the two conditions A1 and A2 of Theorem 1 are satisfied. Then there exists a constant, \( C \), depending only on \( W \) such that
\[
\| \beta \| \leq C \| \tilde{u} \|
\]
whenever

$$W \beta \leq \tilde{u},$$

where $\tilde{u}$ is a $(n + k) \times 1$ vector.

**Proposition 1.** Let $k$ be an even number. For fixed $y \in (0, 1)$ define

$$v_y(\lambda) = \begin{cases} 
(\log \left( \frac{y^{-\lambda} - 1}{\lambda} \right))^k & \text{if } \lambda \neq 0 \\
(\log(-\log y))^k & \text{if } \lambda = 0
\end{cases}.$$

If $\lambda \in [a, b]$, then $v_y(\lambda) \leq v_y(a) + v_y(b)$.

**Proof.** We first consider the case when $\lambda \neq 0$. If $\lambda \neq 0$, we know

$$v_y(\lambda) = \left( \log \left( \frac{y^{-\lambda} - 1}{\lambda} \right) \right)^k.$$

So,

$$\frac{dv_y(\lambda)}{d\lambda} = -k \left( \log \left( \frac{y^{-\lambda} - 1}{\lambda} \right) \right)^{k-1} \frac{\lambda y^{-\lambda} \log y + (y^{-\lambda} - 1)}{y^{-\lambda} - 1}.$$

Since

$$\lambda y^{-\lambda} \log y + (y^{-\lambda} - 1) = - \frac{y^{\lambda} - 1 - \lambda \log y}{y^{\lambda}},$$

and $y \in (0, 1)$ it follows from the inequality $\log x \leq x - 1$ for nonnegative $x$ that $\lambda y^{-\lambda} \log y + (y^{-\lambda} - 1) \leq 0$. Then since $\frac{y^{-\lambda} - 1}{\lambda} \geq 0$ and $k$ is an even number, we have

$$\frac{dv_y(\lambda)}{d\lambda} \geq (\leq) 0 \iff \frac{y^{-\lambda} - 1}{\lambda} \geq (\leq) 1. \quad (11)$$

Now from the above calculation we know that

$$\frac{d}{d\lambda} \left( \frac{y^{-\lambda} - 1}{\lambda} \right) = - \frac{\lambda y^{-\lambda} \log y + (y^{-\lambda} - 1)}{\lambda^2}$$

is nonnegative. So for fixed $y \in (0, 1)$, $\frac{y^{-\lambda} - 1}{\lambda}$ is a non-decreasing function in $\lambda$. Hence from (11) we have

$$\frac{dv_y(\lambda)}{d\lambda} \geq (\leq) 0 \iff \lambda \geq (\leq) y^*$, \quad (12)$$

where $y^*$ is the solution (for $\lambda$) of the equation $y^{-\lambda} - 1 = \lambda$. Finally since

$$\lim_{\lambda \to 0} v_y(\lambda) = \lim_{\lambda \to 0} \left( \log \left( \frac{y^{-\lambda} - 1}{\lambda} \right) \right)^k = (\log(-\log y))^k = v_y(0),$$

it follows from (12)

$$v_y(\lambda) \leq \max(v_y(a), v_y(b)) \text{ for all } \lambda \in [a, b].$$

Therefore the proof is complete since $v_y(\lambda) \geq 0$ for all $\lambda$. 

\[\square\]
Lemma 2. Suppose \( Z \sim H_\lambda(z) \). Then the mgf of \( Z \) is given by

\[
M(\theta) = E(e^{\theta Z}) = \begin{cases} 
\frac{1}{\lambda^{\theta+1}} B \left( \frac{1}{\lambda} - \theta, \theta + 1 \right) & \text{if } \lambda > 0 \\
\frac{1}{|\lambda|^{\theta+1}} B \left( -\frac{1}{\lambda}, \theta + 1 \right) & \text{if } \lambda < 0 \\
\Gamma(1 + \theta) & \text{if } \lambda = 0.
\end{cases}
\]

Proof. We calculate the mgf of \( Z \) separately for the cases \( \lambda > 0 \), \( \lambda < 0 \) and \( \lambda = 0 \).

Case I \( \lambda > 0 \)

Since \( h_\lambda(z) \) is the pdf of \( Z \), we have

\[
M(\theta) = \int_{-\infty}^{\infty} e^{\theta z} h_\lambda(z) dz = \int_{-\infty}^{\infty} e^{\theta z} \frac{e^z}{\lambda e^z + 1}^{1+1/\lambda} dz
\]

Using the transformation \( y = (\lambda e^z + 1)^{-\frac{1}{\lambda}} \), it follows that

\[
M(\theta) = \frac{1}{\lambda^\theta} \int_{0}^{1} (1 - y^\lambda)^\theta y^{-\lambda \theta} dy = \frac{1}{\lambda^\theta+1} \int_{0}^{1} (1 - w)^\theta w^{-\theta - \frac{1}{\lambda} - 1} dw \quad (w = y^\lambda)
\]

