Opuscula Mathematica • Vol. 32 • No. 4 • 2012

http://dx.doi.org/10.7494/OpMath.2012.32.4.761

ON THE MAXIMUM LIKELIHOOD ESTIMATOR IN THE GENERALIZED BETA REGRESSION MODEL

Jerzy P. Rydlewski and Dominik Mielczarek

Abstract. The subject of this article is to present the beta – regression model, where we assume that one parameter in the model is described as a combination of algebraically independent continuous functions. The proposed beta model is useful when the dependent variable is continuous and restricted to the bounded interval. The parameters are obtained by maximum likelihood estimation. We prove that estimators are consistent and asymptotically normal.

Keywords: nonlinear regression, beta distribution, scale parameter, shape parameter, maximum likelihood estimator.

Mathematics Subject Classification: 62J02, 62F10.

1. INTRODUCTION

It is often encountered in practice that the dependent variable takes values only from a finite interval (see [1] and [4]). In this paper we deal with this case as we examine the problem of maximum likelihood estimation for the beta distributed regression model. The maximum likelihood estimators for beta models have been recently studied in Ferrari and Cribari–Neto [1], Cribari–Neto and Vasconcellos [3], Rydlewski [4, 5] and Souza *et al.* [7]. The linear regression model is widely used in applications to analyze data that are considered to be related to other variables. It should not be used in models where the dependent variable is restricted to the interval $(0, 1)$. Moreover, this paper analyses the problems where variables may be time dependent. The dependance on time might be described by a cyclic function, not by a linear function.

Our results are not covered by Wei's monograph on exponential family nonlinear models (see [9] pp. 2–3).

Maximum likelihood estimation for different nonlinear models, as well as references to the relevant literature, are given by Seber and Wild [6].

The Generalized Linear Model applied to beta regression is widely discussed in [3]. However, in [3] the authors do not prove that there exists exactly one maximum likelihood estimator in the model. The application of small sample bias adjustments to the maximum likelihood estimators of these parameters is discussed in [1].

The paper is organised as follows. In Section 2, we describe a model. Section 3 contains results on the existence and uniqueness of the maximum likelihood estimator. In Section 4, we discuss the asymptotics of the model. Section 5 contains some computational aspects of the developed theory.

2. MODEL DEFINITION

The proposed model is based on an assumption that the dependent data is beta distributed. The beta density is given by

$$\pi(y,p,q) = \frac{1}{B(p,q)} y^{p-1}(1-y)^{q-1}, \quad 0 < y < 1, \tag{2.1}$$

where $p > 0$, $q > 0$ and $B(\cdot,\cdot)$ is the beta function. The mean and the variance of y are, respectively,

$$E(y) = \frac{p}{p+q} \tag{2.2}$$

and

$$Var(y) = \frac{pq}{(p+q)^2(p+q+1)}. \tag{2.3}$$

It is also common to define the above model in terms of precision parameters. Hence, if we define $\varphi = p/(p+q)$ and $\phi = p+q$, then we have

$$E(y) = \varphi \quad \text{and} \quad Var(y) = \frac{\varphi(1-\varphi)}{1+\phi}.$$

In this parametrization the beta density is given by

$$\pi(y,\varphi,\phi) = \frac{1}{B(\varphi\phi,(1-\varphi)\phi)} y^{\varphi\phi-1}(1-y)^{(1-\varphi)\phi-1}, \quad 0 < y < 1, \tag{2.4}$$

where $0 < \varphi < 1$ and $\phi > 0$.

Let $y_1,\ldots,y_n$ be independent, beta-distributed random variables observed at the design points $t_1,\ldots,t_n$. We have the following model of the mean

$$E(y_j) = \varphi(t_j),$$

and the variance

$$Var(y_j) = \frac{\varphi(t_j)(1-\varphi(t_j))}{1+\phi}, \quad j = 1,\ldots,n,$$

where φ is a sum of algebraically independent continuous functions. The t_j's may be interpreted as time points. The unknown precision parameter ϕ is independent of t_j.

The model is useful if the response is restricted to the interval (a, b), where $a < b$, because we could then model $(y - a)/(b - a)$.

We consider the special case of the model where

$$0 \leq y_j \leq 1, \quad j = 1, \ldots, n, \tag{2.5}$$

and

$$\varphi(t_j) = K\theta(t_j) + M\left(\beta_0 + \sum_{k=1}^{p} \alpha_k \sin \frac{2\pi k}{T} t_j + \beta_k \cos \frac{2\pi k}{T} t_j\right),$$

where functions of t_j are continuous and algebraically independent and we assume that $0 \leq \varphi(t_j) \leq 1$ for $j = 1, \ldots, n$. The function $\theta(\cdot)$ is responsible for the modelling of the trend. Such a model has been developed in our previous work [4].

In our model we substitute for φ a function of variable t_i, which depends on the multidimensional parameter A. Precisely, we use a set of m algebraically independent continuous functions. Consider

$$p(A, t) = \sum_{k=1}^{m} A_k f_k(t), \tag{2.6}$$

where $A = (A_1, \ldots, A_m)$ and $p(A, t)$ is modelling the parameter p of the beta distribution for each of the observations. Let $y_1, \ldots, y_n$ be independent beta distributed random variables, where each variable has the p parameter in equation (2.1) given by the expression in (2.6). Let $E(y_j) = \varphi(t_j)$ for $j = 1, \ldots, n$. Then, it follows from the aforementioned assumptions that the expected value of y_j has the following form

$$E(y_j) = \sum_{k=1}^{m} \alpha_k f_k(t_j), \qquad j = 1, \ldots, n,$$

with $\alpha_k = A_k/\phi$, where $0 \leq \varphi(t_j) \leq 1$. Absolute continuity of the random variables guarantees that $0 < \varphi(t_j) < 1$, a.e. with respect to the Lebesgue measure.

