Hans Brøns Søren Tolver Jensen

Maximum Likelihood Estimation in the Negative Binomial Distribution

Preprint
December
1988

8

Institute of Mathematical Statistics
University of Copenhagen

ISSN 0902-8846

Hans Brøns and Søren Tolver Jensen

MAXIMUM LIKELIHOOD ESTIMATION IN THE NEGATIVE BINOMIAL DISTRIBUTION

Preprint 1988 No. 8

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

December 1988

Maximum Likelihood Estimation in the Negative Binomial Distribution.

by

Hans Brøns and Søren Tolver Jensen

Institute of Mathematical Statistics

University of Copenhagen

Abstract: For the negative binomial distribution the existence and uniqueness of the maximum likelihood estimator is proved using convexity results.

1. Introduction. Anscombe (1950) conjectured that the maximum likelihood estimator in the negative binomial distribution exists uniquely if and only if the sample variance is greater than the sample mean. In an analytic *tour de force*, this was proved by Simonsen (1976, 1979). Unfortunately, the result is not widely known, see e.g. Kotz and Johnson (1985), p. 173, perhaps because the proof is long and complicated. The importance of the result for the further development of the statistical models based on the negative binomial distribution made us look for an alternative way to prove it. After a considerable effort we have found a simple and short proof, which we shall present here.

Let $x_1,\dots,x_n$ be i.i.d. random variables, $n = 1,2,\dots$, with a negative binomial distribution

$$\binom{\alpha+x-1}{x} p^{\alpha}(1-p)^{x}$$

for $x = 0,1.2,\dots$, where $0 < p < 1$ and $0 < \alpha < \infty$. The likelihood function is

$$L(p,\alpha) = p^{\alpha n}(1-p)^S \prod_{j=1}^{n} \binom{\alpha+x_j-1}{x_j},$$

where

$$S = \sum_{j=1}^{n} x_j .$$

For $S = 0$ we have $L(p,\alpha) = p^{\alpha n}$ and so in this case the maximum likelihood estimator does not exist. In the following it is therefore assumed that $S > 0$. For fixed α the maximum likelihood estimator of p is $\hat{p}(\alpha) = \alpha n/(\alpha n+S) = \alpha/(\alpha+m)$, where $m = S/n$ is the sample mean. The profile function of the negative log likelihood function becomes

$$f(\alpha) = -\log L(\hat{p}(\alpha),\alpha)$$
$$= n((\alpha+m)\log(\alpha+m)-\alpha\log(\alpha)-m\log(m)) - \sum_{j=1}^{n} \sum_{k=0}^{x_j-1} (\log(\alpha+k)-\log(k+1)).$$

The difficult point is to show that the likelihood equation $Df(\alpha) = 0$ has at most one solution. Anscombe's conjecture is then an easy consequence, see sect.3. Simonsen proved that Df is strictly quasi concave. Since $Df(\alpha) \to 0$ for $\alpha \to \infty$, it follows that the likelihood equation has at most one solution. We shall prove the stronger result that the profile function f is strictly convex as a function of $\beta = \alpha(\log(\alpha+m) - \log(\alpha))/m$.

<u>2.Transformation to convex functions</u>. It was conjectured by one of the authors (Brøns) and proved by Johansen (1972) that if a family of quasi

convex functions defined on $\mathbb{R}$ is closed under addition, then it could be transformed into a family of convex functions by as monotone transformation of the domain of definition. It is clear that the family of profile functions with a fixed m is closed under addition. The uniqueness of the maximum likelihood estimator should therefore be proved by a transformation to convex functions. The problem is then first to find the transformation and then to prove that the function given by (1) below is increasing.

Theorem. The profile function f is a convex function in the parameter $\beta = \alpha(\log(\alpha+m) - \log(\alpha))/m$.

Proof. Differentiating f we get

$$Df(\alpha) = n(\log(\alpha+m) - \log(\alpha)) - \sum_{j=1}^{n} \sum_{k=0}^{x_j-1} \frac{1}{\alpha+k}$$

and

$$D^2f(\alpha) = n(\frac{1}{\alpha+m} - \frac{1}{\alpha}) + \sum_{j=1}^{n} \sum_{k=0}^{x_j-1} \frac{1}{(\alpha+k)^2}.$$

The substitution β: $]0,\infty[\rightarrow]0,1[$ is a strictly increasing and strictly concave function with

$$\begin{aligned} D\beta(\alpha) &= \frac{1}{m}((\log(\alpha+m) - \log(\alpha)) - \frac{1}{\alpha+m} \\ &= \int_{\alpha}^{\infty} (\frac{1}{m}(\frac{1}{t} - \frac{1}{t+m}) - (\frac{1}{t+m})^2)dt \\ &= \int_{\alpha}^{\infty} m/(t(t+m)^2)dt > 0 \end{aligned}$$

and

$$D^2\beta(\alpha) = - m/(\alpha(\alpha+m)^2) < 0.$$

We have

$$Df(\alpha) = D_\beta f(\alpha)\ D\beta(\alpha)$$

and

$$\begin{aligned} D^2f(\alpha) &= D_\beta^2 f(\alpha)(D\beta(\alpha))^2 + D_\beta f(\alpha) D^2\beta(\alpha) \\ &= D_\beta^2 f(\alpha)(D\beta(\alpha))^2 + Df(\alpha)D^2\beta(\alpha)/D\beta(\alpha). \end{aligned}$$

