Opuscula Mathematica • Vol. 29 • No. 3 • 2009

Jerzy P. Rydlewski

A NOTE ON THE MAXIMUM LIKELIHOOD ESTIMATOR IN THE GAMMA REGRESSION MODEL

Abstract. This paper considers a nonlinear regression model, in which the dependent variable has the gamma distribution. A model is considered in which the shape parameter of the random variable is the sum of continuous and algebraically independent functions. The paper proves that there is exactly one maximum likelihood estimator for the gamma regression model.

Keywords: gamma regression, nonlinear regression, maximum likelihood estimator, shape parameter.

Mathematics Subject Classification: 62J02, 62F10.

1. INTRODUCTION

In classical models of regression the following relationship is adopted

$$x_i = at_i + b + \varepsilon_i,$$

where the random variables ε_i are independent and have a normal distribution with average equal to zero. A somewhat more general form of this model can be expressed by the formula

$$x_i = \varphi(t_i, \theta) + \varepsilon_i,$$

where $\varphi(\cdot, \theta)$ is a function depending on the estimated parameter θ. This model can also be expressed in a different way, by assuming that the random variable x_i has a distribution of the form $N(\varphi(t_i, \theta), \sigma^2)$.

In some situations it is not natural to assume that the variable can take any real value. It may happen that the variable takes values only from a certain interval (see [2] and [3]), or only positive values. In this paper we will deal only with the latter case.

We assume that the random variable x_i has a distribution of the form $f(\varphi(\theta_1, t_i), \theta_2)$, where $\varphi(\theta_1, t_i)$ is the distribution mean and θ_2 is another parameter. A model of gamma regression is considered, i.e., the random variable x_i has a distribution of the form $\gamma(p, r)$. The random variable density function x_i with a gamma distribution has the form

$$f(t, p, r) = \frac{r^p}{\Gamma(p)} t^{p-1} e^{-rt}, \quad t > 0,$$

where $r > 0$ is the scale parameter, $p > 0$ is the shape parameter, and $\Gamma(\cdot)$ is the gamma function. The expected value of the random variable X with the gamma distribution is $E(X) = \frac{p}{r}$, and the variance $Var(X) = \frac{p}{r^2}$.

Questions of existence and uniqueness of the maximum likelihood estimates of the shape parameter in generalized linear models for one–parameter gamma distributed random variables have been studied by Wedderburn [5]. This paper gives a generalization of some of his results because the scale parameter is estimated as well.

The maximum likelihood estimation for two–parameter gamma distribution was widely discussed by Bowman and Shenton [1]. The authors did not, however, discuss the model, where shape or scale parameter is modelled by a function. The results are not covered by the Wei's monograph on exponential family nonlinear models (see [6] pp. 2–3).

Maximum likelihood estimation in different nonlinear models, as well as references to this literature, are given by Seber and Wild [4].

In our model we substitute p with a function of variable t_i, which depends on the multidimensional parameter A. Precisely, we use the set of m continuous and algebraically independent functions. Then,

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

Let $x_1, x_2, \ldots, x_n$ be independent random variables with a gamma distribution. It follows from the assumptions made that the expected values of the random variables x_j have the form

$$E(x_j) = \sum_{k=1}^{m} \alpha_k f_k(t_j), \qquad j = 1, 2, \ldots, n,$$

with $\alpha_k = \frac{A_k}{r}$.

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

$$L(x_1, x_2, \ldots, x_n, t_1, t_2, \ldots, t_n, A, r) = \prod_{j=1}^{n} \frac{1}{\Gamma(p(A, t_j))} r^{p(A,t_j)} x_j^{p(A,t_j)-1} e^{-rx_j}, \quad (1.1)$$

and hence the logarithm of the likelihood function is

$$\log L = \sum_{j=1}^{n} \Big(-\log \Gamma(p(A, t_j)) + p(A, t_j) \log r + (p(A, t_j) - 1) \log x_j - r x_j \Big). \quad (1.2)$$

We can write out our parameters in a different way, taking advantage of the properties of the gamma distribution. Let $a = (\alpha_1, \alpha_2, \ldots, \alpha_m)$. Then,

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

Obviously

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

where

$$\log L_j(ra, r) = -\log \Gamma\left(r\varphi(a, t_j)\right) + r\varphi(a, t_j) \log r + (r\varphi(a, t_j) - 1) \log x_j - rx_j.$$

In such a case the expected value of the variable x_j has the form $E(x_j) = \frac{p}{r} = \varphi(a, t_j)$, where $|\varphi(a, t_j)| < M$ for a fixed M and $j = 1, 2, \ldots, n$.