In order to model the parameter $p = \phi\varphi$ we take m algebraically independent continuous functions.

To estimate the parameters we will use the maximum likelihood estimation. In our beta regression model, the likelihood function has the form

$$\begin{aligned} &L(y_1, \ldots, y_n, t_1, \ldots, t_n, A, \phi) = \\ &= \prod_{j=1}^{n} \frac{1}{B(p(A, t_j), \phi - p(A, t_j))} y_j^{p(A,t_j)-1} (1 - y_j)^{\phi - p(A,t_j)-1}, \end{aligned} \tag{2.7}$$

where $B(\cdot, \cdot)$ is the beta function. Hence, the logarithm of the likelihood function is

$$\begin{aligned} \log L &= \sum_{j=1}^{n} \log L_j = \\ &= \sum_{j=1}^{n} -\log B(p(A, t_j), \phi - p(A, t_j)) + (p(A, t_j) - 1) \log y_j + (\phi - p(A, t_j) - 1) \log(1 - y_j). \end{aligned} \tag{2.8}$$

We can carry out a reparametrization analogous to the one described at the beginning of this section. Let $a = (\alpha_1, \ldots, \alpha_m)$. Then,

$$\varphi(a, t_j) = \frac{p(A, t_j)}{\phi} = \sum_{k=1}^{m} \alpha_k f_k(t_j).$$

Obviously

$$\log L(\phi a, \phi) = \sum_{j=1}^{n} \log L_j(\phi a, \phi),$$

where

$$\log L_j(\phi a, \phi) = -\log B(\phi\varphi(a, t_j), \phi(1 - \varphi(a, t_j)))+ \\ + (\phi\varphi(a, t_j) - 1)\log y_j + (\phi(1 - \varphi(a, t_j)) - 1)\log(1 - y_j).$$

Later on, we will prove that the maximum likelihood estimator is determined uniquely. Ferrari and Cribari-Neto [3] defined a regression structure for beta distributed responses that differs from (2.4). Our model is equivalent to the model developed by Ferrari and Cribari-Neto [3] through reparametrization

$$(A_1, \ldots, A_m, \phi) \leftrightarrow \left(\frac{A_1}{\phi}, \ldots, \frac{A_m}{\phi}, \phi\right).$$

However, in [3] the proof of uniqueness and existence of the maximum likelihood estimator has not been established. Furthermore, the results established in [3] does not include explicit proofs of consistency and asymptotic normality. The proofs are also absent in [7].

3. MAXIMUM LIKELIHOOD ESTIMATION

Lemma 3.1. *The function* $\log B(x, y)$ *is a strongly convex function of* x *and* y.

Proof. Let $x_1, x_2, y_1, y_2 > 0$ and let $\lambda \in (0, 1)$. We obtain

$$B(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2) = \int_0^1 s^{\lambda x_1 + (1-\lambda)x_2 - 1}(1-s)^{\lambda y_1 + (1-\lambda)y_2 - 1} ds = \\ = \int_0^1 \left[s^{x_1-1}(1-s)^{y_1-1}\right]^{\lambda} \left[s^{x_2-1}(1-s)^{y_2-1}\right]^{1-\lambda} ds.$$

In light of the Hölder inequality with Hölder conjugates $\frac{1}{\lambda}$ and $\frac{1}{1-\lambda}$ we get

$$B(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2) \leq \\ \leq \left[\int_0^1 s^{x_1-1}(1-s)^{y_1-1} ds\right]^{\lambda} \left[\int_0^1 s^{x_2-1}(1-s)^{y_2-1} ds\right]^{1-\lambda}.$$

Thus

$$B\left(\lambda x_1+(1-\lambda)x_2,\lambda y_1+(1-\lambda)y_2\right)\le B\left(x_1,y_1\right)^{\lambda}B\left(x_2,y_2\right)^{1-\lambda}$$

and finally we have

$$\log B(\lambda x_1+(1-\lambda)x_2,\lambda y_1+(1-\lambda)y_2)\le\lambda\log B(x_1,y_1)+(1-\lambda)\log B(x_2,y_2). \quad (3.1)$$

Equality in (3.1) holds if and only if there exists a nonzero constant $c\in\mathbb{R}$ such that

$$s^{x_2-1}(1-s)^{y_2-1}=c\left(s^{x_1-1}(1-s)^{y_1-1}\right)$$

for $s\in[0,1]$. In other words, the equality is valid if and only if $x_1=x_2$ and $y_1=y_2$. The proof of the lemma is complete. □

Lemma 3.2. *Let $[c,d]$ be a closed and bounded interval. Let $f_1,\dots,f_m$ be a set of continuous algebraically independent functions on $[c,d]$. The set of all parameters $\boldsymbol{A}$ of the form $(A_1,\dots,A_m,\phi)\in\mathbb{R}^{m+1}$ such that for any $t\in[c,d]$,*

$$0\le\sum_{k=1}^{m}A_kf_k(t)\le M\phi \quad (3.2)$$

is non-empty, closed and convex in $\mathbb{R}^{m+1}$.