Hence

$$D_\beta^2 f(\alpha) = (D^2f(\alpha)D\beta(\alpha) - Df(\alpha)D^2\beta(\alpha))/(D\beta(\alpha))^3,$$

and so $D_\beta^2 f(\alpha) > 0$ if and only if

$$D^2f(\alpha)D\beta(\alpha) > Df(\alpha)D^2\beta(\alpha),$$

If we put

(1) $$h(\alpha) = D^2f(\alpha)/D^2\beta(\alpha)$$

and notice that both $Df(\alpha)$ and $D\beta(\alpha)$ tend to zero for $\alpha \to \infty$ the inequality becomes

$$\int_\alpha^\infty h(\alpha)\ D^2\beta(\alpha)D^2\beta(t)dt < \int_\alpha^\infty h(t)\ D^2\beta(t)D^2\beta(\alpha)dt$$

which is true if $h(\alpha)$ is strictly increasing in α.

We have

$$\begin{aligned} h(\alpha) &= -\frac{\alpha(\alpha+m)^2}{m}\left(n\left(\frac{1}{\alpha+m} - \frac{1}{\alpha}\right) + \sum_{j=1}^{n}\sum_{k=0}^{x_j-1}\frac{1}{(\alpha+k)^2}\right) \\ &= n((\alpha+m) - (\alpha+m)^2 g(\alpha)) \end{aligned}$$

where

(2) $$g(\alpha) = \frac{1}{mn}\sum_{j=1}^{n}\sum_{k=0}^{x_j-1}\frac{\alpha}{(\alpha+k)^2}.$$

Rearranging we get

$$h(\alpha) = \frac{1}{m}\sum_{j=1}^{n}\sum_{k=0}^{x_j-1}\left((\alpha+m) - \frac{\alpha(\alpha+m)^2}{(\alpha+k)^2}\right)$$
$$= \frac{1}{m}\sum_{j=1}^{n}\sum_{k=0}^{x_j-1}\frac{(\alpha+m)(2k\alpha+k^2-m\alpha)}{(\alpha+k)(\alpha+k)},$$

which shows that

$$(3) \qquad h(\alpha) \to \frac{1}{m}\sum_{j=1}^{n}\sum_{k=0}^{x_j-1}(2k-m)$$
$$= \frac{1}{m}\sum_{j=1}^{n}(x_j(x_j-1)-x_jm) \quad = n(s^2-m)/m$$

for $\alpha \to \infty$, where $s^2 = \frac{1}{n}\sum_{j=1}^{n}(x_j-m)^2$ is the sample variance.

To prove that h is strictly increasing it therefore suffices to show that h is strictly concave, but $h(\alpha) = n((\alpha+m) - (\alpha+m)^2 g(\alpha))$ is strictly concave if and only if $(\alpha+m)^2 g(\alpha)$ is strictly convex.

The set G of functions $g:]0,\infty[\to]0,\infty[$ such that $(\alpha+m)^2 g(\alpha)$ is strictly convex is evidently a convex cone, and so by (2) we only have to show that the functions

$$g_i(\alpha) = \sum_{k=0}^{i-1}\frac{\alpha}{(\alpha+k)^2},$$

belong to G for $i = 1,2,\ldots$.

For any $j = 0,1,\ldots,i-1$

$$\sum_{k=0}^{j}\frac{\alpha}{(\alpha+k)^2} = \sum_{k=0}^{j}\frac{1}{\alpha+k} - \sum_{k=1}^{j}\frac{k}{(\alpha+k)^2}$$
$$= \sum_{k=0}^{j}\frac{1}{\alpha+k} - \sum_{k=1}^{j}\left(\frac{k-1+1}{\alpha+k-1} - \frac{k}{\alpha+k}\right)\left(1 - \frac{1}{\alpha+k}\right)$$
$$= \sum_{k=0}^{j}\frac{1}{\alpha+k} - \sum_{k=1}^{j}\frac{1}{\alpha+k-1} + \frac{j}{\alpha+j} + \sum_{k=1}^{j}\frac{k}{(\alpha+k)^2(\alpha+k-1)}$$
$$= \frac{j+1}{\alpha+j} + \sum_{k=1}^{j}\frac{k}{(\alpha+k)^2(\alpha+k-1)}.$$

Hence

$$(\alpha+m)^2 \sum_{k=0}^{i-1} \frac{\alpha}{(\alpha+k)^2} = (\alpha+m)^2 \frac{j+1}{\alpha+j} + \sum_{k=1}^{j} \frac{(\alpha+m)^2 k}{(\alpha+k)^2(\alpha+k-1)} + \sum_{k=j+1}^{i-1} \alpha \left(\frac{\alpha+m}{\alpha+k}\right)^2.$$