Later on, we will prove that the maximum likelihood estimator is determined uniquely.

2. MAXIMUM LIKELIHOOD ESTIMATION

Lemma 2.1. *Let $[c, d]$ be a closed and bounded interval. Let $f_1, f_2, \ldots, f_m$ be the set of algebraically independent functions continuous on $[c, d]$. The set of all parameters $\boldsymbol{A}$ of the form $(A_1, \ldots, A_m, r) \in \mathbb{R}^{m+1}$ such that for any $t \in [c, d]$,*

$$0 \leq \sum_{k=1}^{m} A_k f_k(t) \leq Mr \tag{2.1}$$

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

Proof. The fact that $\mathbf{A}$ is a non-empty set is obvious. Let $\mathbf{A}_t$ be the set $(A_1, \ldots, A_m, r) \in \mathbb{R}^{m+1}$ satisfying condition (2.1) for a fixed t. The set

$$\bigcap_{t \in [c,d]} A_t$$

is closed as the intersection of closed sets.

Let $\mathbb{A}_1 = (A_{10}, A_{11}, \ldots, A_{1m}, r_1) \in \mathbf{A}$ and $\mathbb{A}_2 = (A_{20}, A_{21}, \ldots, A_{2m}, r_2) \in \mathbf{A}$ be two vectors, and let $\lambda \in [0, 1]$. It can be easily shown that $\lambda \mathbb{A}_1 + (1 - \lambda)\mathbb{A}_2 \in \mathbf{A}$. □

Lemma 2.2. *Let $f_1, f_2, \ldots, f_m : \mathbb{R} \longrightarrow \mathbb{R}$ be the set of continuous and algebraically independent functions. The set $\mathbf{a}$ of all parameters $a = (\alpha_1, \alpha_2, \ldots, \alpha_m) \in \mathbb{R}^m$ satisfying the equation*

$$0 \leq \sum_{k=1}^{m} \alpha_k f_k(t) \leq M \tag{2.2}$$

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

Proof. By *reductio ad absurdum*, if the set $\mathbf{a}$ is unbounded, then there exists a sequence $\{(\alpha_1^n, \alpha_2^n, \ldots, \alpha_m^n)\}_{n\in\mathbb{N}}$ of elements of the set $\mathbf{a}$ such that

$$(\alpha_1^n, \alpha_2^n, \ldots, \alpha_m^n) \longrightarrow +\infty, \; n \to +\infty.$$

We have

$$\frac{(\alpha_1^n, \alpha_2^n, \ldots, \alpha_m^n)}{\| (\alpha_1^n, \alpha_2^n, \ldots, \alpha_m^n) \|} \in S^{m-1},$$

where S^{m-1} is a unit $m-$dimensional sphere. We can choose a subsequence

$$\left\{ \frac{\left(\alpha_1^{l_n}, \alpha_2^{l_n}, \ldots, \alpha_m^{l_n}\right)}{\| \left(\alpha_1^{l_n}, \alpha_2^{l_n}, \ldots, \alpha_m^{l_n}\right) \|} \right\}_{n\in\mathbb{N}}$$

convergent to some $\left(\alpha_1^0, \alpha_2^0, \ldots, \alpha_m^0\right)$, and we get

$$\forall t \in \mathbb{R} \qquad 0 \le \sum_{k=1}^{m} \alpha_k^{l_n} f_k(t) \le M.$$

We have

$$0 \le \frac{\sum_{k=1}^{m} \alpha_k^{l_n} f_k(t)}{\| (\alpha_1^n, \alpha_2^n, \ldots, \alpha_m^n) \|} \le \frac{M}{\| (\alpha_1^n, \alpha_2^n, \ldots, \alpha_m^n) \|}.$$