Proof. Proof of this lemma can be found in [5]. □

Lemma 3.3. *Let $f_1,\dots,f_m:\mathbb{R}\longrightarrow\mathbb{R}$ be a set of algebraically independent continuous functions. The set* **a** *of all parameters $a=(\alpha_1,\dots,\alpha_m)\in\mathbb{R}^m$ satisfying the equation*

$$0\le\sum_{k=1}^{m}\alpha_kf_k(t)\le M \quad (3.3)$$

for any $t\in\mathbb{R}$ is non-empty and compact in $\mathbb{R}^m$.

Proof. Proof of this lemma can be found in [5] as well. □

The proof of the next lemma is similar in spirit to that of Lemma 3.3 for the gamma regression model in [5]. Although these proofs run along similar lines, there are subtle adjustments necessary to fit the argument to each new situation.

Lemma 3.4. *Exactly one of the conditions specified below is true*

(i) *for all $j=1,\dots,n$*

$$y_j=\sum_{k=1}^{m}\alpha_kf_k(t_j);$$

(ii)

$$\lim_{\phi\to+\infty}\frac{d}{d\phi}\log L(\phi a,\phi)<0.$$

Proof. In the beta regression model we have

$$\begin{aligned}\lim_{\phi\to\infty}\frac{d}{d\phi}\log L(\phi a,\phi) &= \lim_{\phi\to\infty}\frac{d}{d\phi}\sum_{j=1}^{n}\log L_j(\phi a,\phi) = \\ &= \lim_{\phi\to\infty}\frac{d}{d\phi}\sum_{j=1}^{n}(-\log B(\phi\varphi(a,t_j),\phi(1-\varphi(a,t_j)))+ \\ &\quad + (\phi\varphi(a,t_j)-1)\log y_j + (\phi(1-\varphi(a,t_j))-1)\log(1-y_j)).\end{aligned}$$

Let $w_j = \varphi(a,t_j)$. Consequently

$$\log L_j(\phi a,\phi) = -\log B(\phi w_j,\phi(1-w_j)) + (\phi w_j - 1)\log y_j + (\phi(1-w_j)-1)\log(1-y_j).$$

We take advantage of the B function property in order to get

$$\begin{aligned}\log L_j(\phi a,\phi) = &-\log\Gamma(\phi w_j) - \log\Gamma(\phi(1-w_j)) + \log\Gamma(\phi)+ \\ &+ (\phi w_j - 1)\log y_j + (\phi(1-w_j)-1)\log(1-y_j).\end{aligned}$$

Hence

$$\begin{aligned}\frac{d}{d\phi}\log L_j(\phi a,\phi) &= -\frac{d}{d\phi}\log\Gamma(\phi w_j) - \frac{d}{d\phi}\log\Gamma(\phi(1-w_j))+ \\ &\quad + \frac{d}{d\phi}\log\Gamma(\phi) + w_j\log y_j + (1-w_j)\log(1-y_j) = \\ &= -w_j\Psi(\phi w_j) - (1-w_j)\Psi(\phi(1-w_j))+ \\ &\quad + \Psi(\phi) + w_j\log y_j + (1-w_j)\log(1-y_j),\end{aligned}$$

where $\Psi(y) = \frac{d}{dy}\log\Gamma(y)$. Using the well–known equality

$$\lim_{y\to\infty}(\Psi(y) - \log y) = 0,$$

we obtain

$$\lim_{\phi\to\infty}\frac{d}{d\phi}\log L_j(\phi a,\phi) = -w_j\log w_j - (1-w_j)\log(1-w_j) + w_j\log y_j + (1-w_j)\log(1-y_j).$$

Define

$$h(y) = w_j\log y + (1-w_j)\log(1-y).$$

The function h takes the largest value at $y = w_j$, and hence

$$\lim_{\phi\to\infty}\sum_{j=1}^{n}\frac{d}{d\phi}\log L_j(\phi a,\phi) \le 0$$

and if for at least one $j \in \{1,\ldots,n\}$ we have $y_j \ne w_j$, then we obtain

$$\lim_{\phi\to\infty}\sum_{j=1}^{n}\frac{d}{d\phi}\log L_j(\phi a,\phi) < 0.$$

□

Lemma 3.5. *The function* $\log L(\phi a, \phi)$ *as a function of the parameter* ϕ *is strictly concave.*

Proof. It suffices to show that function

$$\log B\left(\phi\varphi(a,t_j), \phi(1-\varphi(a,t_j))\right)$$

is strongly convex for at least one $j \in \{1,\dots,n\}$. Strong convexity now follows from the definition of convexity and the Hölder inequality, like that of the proof of Lemma 3.1 □

Let $J \in \mathbb{R}^{n\times m}$ be a real matrix

$$\begin{bmatrix} f_1(t_1) & \dots & \dots & \dots & f_m(t_1) \\ & & & & \\ \dots & \dots & & & \dots \\ \dots & & \dots & & \dots \\ \dots & & & \dots & \dots \\ & & & & \\ f_1(t_n) & \dots & \dots & \dots & f_m(t_n) \end{bmatrix}. \tag{3.4}$$

Parameters A_k and A_l are defined to be orthogonal if the (A_k, A_l) component of the Fisher information matrix is zero. Note that if the functions f_k and f_l are orthogonal, in the sense that $f_k(t_j)f_l(t_j) = 0$ if and only if $k \neq l$ for $j = 1,\dots,n$, then the parameters A_k and A_l are orthogonal.

