International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:6, No:2, 2012

Bootstrap Confidence Intervals and Parameter Estimation for Zero Inflated Strict Arcsine Model

Y. N. Phang, E. F. Loh

***Abstract*—**Zero inflated Strict Arcsine model is a newly developed model which is found to be appropriate in modeling overdispersed count data. In this study, maximum likelihood estimation method is used in estimating the parameters for zero inflated strict arcsine model. Bootstrapping is then employed to compute the confidence intervals for the estimated parameters.

***Keywords*—**overdispersed count data**,** maximum likelihood estimation, simulated annealing, BCa confidence intervals.

I. Introduction

LETAC and Mora [10] introduced the strict arcsine (SA) model. Kokonendji [8] compared the strict arcsine distributiuon with Poisson, negative binomial, Poisson inverse gaussian, and generalized Poisson models by using the moment method to estimate the parameters. The SA model is found to be overdispersed, skewed and leptokurtic. Marque and Kokonendji [9] studied the strict arcsine regression model. Phang and Loh [11] developed zero inflated strict arcsine (ZISA) model and fitted it to a simulated and a real life data sets. The study showed that this developed model can be used as an alternative model in modeling overdispersed count data. In this paper, we apply the maximum likelihood estimation method through a global optimization routine to estimate the parameters for ZISA model. Bootstrap methods such as simple percentile, normal and BCa (bootstrap Bias Corrected-adjusted) methods are used in computing the confidence interval for all the estimated parameters for ZISA model. This is to ensure that the accuracy of the obtained estimated parameters is achieved in order to avoid misleading inference. Bootstrap confidence interval methods are used because they do not depend on the normality assumption. Some practical examples of confidence interval construction are found in Efron and Tibshirani [5] and Davison and Hinkley [3]. DiCicco and Efron [4] give a detailed review of the methods discussed in this paper. Carpenter and Bithell [2] provide a good discussion on the application of bootstrap confidence intervals. Section 2 of the paper discusses the properties of the SA and ZISA models, section 3 considers statistical inference which involves parameter estimation and computing confidence intervals.

Y. N. Phang is with Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Kampus Bandaraya Melaka,Malaysia (e-mail:phang@melaka.uitm.edu.my)

E.F.Loh is with Academy of Language Study, Universiti Teknologi MARA, Kampus Bandaraya Melaka, Malaysia (email: david_loh@melaka.uitm.edu.my)

The results are discussed in section 4. Section 5 provides a short conclusion.

II. Properties Of The Distributions

A. The Strict Arcsine Distribution

The SA distribution is introduced by Letac and Mora [10]. Kokonendji [8] studied the properties of the strict arcsine distribution and found that the SA distribution is overdispersed, skewed to the right and leptokurtic.

The pmf of SA is given by

$$\Pr_{SA}(x) = \frac{A(x;\alpha)}{x!}\, p^x \exp\{-\alpha \arcsin(p)\}, \qquad x = 0, 1, 2, \ldots\ldots \tag{1}$$

where $0 < \alpha$, $0 < p < 1$, and $A(x; \alpha)$ is defined as

$$A(x;\alpha) = \begin{cases} \prod_{k=0}^{z-1} (\alpha^2 + 4k^2) & \text{if } x = 2z; \text{and} A(0,\alpha) = 1 \\ \alpha \prod_{k=0}^{z-1} (\alpha^2 + (2k+1)^2) & \text{if } x = 2z+1; \text{and} A(1,\alpha) = \alpha \end{cases} \tag{2}$$

The recurrence formula of SA is

$$\Pr(x+1) = \frac{A(x+1;\alpha)}{A(x;\alpha)} \cdot \frac{p}{x+1} \Pr(x), \qquad x = 0, 1, 2, \ldots\ldots \tag{3}$$

with

$$\Pr(0) = \exp(-\alpha \arcsin(p)) \text{ and } \Pr(1) = \alpha p \exp(-\arcsin(p)).$$

The likelihood L is given by

$$L_{SA} = \prod_{k=0}^{x} \Pr_{SA}(k)^{F_k}, \qquad x = 0, 1, 2, \ldots\ldots \tag{4}$$

and the log-likelihood is

$$\ln L_{SA} = \sum_{k=0}^{x} F_k \ln \Pr_{SA}(k). \tag{5}$$

The likelihood score functions are given below

$$\frac{\partial \ell_{SA}}{\partial \alpha} = \sum_{k=0}^{x} F_k \frac{\partial \log A(k,\alpha)}{\partial \alpha} - \arcsin(p), \quad x = 0, 1, 2, \ldots\ldots \tag{6}$$

where

$$\frac{\partial \log A(x,\alpha)}{\partial \alpha} = \begin{cases} \sum_{k=0}^{z-1} \frac{2\alpha}{(\alpha^2+4k^2)}, & \text{if } x=2z \text{ and } \frac{\partial \log A(0,\alpha)}{\partial \alpha}=0 \\ -\frac{1}{\alpha^2} \sum_{k=0}^{z-1} \frac{2(2k+1)^2-2\alpha^2}{\left[\alpha^2+(2k+1)^2\right]^2}, & \text{if } x=2z+1 \text{ and } \frac{\partial \log A(1,\alpha)}{\partial \alpha}=1 \end{cases}$$

$$\frac{\partial \ell_{SA}}{\partial p} = \sum_{k=0}^{x} F_k \left[\frac{k}{p} - \frac{\alpha}{\sqrt{1-p^2}} \right], \qquad x = 0, 1, 2, \ldots \tag{7}$$

B.The Zero Inflated Strict Arcsine Model

Phang and Loh [11] developed the zero inflated strict arcsine model and fitted it to a simulated and a real life data sets. The study shown that this model can be used in modeling zero inflated count data.