As $n \to +\infty$, we get $\sum_{k=1}^{m} \alpha_k^0 f_k(t) = 0$. Taking m suitable values $t_1, t_2, \ldots, t_m$ gives $\alpha_k^0 = 0$ for $k = 1, 2, \ldots, m$, which contradicts the fact that $\left(\alpha_1^0, \alpha_2^0, \ldots, \alpha_m^0\right) \in S^{m-1}$. From this it follows that the set $\mathbf{a}$ is bounded. Clearly, it is also non-empty and closed. □

Lemma 2.3. *Exactly one of the conditions specified below is true*

(i) *For all* $j = 1, 2, \ldots, n$

$$x_j = \sum_{k=1}^{m} \alpha_k f_k(t_j).$$

(ii)

$$\lim_{r\to+\infty} \frac{d}{dr} \log L(ra, r) < 0.$$

Proof. Let $y_j = \varphi(a, t_j)$. Then

$$\frac{d}{dr} \log L_j(ra, r) = -y_j \Psi\left(ry_j\right) + y_j \log r + y_j + y_j \log x_j - x_j,$$

where $\Psi(x) = \frac{d}{dx} \log \Gamma(x)$. We have

$$\lim_{x\to+\infty} \left(\Psi(x) - \log x\right) = 0,$$

and hence

$$\lim_{r\to+\infty}\frac{d}{dr}\log L_j(ra,r) = -y_j\log y_j + y_j + y_j\log x_j - x_j.$$

Let

$$g(x) = -y_j\log x + x.$$

The function g takes the smallest value at $x = y_j$, and hence

$$\lim_{r\to+\infty}\sum_{j=1}^{n}\frac{d}{dr}\log L_j(ra,r) \le 0$$

and if for at least one j we have $x_j \neq y_j$, then we obtain

$$\lim_{r\to+\infty}\sum_{j=1}^{n}\frac{d}{dr}\log L_j(ra,r) < 0.$$

□

Lemma 2.4. *The function* $\log L(ra,r)$ *as a function of the parameter* r *is strictly concave.*

Proof. We have

$$\log L(ra,r) = \sum_{j=1}^{n}\Big(-\log\Gamma(r\varphi(a,t_j)) + r\varphi(a,t_j)\log r + (r\varphi(a,t_j)-1)\log x_j - rx_j\Big).$$

Hence

$$\frac{d}{dr}\log L(ra,r) = \sum_{j=1}^{n}\Big(-\frac{d}{dr}(\log\Gamma(r\varphi(a,t_j))) + \varphi(a,t_j)\log r + \varphi(a,t_j)(1+\log x_j) - x_j\Big)$$

and

$$\frac{d^2}{dr^2}\log L(ra,r) = \sum_{j=1}^{n}\Big(-\frac{d^2}{dr^2}(\log\Gamma(r\varphi(a,t_j))) + \frac{\varphi(a,t_j)}{r}\Big) =$$

$$= \sum_{j=1}^{n}\Big(-(\varphi(a,t_j))^2\,\Psi'(r\varphi(a,t_j)) + \frac{\varphi(a,t_j)}{r}\Big).$$

It is sufficient to check if

$$\Psi'(r\varphi(a,t_j)) > \frac{1}{r\varphi(a,t_j)},\quad j = 1,\dots,n.$$

It is known that

$$\Psi'(y) = \sum_{n=0}^{+\infty}\frac{1}{(y+n)^2},$$

and

$$\sum_{n=0}^{+\infty} \frac{1}{(y+n)^2} > \int_0^{+\infty} \frac{ds}{(y+s)^2} = \frac{1}{y},$$

which proves the lemma. □

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

$$\begin{bmatrix} f_1(t_1) & f_2(t_1) & \cdots & f_m(t_1) \\ f_1(t_2) & f_2(t_2) & \cdots & f_m(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ f_1(t_n) & f_2(t_n) & \cdots & f_m(t_n) \end{bmatrix}. \tag{2.3}$$

Lemma 2.5. *If the number n of observations is sufficient, i.e., $n \geq m$, and the rank of the matrix J is maximal, i.e., rank$J = m$, then the function $\log L(A_1, \ldots, A_m, r)$ is strictly concave.*