Lemma 3.6. *If* n*–the number of observations is sufficient, i.e.,* $n \geq m$*, and the rank of the matrix* J *is maximal, i.e.,* $\mathrm{rank}J = m$*, then the function* $\log L(A_1,\dots,A_m,\phi)$ *is strictly concave.*

Proof. It follows from Lemma 3.1, that the function $\log B(x,y)$ is strictly convex. The function

$$\log L(A_1,\dots,A_m,\phi) =$$
$$= \sum_{j=1}^{n} -\log B(p(A,t_j), \phi - p(A,t_j)) + (p(A,t_j)-1)\log y_j + (\phi - p(A,t_j)-1)\log(1-y_j)$$

is concave as the sum of concave and linear functions. As we have assumed, that the rank of the matrix J is maximal, the intersection of all hyperplanes

$$p(A,t_j) = p(A_1,\dots,A_m,t_j) = const$$

is at most a single point. Following the same lines as the last paragraph of the proof of Lemma 3.1, we find out that for the above fact it suffices to prove that the $\log L(A_1,\dots,A_m,\phi)$ function is strictly concave. □

Theorem 3.7. *Let $n \geq m$ and let for a given $t_1, \ldots, t_n \in [c,d]$ the rank of matrix J defined in* (3.4) *be maximal. Then for given $t_1, \ldots, t_n \in [c,d]$ with probability* 1 *there exists exactly one $(\widehat{A}, \widehat{\phi}) \in \boldsymbol{A}$ such that*

$$L(\widehat{A}, \widehat{\phi}) = \max_{(A,\phi)\in \boldsymbol{A}} L(A, \phi),$$

where L is the likelihood function defined in (2.7).

Proof. The proof proceeds along the same lines as the proof of Theorem 3.7 in [5]. □

4. CONSISTENCY AND ASYMPTOTIC NORMALITY

Let $\theta \in \Theta$ be now a set of parameters. $J_n(\theta)$ is the observed information matrix at θ. We should make some assumptions.

(AN 1) Let the true parameter $\theta_0 \in \text{int}\, \Theta$.

(AN 2) Let

$$\frac{1}{n} J_n(\theta) \to K(\theta),\ n \to +\infty,$$

uniformly in $\overline{N}(\delta)$, where $\overline{N}(\delta)$ is a neighbourhood of the true θ with radius δ and $K(\theta)$ is some positive definite matrix. The assumption is widely accepted [2, 9].

Theorem 4.1. *Under* (AN 1) *and* (AN 2)*, the maximum likelihood estimator in the beta–regression model is strongly consistent.*

Proof. We show that

$$\exists \delta^* > 0 \forall \delta \in (0, \delta^*]\, \exists n^* \in \mathbb{N} \forall n > n^* \forall \theta \in \partial N(\delta) : P\left(l_n(\theta) - l_n(\theta_0) < 0\right) = 1, \quad (4.1)$$

where $l_n(\cdot)$ is the log-likelihood for n observations, $N(\delta) = \{\theta : \|\theta - \theta_0\| < \delta\}$, $\partial N(\delta) = \{\theta : \|\theta - \theta_0\| = \delta\}$ and $\overline{N}(\delta) = \{\theta : \|\theta - \theta_0\| \leq \delta\}$.

It means that $\widehat{\theta}_n$, that maximizes $l_n(\theta)$ must be inside $\overline{N}(\delta)$. Because $\delta \leq \delta^*$ and δ is arbitrarily small, we get $\widehat{\theta}_n \to \theta_0$ a.e.

Let $\lambda = \frac{\theta - \theta_0}{\delta}$ and Taylor theorem gives us that

$$l_n(\theta) - l_n(\theta_0) = \delta \lambda^T l_n'(\theta_0) + \frac{1}{2}\delta^2 \lambda^T l_n''(\widehat{\theta}_n)\lambda,$$

for some $\widehat{t}_n \in [0,1]$ is $\widehat{\theta}_n = \widehat{t}_n \theta_0 + (1 - \widehat{t}_n)\theta$. (4.1) is equivalent to

$$\exists \delta^* > 0 \forall \delta \in (0, \delta^*]\, \exists n^* \in \mathbb{N} \forall n > n^* \forall \theta \in \partial N(\delta) :$$

$$P\left(\frac{1}{n}\lambda^T l_n'(\theta_0) < \frac{1}{2n}\delta \lambda^T \left(-l_n''(\widehat{\theta}_n)\right)\lambda\right) = 1. \quad (4.2)$$

Let J_n be the Fisher information matrix from n observations. Define $R_n(\widehat{\theta}_n) = l_n''(\widehat{\theta}_n) + J_n(\widehat{\theta}_n)$. In our model $R_n(\widehat{\theta}_n) = 0$ and consequently we obtain

$$\exists \delta^* > 0 \forall \delta \in (0, \delta^*]\, \exists n^* \in \mathbb{N} \forall n > n^* \forall \theta \in \partial N(\delta):$$

$$P\left(\frac{1}{n}\lambda^T l_n'(\theta_0) < \frac{1}{2n}\delta\lambda^T J_n(\widehat{\theta}_n)\lambda\right) = 1.$$