If we now choose the integer j such that $j = 0$ for $m < 1$, $m-1 < j \leq m$ for $1 \leq m < i-1$, and $j = i-1$ for $i-1 \leq m$, it is clear that the three terms are all convex functions. First

$$(\alpha+m)^2/(\alpha+j) = \alpha+j + 2(m-j) + (m-j)^2/(\alpha+j)^2$$

is convex, secondly

$$(\alpha+m)^2/((\alpha+k)^2(\alpha+k-1))$$

is even logaritmic convex for $1 \leq k \leq m$, and finally

$$\alpha\ (\alpha+m)^2/(\alpha+k)^2 = \alpha + 2(m-k) + (m-k)(m-3k)/(\alpha+k) - k(m-k)^2/(\alpha+k)^2$$

is convex for $k > m$, because then the second derivative

$$(2(m-k)(m-3k)\alpha + 4mk(k-m))/(\alpha+k)^4$$

is positive.

<u>3. Proof of Anscombe's conjecture</u>. Since $D_\beta f(\alpha) = Df(\alpha)/D\beta(\alpha)$ is strictly increasing, the likelihood equation $Df(\alpha) = 0$ has at most one solution.

It is clear that $\lim_{\alpha \to 0} D_\beta f(\alpha) \leq 0$, and, since both $Df(\alpha)$ and $D\beta(\alpha)$ tend to 0 for $\alpha \to \infty$, we have by (3)

$$\lim_{\alpha \to \infty} D_\beta f(\alpha) = \lim_{\alpha \to \infty} Df(\alpha)/D\beta(\alpha) = \lim_{\alpha \to \infty} D^2 f(\alpha)/D^2\beta(\alpha),$$
$$= \lim_{\alpha \to \infty} h(\alpha) = n(s^2 - m)/m.$$

Hence a solution exists if and only if $s^2 > m$.

References.

Anscombe, F.J. (1950). Sampling theory of the negative and logarithmic series distributions. Biometrika 37, 358-382.

Johansen, S. (1972). A representation theorem for a convex cone of quasi convex functions. Math. Scand. 30, 297-312.

Kotz, S. and N.L.Johnson (1985). Encyclopedia of Statistical Sciences, Volume 6. Wiley, New York.

Simonsen, W. (1976). On the Solution of a Maximum-Likelihood Equation of the Negative Binomial Distribution. Scand. Actuarial J. 1976:220-231.

Simonsen, W. (1979). Correction note. Scand. Actuarial J. 1979:228-229.

PREPRINTS 1987

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK, TELEPHONE + 45 1 35 31 33.

No. 1 Jensen, Søren Tolver and Johansen, Søren: Estimation of Proportional Covariances.

No. 2 Rootzén, Holger: Extremes, Loads, and Strengths.

No. 3 Bertelsen, Aksel: On the Problem of Testing Reality of a Complex Multivariate Normal Distribution.

No. 4 Gill, Richard D. and Johansen, Søren: Product-Integrals and Counting Processes.

No. 5 Leadbetter, M.R. and Rootzén, Holger: Extremal Theory for Stochastic Processes.

No. 6 Tjur, Tue: Block Designs and Electrical Networks.

No. 7 Johansen, Søren: Statistical Analysis of Cointegration Vectors.

No. 8 Bertelsen, Aksel: On the Problem of Testing Reality of a Complex Multivariate Normal Distribution, II.

No. 9 Andersson, S.A. and Perlman, M.D.: Group-Invariant Analogues of Hadamard's Inequality.

No. 10 Hald, Anders: Two Generalizations of the Problem of Points by Bernoulli, de Moivre and Montmort.

No. 11 Andersson, Steen Arne: The Lattice Structure of Orthogonal Linear Models and Orthogonal Variance Component Models.

PREPRINTS 1988

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK, TELEPHONE + 45 1 35 31 33.

No. 1 Jacobsen, Martin: Discrete Exponential Families: Deciding when the Maximum Likelihood Estimator Exists and Is Unique.

No. 2 Johansen, Søren and Juselius, Katarina: Hypothesis Testing for Cointegration Vectors - with an Application to the Demand for Money in Denmark and Finland.

No. 3 Jensen, Søren Tolver, Johansen, Søren and Lauritzen, Steffen L.: An Algorithm for Maximizing a Likelihood Function.

No. 4 Bertelsen, Aksel: On Non-Null Distributions Connected with Testing that a Real Normal Distribution Is Complex.

No. 5 Tjur, Tue: Statistical Tables for Personal Computer Users.

No. 6 Tjur, Tue: A New Upper Bound for the Efficiency of a Block Design.

No. 7 Bunzel, Henning, Høst, Viggo and Johansen, Søren: Some Simple Non-Parametric Tests for Misspecification of Regression Models Using Sign Changes of Residuals.

No. 8 Brøns, Hans and Jensen, Søren Tolver: Maximum Likelihood Estimation in the Negative Binomial Distribution.