Proof. Let

$$F(x, r) = -\log \Gamma(x) + x \log r.$$

The Hessian matrix of the function F is

$$H_F = \begin{bmatrix} -\frac{x}{r^2} & \frac{1}{r} \\ \frac{1}{r} & -\Psi'(x) \end{bmatrix}.$$

As $x\Psi'(x) > 1$, the matrix H_F is negative definite and thus $F(x, r)$ is strictly concave. From (1.2) we have

$$\log L = \sum_{j=1}^{n} \log L_j.$$

Each function $\log L_j$ is the sum of a linear function and a function, which is the composition of F and the multilinear function $p(A_1, \ldots, A_m, t_j)$. Using only the definition of concavity we can easily prove that each $\log L_j$ is concave. Since we have assumed that the rank of the matrix J is maximal, the intersection of all hyperplanes

$$p(A_1, \ldots, A_m, t_j) = const$$

is at most a single point. Thus $\log L(A_1, \ldots, A_m, r)$ is a strictly concave function. □

Theorem 2.6. *Let $n \geq m$ and for given $t_1, t_2, \ldots, t_n \in [c, d]$ let the rank of matrix J defined in (2.3) be maximal. Then for given $t_1, t_2, \ldots, t_n \in [c, d]$ and $x_1, x_2, \ldots, x_n$ there exists exactly one $(\widehat{A}, \widehat{r}) \in \boldsymbol{A}$ such that*

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

with probability 1, where L is the likelihood function defined in (1.1).

Proof. It follows from Lemma 2.3 that, with probability one, condition (ii) of the Lemma holds true. Let

$$g(a) = \lim_{r \to +\infty} \frac{d}{dr} \log L(ra, r),$$

and

$$K(r) = \left\{ a \in \mathbf{a} : \frac{d}{dr} \log L(ra, r) \geq g(a) + \frac{|g(a)|}{2} \right\}.$$

The function $g(a)$ is continuous, as the proof of Lemma 2.3 shows. This, and Lemma 2.2 give us that every set $K(r)$ is compact. The function $\frac{d}{dr} \log L(ra, r)$ is decreasing as a derivative of a strictly concave function. The set–valued map K is also decreasing. From Lemma 2.3 we obtain that

$$\bigcap_{r>0} K(r) = \emptyset.$$

As every set $K(r)$ is compact, by the Riesz theorem there exists $r_0 > 0$, such that $K(r_0) = \emptyset$. Using Lemma 2.5, we get that there exists exactly one $\left(\widehat{A}, \widehat{r}\right) \in \mathbf{A}$, such that

$$\log L(\widehat{A}, \widehat{r}) = \max_{(A,r) \in \mathbf{A}_0} \log L(A, r),$$

where $\mathbf{A}_0 = \{(A, r) \in \mathbf{A} : r \leq r_0\}$ is a convex and compact set.

Let $(A, r) \in \mathbf{A} \backslash \mathbf{A}_0$ and $\alpha_i = \frac{A_i}{r}$, for $i = 1, \ldots, m$. As K is a decreasing set–valued map, for any $r \geq r_0$ we have $K(r) = \emptyset$. Consequently, for any $r \geq r_0$ and for any $a \in \mathbf{a}$ we have $\frac{d}{dr} \log L(ra, r) < 0$ and thus $\log L(ra, r)$ is a decreasing function of the argument r for $r \geq r_0$. From this we find that

$$\log L(A, r) = \log L(ra, r) < \log L(r_0 a, r_0) \leq \log L(\widehat{A}, \widehat{r}).$$

By proving that there exists a global maximum of the function $\log L$ we completed the proof of Theorem 2.1. □

Acknowledgements

Financial support from AGH University of Science and Technology local grant No. 10.420.03. is gratefully acknowledged.

REFERENCES

[1] K.O. Bowman, L.R. Shenton, *Properties of Estimators for the Gamma Distribution*, Marcel Dekker, New York, 1988.

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

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

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

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

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

J.P. Rydlewski
ry@agh.edu.pl

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

Jagiellonian University
Department of Mathematics
ul. Łojasiewicza 6, 30-348 Krakow, Poland

Received: March 20, 2009.
Revised: April 12, 2009.
Accepted: April 20, 2009.