Function $J_n(\theta)$ is continuous in θ_0 and assumption (AN 2) gives us

$$\forall \epsilon > 0 \exists \delta_1 > 0 \exists n_1 \in \mathbb{N} \forall \theta \in N(\delta_1) \forall n > n_1:$$

$$\left|\frac{1}{n}\lambda^T J_n(\theta)\lambda - \lambda^T K(\theta_0)\lambda\right| \leq \left|\frac{1}{n}\lambda^T J_n(\theta)\lambda - \lambda^T K(\theta)\lambda\right| + \left|\lambda^T K(\theta)\lambda - \lambda^T K(\theta_0)\lambda\right| < \epsilon.$$

Let $c = \lambda_{\min}(K(\theta_0))$ be the smallest eigenvalue of matrix $K(\theta_0)$. From inequality $\lambda_{\min}(K(\theta_0)) \leq \lambda^T K(\theta_0)\lambda$, for $\lambda^T\lambda = 1$, we obtain

$$\text{(i)} \quad \forall \epsilon > 0 \exists \delta_1 > 0 \exists n_1 \in \mathbb{N} \forall \theta \in N(\delta_1) \forall n > n_1: \frac{1}{n}\lambda^T J_n(\theta)\lambda > c - \epsilon \text{ a.s.}$$

We need to prove that (ii) $\lim_{n\to+\infty} \frac{1}{n}\lambda^T l_n'(\theta_0) \to 0$ a.s., for any λ, $\lambda^T\lambda = 1$.

It is true due to the Strong Law of Large Numbers and inequality

$$\left(\lambda^T l_n'(\theta_0)\right)^2 \leq (\lambda^T\lambda)(l_n'^T l_n') = \|\lambda\|^2 \|l_n'\|^2.$$

From (i) and (ii) we obtain

$$\forall \epsilon > 0\, \exists \delta^* > 0\, \exists n_2 \in \mathbb{N}\, \forall \theta \in N(\delta^*)\, \forall n > n_2: \frac{1}{n}\lambda^T\left(-l''(\theta)\right)\lambda = \frac{1}{n}\lambda^T J_n(\theta)\lambda > c - \epsilon \text{ a.e.}$$

Now we have $\widehat{\theta}_n \in N(\delta^*)$, for $n > n_2$. (ii) gives us that

$$\forall \epsilon > 0 \exists \delta^* > 0 \forall \delta \in (0, \delta^*]\, \exists n^* \geq n_2 \forall \theta \in N(\delta) \forall n > n^*: \frac{1}{n}\lambda^T l_n'(\theta_0) < \frac{1}{2}\delta(c - \epsilon) \text{ a.e.}$$

which completes the proof. □

We will concern asymptotic normality.

(AN 3) Let $\delta > 0$

$$\max_{\theta \in U_n(\delta)} \|V_n(\theta) - I\| \to 0,$$

where $V_n(\theta) = -J_n^{-\frac{1}{2}}(\theta_0) l_n''(\theta) J_n^{-\frac{1}{2}}(\theta_0)$ and
$U_n(\delta) = \left\{\theta \in \Theta : \|J_n^{\frac{1}{2}}(\theta_0)(\theta - \theta_0) \leq \delta\|\right\}$ for $n \in \mathbb{N}$.

The following lemma is an adaptation of Lemma 1 in [2].

Lemma 4.2. *Under assumptions* (AN 1)–(AN 3)

$$J_n^{-\frac{1}{2}} l_n' \xrightarrow{D} \mathcal{N}(0, I),\ n \to +\infty,$$

where $J_n = J_n(\theta_0)$, $l_n = l_n'(\theta_0)$ *and symbol* $\xrightarrow{D}$ *means convergence in the distribution.*

Proof. We will show that the moment generating function of the random variable $\lambda^T J_n^{-\frac{1}{2}} l_n'$ converges to a standard normal distribution moment generating function if $\lambda^T \lambda = 1$. Let us fix $\delta > 0$ and λ, $\lambda^T \lambda = 1$. If $\theta_n = \theta_0 + \delta J_n^{-\frac{1}{2}} \lambda$, $n \in \mathbb{N}$, then $\theta_n \in U_n(\delta)$. The Taylor expansion of the log likelihood becomes

$$l_n(\theta_n) = l_n(\theta_0) + (\theta_n - \theta_0)^T l_n' + \frac{1}{2}(\theta_n - \theta_0)^T l_n''(\overline{\theta}_n)(\theta_n - \theta_0),$$

where $\overline{\theta}_n = t_n \theta_n + (1 - t_n)\theta_0$ for $t_n \in [0, 1]$. We get

$$\exp\left\{\lambda^T J_n^{-\frac{1}{2}} \left(-l_n''(\overline{\theta}_n)\right) J_n^{-\frac{1}{2}} \lambda \frac{\delta^2}{2}\right\} L_n(\theta_n) = \exp\left\{\delta \lambda^T J_n^{-\frac{1}{2}} l_n'\right\} L_n(\theta_0),$$

and

$$\exp\left\{\lambda^T V_n(\overline{\theta}_n)\lambda \frac{\delta^2}{2}\right\} L_n(\theta_n) = \exp\left\{\delta \lambda^T J_n^{-\frac{1}{2}} l_n'\right\} L_n(\theta_0),$$

where L_n is a likelihood function. The left side of the equality is integrable, because $\exp\left\{\lambda^T V_n(\theta)\lambda \frac{\delta^2}{2}\right\}$ is a continuous function with respect to θ, and so it is bounded on a compact line segment $[\theta_0, \theta_n]$. This upper bound is not a function of $y_1, \ldots, y_n$, because $J_n(\theta) = -l_n''(\theta)$. Integrating the identity we obtain

$$E_{\theta_n} \exp\left\{\lambda^T V_n(\overline{\theta}_n)\lambda \frac{\delta^2}{2}\right\} = E \exp\left\{\delta \lambda^T J_n^{-\frac{1}{2}} l_n'\right\}.$$