Hence in this case the mgf \( M(\theta) \) exists if \( \theta \) lies in the following neighborhood of 0

\[-1 < \theta < \frac{1}{\lambda}.\]

Case II \( \lambda < 0 \)

In this case we know

\[
M(\theta) = \int_{-\infty}^{-\log(-\lambda)} e^{\theta z} \frac{e^z}{(\lambda e^z + 1)^{1+1/\lambda}} dz.
\]

Here again taking the transformation \( y = (\lambda e^z + 1)^{-\frac{1}{\lambda}} \) we have

\[
M(\theta) = \frac{1}{\lambda^\theta} \int_{0}^{1} (1 - y^\lambda)^\theta y^{-\lambda \theta} dy = \frac{(-1)^\theta}{\lambda^{\theta+1}} \int_{0}^{1} (1 - w)^\theta w^{-\frac{1}{\lambda} - 1} dw
\]

\[
= \frac{1}{|\lambda|^{\theta+1}} B \left( -\frac{1}{\lambda}, \theta + 1 \right).
\]
Hence, in this case the mgf $M(\theta)$ exists as long as $\theta > -1$.

**Case III $\lambda = 0$**

Note that, if $Z \sim H_0(\cdot)$ then $-Z$ follows the standard Gumbel distribution. So we know that $E(e^{-\theta Z}) = \Gamma(1 - \theta)$ and hence the moment generating function of $H_0(\cdot)$ is $M(\theta) = E(e^{\theta Z}) = \Gamma(1 + \theta)$.

**Lemma 3.** The family of distribution functions $\{H_\lambda(\cdot)\}$ is stochastically increasing.

**Proof.** We know that a distribution function $H_\lambda(\cdot)$ is stochastically larger than a distribution function $H_{\lambda'}(\cdot)$ if $H_\lambda(t) \leq H_{\lambda'}(t)$ for all $t$. In order to prove that $\{H_\lambda(\cdot)\}$ is stochastically increasing family, we show that for given $t$, $H_\lambda(t)$ is a strictly decreasing function in $\lambda$.

We first show that for given $t$, $H_\lambda(t)$ is a decreasing function in $\lambda$ on both negative and positive half-line i.e., when $\lambda < 0$ or $\lambda > 0$. Then using mean value theorem, we will prove that $H_\lambda(t)$ is a decreasing function in $\lambda$ on the entire real line. Note that, when $\lambda \neq 0$, $H_\lambda(t)$ is given by

$$H_\lambda(t) = 1 - \frac{1}{\{\lambda e^t + 1\}^\frac{1}{\lambda}}.$$ 

So, $H_\lambda(t)$ is decreasing in $\lambda$ if and only if $g_t(\lambda) := \{\lambda e^t + 1\}^{1/\lambda}$ is decreasing in $\lambda$. Now,

$$\log g_t(\lambda) = \frac{1}{\lambda} \log(\lambda e^t + 1).$$

So,

$$g_t'(\lambda) \over g_t(\lambda) = \frac{\lambda e^t}{\lambda e^t + 1} - \log(\lambda e^t + 1)$$

$$= \frac{\{(\lambda e^t + 1) - (\lambda e^t + 1) \log(\lambda e^t + 1)\} - 1}{\lambda^2(\lambda e^t + 1)}$$

Notice that $\lambda e^t + 1$ when $t$ lies in the support of $H_\lambda(t)$. (Obviously, $\lambda e^t + 1 > 0$ if $\lambda > 0$. If $\lambda < 0$ we know that $t < -\log(-\lambda)$ i.e, $\lambda e^t + 1 > 0$.) So, $g_t(\lambda)$ is also positive. Then from the inequality $x - x \log x \leq 1$ for nonnegative $x$, it follows that $g_t'(\lambda) < 0$ if $\lambda > 0$ or $\lambda < 0$. So, for fixed $t$, $H_\lambda(t)$ is a strictly decreasing function of $\lambda$ in the region $\{\lambda : \lambda \neq 0\}$.

Now to show that $H_\lambda(t)$ is a decreasing function in $\lambda$ on the entire real line, first note that for $\lambda \neq 0$, 

$$\frac{d}{d\lambda} H_\lambda(t) \left( = \frac{g_t'(\lambda)}{(g_t(\lambda))^2} \right) < 0.$$ 

Then, since $H_\lambda(t)$ is a continuous function in $\lambda$ on the entire real line, an application of mean value theorem shows that

$$H_\lambda(t) < H_0(t) \text{ if } \lambda > 0$$

23
and
\[ H_\lambda(t) > H_0(t) \text{ if } \lambda < 0. \]

Hence, we have \( H_\lambda(t) < H_{\lambda'}(t) \) when \( \lambda > \lambda' \).

Acknowledgments

The authors thank the editor and a reviewer for helpful comments and valuable suggestions which led to several improvements in the manuscript.

References