Assumption (AN 3) and $\overline{\theta}_n \in U_n(\delta)$ give us

$\forall \epsilon > 0 \exists n_1 \in \mathbb{N} \forall n \geq n_1 :$

$$\left|\exp\left\{\lambda^T V_n(\overline{\theta}_n)\lambda \frac{\delta^2}{2}\right\} - \exp\left\{\frac{\lambda^T \lambda I \delta^2}{2}\right\}\right| = \left|\exp\left\{\lambda^T V_n(\overline{\theta}_n)\lambda \frac{\delta^2}{2}\right\} - \exp\left\{\frac{\delta^2}{2}\right\}\right| \leq \epsilon.$$

Integrating the inequality, we obtain by the dominated convergence theorem

$$E_{\theta_n} \exp\left\{\lambda^T V_n(\overline{\theta}_n)\lambda \frac{\delta^2}{2}\right\} \to E \exp\left\{\frac{\delta^2}{2}\right\},\ n \to \infty.$$

Therefore

$$E \exp\left\{\delta \lambda^T J_n^{-\frac{1}{2}} l_n'\right\} \to E \exp\left\{\frac{\delta^2}{2}\right\},\ n \to \infty,$$

when $\lambda^T \lambda = 1$. The theorem on the continuity of moment generating functions ensure us that $\lambda J_n^{-\frac{1}{2}} l_n'$, $\lambda \in \mathbb{R}^{m+1}$ is asymptotically standard gaussian. Because λ is arbitrary such that $\|\lambda\| = 1$, the proof is completed. □

Theorem 4.3. *Under* (AN 1), (AN 2) *and* (AN 3), *the maximum likelihood estimator in the beta–regression model is asymptotically normal.*

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \xrightarrow{D} \mathcal{N}\left(0, K^{-1}(\theta_0)\right).$$

Proof. Taylor's expansion of $l_n'(\theta_0)$ around $\widehat{\theta}_n$, where θ_0 is the true parameter and $\widehat{\theta}_n$ is a solution of the likelihood equations when the number of observations is n. We obtain

$$l_n'(\theta_0) = l_n'(\widehat{\theta}_n) + l''(\theta_n^*)(\theta_0 - \widehat{\theta}_n) = -l_n''(\theta_n^*)(\widehat{\theta}_n - \theta_0),$$

where $\theta_n^* = t_n^*\theta_0 + (1 - t_n^*)\widehat{\theta}_n$ for some $t_n^* \in [0, 1]$. We have also $\theta_n^* \to \theta_0$, $n \to +\infty$, because $\widehat{\theta}_n \to \theta_0$. Rewrite our equation

$$J_n^{-\frac{1}{2}} l_n' = J_n^{-\frac{1}{2}} \left\{-l_n''(\theta_n^*)\right\} J_n^{-\frac{1}{2}} J_n^{\frac{1}{2}} (\widehat{\theta}_n - \theta_0).$$

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) = \left(\frac{1}{n}J_n\right)^{-\frac{1}{2}} G_n^{-1} J_n^{-\frac{1}{2}} l_n', \tag{4.3}$$

where

$$G_n = J_n^{-\frac{1}{2}} \left\{-l_n''(\theta_n^*)\right\} J_n^{-\frac{1}{2}}.$$

We shall prove that $G_n \to I_{m+1}$ a.e., $n \to +\infty$. We already have

$$G_n = \left(\frac{1}{n}J_n\right)^{-\frac{1}{2}} \left\{\frac{1}{n}J_n(\theta_n^*) - \frac{1}{n}R_n(\theta_n^*)\right\} \left(\frac{1}{n}J_n\right)^{-\frac{1}{2}},$$

where matrix $R_n(\cdot) = 0$ as in the proof of Theorem 3.7 From assumption (AN 2)

$$\frac{1}{n}J_n \to K(\theta_0) \text{ and } \left(\frac{1}{n}J_n\right)^{-\frac{1}{2}} \to K^{-\frac{1}{2}}(\theta_0), n \to +\infty.$$

Because $\widehat{\theta}_n \to \theta_0$ a.e., so $\frac{1}{n}J_n(\theta_n^*) \to K(\theta_0)$. We obtain $G_n \to I_{m+1}$, $n \to +\infty$ a.e. The preceding lemma gives us

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \xrightarrow{D} \mathcal{N}\left(0, K^{-\frac{1}{2}}(\theta_0)K^{-\frac{1}{2}}(\theta_0)\right) = \mathcal{N}\left(0, K^{-1}(\theta_0)\right). \qquad \square$$

5. APPLICATIONS

This section contains an application of the beta regression model proposed in Section 2. All computations were carried out using *Mathematica 7.0.* After computing MLEs (maximum likelihood estimators), it is important to provide confidence intervals and to perform some diagnostic analyses in order to check the goodness-of-fit of the estimated model.

An approximate $(1 - \alpha)100\%$ confidence interval for θ_k, $k = 1, \ldots, m + 1$ and $\alpha \in (0, 0.5)$, has limits given by $\widehat{\theta}_k \pm q_{1-\frac{\alpha}{2}} \left(J_n^{-1}(\widehat{\theta})^{kk}\right)^{0.5}$, where $J_n^{-1}(\widehat{\theta})^{kk}$ is the kth diagonal element of $J_n^{-1}(\widehat{\theta})$ and q_γ represents the γ quantile of the $N(0, 1)$ distribution.

The discrepancy of fit can be measured as twice the difference between the maximum achievable log–likelihood and that attained under the fitted model. In other words, $D(y,\theta)=\sum_{i=j}^{n}2(l_j(\widehat{\theta})-l_j(\theta))$. We also have $D(y,\theta)=\sum_{j=1}^{n}(r_j^d)^2$, where

$$r_j^d = sign(y_j - \widehat{y_j})2(l_j(\widehat{\theta}) - l_j(\theta))^{0.5}.$$

The observation with a large absolute value of r_j^d may be viewed as discrepant. The r_j^d are called the jth deviance residuals. We shall also consider the residuals $r_j = y_j - E(\widehat{y_j})$.

Example 5.1. We draw a sample of size 80 from a beta distributed random variable, where $p(A_1,t_j)=A_1=1$ and $\phi=10$, $j=1,\dots,80$.

MLEs of A_1 and ϕ computed with *Mathematica 7.0* and Newton–Raphson method are $\widehat{A}_1 = 0.964\,269$ and $\widehat{\phi}=11.6337$. We have the following confidence intervals for A_1: $(0.6189, 1.3097)$, $(0.7019, 1.2668)$ and $(0.7447, 1.1838)$ for $\alpha=0.01$, $\alpha=0.05$ and $\alpha=0.1$, respectively. The confidence intervals for ϕ are $(6.5600, 16.7074)$, $(7.7793, 15.4881)$ and $(8.4086, 14.8588)$ for $\alpha=0.01$, $\alpha=0.05$ and $\alpha=0.1$, respectively. We have $D(y,\theta)=3.2104$.

Example 5.2. We draw a sample of size 80 from a beta distributed random variable, where $p(A_1,t_j)=A_1(\sin\frac{\Pi}{8}t_j+1.2)$, where $A_1=1$ and $\phi=30$, $j=1,\dots,80$. In such a way we could check periodicity in a shape parameter.

MLEs of A_1 and ϕ computed with *Mathematica 7.0* and Newton–Raphson method are $\widehat{A}_1=0.983\,251$ and $\widehat{\phi}=30.6473$. We have the following confidence intervals for A_1: $(0.6356, 1.3309)$, $(0.7191, 1.2474)$ and $(0.7623, 1.2043)$ for $\alpha=0.01$, $\alpha=0.05$ and $\alpha=0.1$, respectively. The confidence intervals for ϕ are $(17.4421, 43.8525)$, $(20.6155, 40.6791)$ and $(22.2530, 39.0413)$ for $\alpha=0.01$, $\alpha=0.05$ and $\alpha=0.1$, respectively. We also get $D(y,\theta)=0.1428$.

Example 5.3. We draw a sample of size 80 from a beta distributed random variable, where $p(A_1,t_j)=A_1(|\frac{1}{2}\sin\frac{\Pi}{8}t_j|+2)$, where $A_1=1$ and $\phi=10$, $j=1,\dots,80$.

MLEs of A_1 and ϕ computed with *Mathematica 7.0* and Newton–Raphson method are $\widehat{A}_1=1.02276$ and $\widehat{\phi}=9.66938$. We have the following confidence intervals for A_1: $(0.6304, 1.4151)$, $(0.7247, 1.3208)$ and $(0.7733, 1.2722)$ for $\alpha=0.01$, $\alpha=0.05$ and $\alpha=0.1$, respectively. The confidence intervals for ϕ are $(5.8447, 13.4941)$, $(6.7638, 12.5750)$ and $(7.2380, 12.1006)$ for $\alpha=0.01$, $\alpha=0.05$ and $\alpha=0.1$, respectively. We receive $D(y,\theta)=0.8709$.

Example 5.4. We draw a sample of size 40 from a beta distributed random variable, where $p(A_1,t_j)=A_1\sin\frac{\Pi}{8}t_j+A_2\frac{t_j}{8}+2$, where $A_1=1$, $A_2=1$ and $\phi=1000$, $j=1,\dots,40$. We consider periodical data with a trend.

MLEs of A_1, A_2 and ϕ computed with *Mathematica 7.0* and Newton–Raphson method are $\widehat{A}_1=1.05728$, $\widehat{A}_2=1.06575$ and $\widehat{\phi}=1141.7$. We have the following confidence intervals for A_1: $(0.1100, 2.0036)$, $(0.2654, 1.8490)$ and the following for A_2: $(0.4200, 1.7115)$, $(0.5254, 1.6060)$ for $\alpha=0.05$ and $\alpha=0.1$, respectively. The confidence intervals for ϕ are $(709, 1573.9)$ and $(780, 1503)$ for $\alpha=0.05$ and $\alpha=0.1$, respectively. We also have $D(y,\theta)=0.5063$.

Note that the computed discrepancies are small. The following plots (Fig. 1) of the residuals show no detectable pattern. However, the approximate confidence intervals are not very satisfying. It is due to the fact that we have a relatively small number of observations.

Crucially, the differences $\widehat{\theta}_k - \theta_k$ are relatively small. This is the main result of this section, which motivates our approach.

When the number of parameters exceed 3 the computations of maximum likelihood estimators do not give satisfying results or the computations are simply aborted.

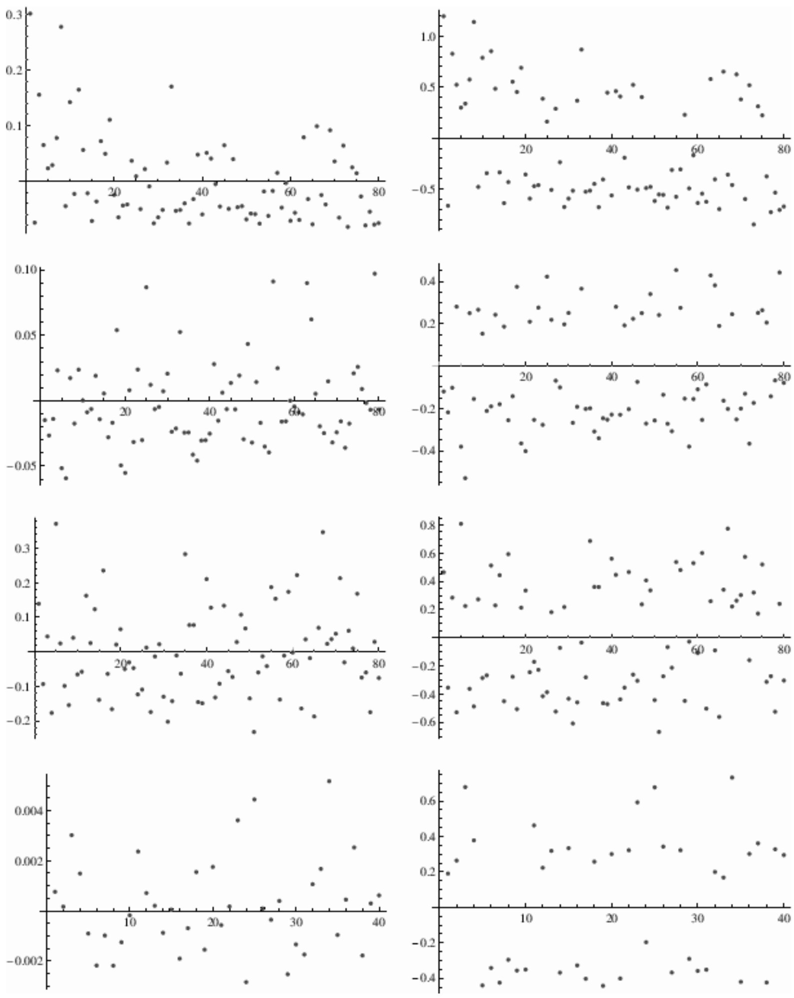

Fig. 1. Eight residuals plots for the four above examples. The left panels plot the residuals against the index of observation, while the right plot the deviance residuals against the index of observation

Acknowledgments

We thank very much the anonymous referees for their valuable comments and suggestions which helped to improve the manuscript.

The research of the first author has been partially supported by the Polish Ministry of Science and the Higher Education grant N N201 525838.

REFERENCES

[1] F. Cribari–Neto, K.L.P. Vasconcellos, *Nearly unbiased maximum likelihood estimation for the beta distribution*, J. Stat. Comput. Sim. **72** (2002), 107–118.

[2] L. Fahrmeir, H. Kaufmann, *Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models*, Ann. Statist. **13** (1985), 342–368.

[3] S.L.P. Ferrari, F. Cribari-Neto, *Beta regression for modelling rates and proportions*, J. Appl. Stat. **31** (2004) 7, 799–815.

[4] J.P. Rydlewski, *Beta–regression model for periodic data with a trend*, Univ. Iagel. Acta Math. **XLV** (2007), 211–222.

[5] J.P. Rydlewski, *A note on the maximum likelihood estimator in the gamma regression model*, Opuscula Math. **29** (2009) 3, 305–312.

[6] G.A.F. Seber, C.J. Wild, *Nonlinear Regression*, Wiley, New York, 2003.

[7] A.B. Simas, W. Barreto–Souza, A.V. Rocha, *Improved estimators for a general class of beta regression models*, Comput. Stat. Data Anal. **54** (2010), 348–366.

[8] R.W.M. Wedderburn, *On the existence and uniqueness of the maximum likelihood estimates for certain generalized linear models*, Biometrika **63** (1976), 27–32.

[9] B.C. Wei, *Exponential Family Nonlinear Models*, Springer, Singapore, 1998.

Jerzy P. Rydlewski
ry@agh.edu.pl

AGH University of Science and Technology
Faculty of Applied Mathematics
al. A. Mickiewicza 30, 30-059 Krakow, Poland

Dominik Mielczarek
dmielcza@mat.agh.edu.pl

AGH University of Science and Technology
Faculty of Applied Mathematics
al. A. Mickiewicza 30, 30-059 Krakow, Poland.

Received: June 29, 2011.
Revised: February 25, 2012.
Accepted: September 26, 2012.