The pmf for ZISA is given by

$$P_{ZISA}(Y = 0) = \omega + (1-\omega)\exp(-\alpha \arcsin(p))$$

$$P_{ZISA}(Y = y) = (1-\omega)\,\Pr_{SA}(y), \quad y = 1,2,3,\ldots \tag{8}$$

The likelihood L is given by

$$L_{ZISA} = \prod_{k=0}^{x} \Pr_{ZISA}(k)^{F_k}, \qquad x = 0, 1, 2, \ldots \tag{9}$$

and the log-likelihood is

$$\ln L_{ZISA} = \sum_{k=0}^{x} F_k \ln \Pr_{ZISA}(k). \tag{10}$$

III. Statistical Inference

A. Parameter Estimation

In this study, simulated annealing [7], a global optimization routine, is used to obtain the maximum likelihood estimate. The method of maximum likelihood is used for estimating the unknown parameters because of the desirable qualities of the maximum likelihood estimates like best asymtotically normal [12] under certain regularity conditions. Moreover, it is more suitable for statistical inference such as constructing confidence intervals or testing hypotheses. Simulated annealing applied the concept of cooling process where at high temperature, molecules are free to move, but the mobility of the molecules drops as the temperature decreases and the molecules tend to line themselves up in a rigid structure which in fact is a stage of minimum energy. The advantage of this approach is that derivatives of the likelihood function are not needed. To check that a global optimum is achieved, various seeds from the random generator RANMAR and temperature reduction factor are used. Convergence is evaluated at each step by the difference between two function values lower than 10^{-6} (which is the convergence criteria). The maximum likelihood estimates are validated by substituting these estimates into the likelihood score equations.

We find the parameter estimates for a real life data set using the above-mentioned method. Table I shows the data set of Gossiaux & Lemaire [6]: Automobile claim frequency data.

TABLE I
AUTOMOBILE CLAIM FREQUENCY DATA, GOSSIAUX AND LEMAIRE (1981)

	Observed frequency	Expected frequency
		ZISA
0	103704	103704.00
1	14075	14073.16
2	1766	1763.25
3	255	265.49
4	45	38.85
5	6	6.85
6	2	1.40
Total	119853	
-Log-likelihood		54609.69
χ^2		1.58

$\hat{p}_{ZISA} = 0.2244$ $\hat{\alpha}_{ZISA} = 1.11625$ $\hat{\omega} = 0.3967$
Mean = 17122
Variance =1.484×10^9

B. Bootstrap Confidence Interval

SA and ZISA are found to be positively skewed [8]. According to Ader et al [1], the bootstrap procedure is recommended when the theoretical distribution of a statistic of interest is complicated or unknown. In this study, we consider the non-parametric bootstrap method. The parameters for each generated bootstrap sample are computed. The estimated parameters are then resample to construct the confidence intervals using percentile, normal and Bca methods.

The steps involved are as follow:

1. Resample the original data set 1000 times.
2. Estimate the parameters for each bootstrap sample using maximum likelihood estimation method and

let the estimated parameters be $\hat{\theta}_i^{(r)}$ where r = 1,2.....1000 and i=1 ...n

3. Resample the estimated parameters 1000 times.
4. Calculate the normal, percentile and BCa confidence intervals for the parameter estimates.

IV. Results

The results show that the estimated parameters for the original data set fell within the specified limits of all confindence intervals constructed using the bootstrap methods.

Table I shows the expected frequencies and estimated parameters for the data set. Table II shows the confidence intervals for parameters of ZISA constructed using simple percentile method, BCa confidence limits and normal methods.

TABLE II
CONFIDENCE INTERVALS FOR ESTIMATED PARAMETERS FOR ZISA MODEL

		Lower bound	Upper bound
p	CI_1	0.2228	0.2244
	CI_2	0.2228	0.2244
	CI_3	0.2228	0.2244
α	CI_1	1.1200	1.1310
	CI_2	1.1200	1.1308
	CI_3	1.1200	1.1308
ω	CI_1	0.3959	0.3976
	CI_2	0.3959	0.3976
	CI_3	0.3959	0.3976

Legend: CI_1 simple percentile; CI_2 BCa ; CI_3 normal 95% confidence intervals

V. CONCLUDING REMARKS

All the estimated parameters of ZISA from the original data fell in all the confidence intervals constructed using the bootstrap methods. The confidence intervals obtained using percentile, normal and BCa methods appear to be the same. The widths of the confidence intevals are a bit narrow. The results show that the estimating method used in estimating the parameters for ZISA is valid. This study also confirm the finding found in Phang and Loh [11] that ZISA can be used as an alternative model in modelling overdispersed count data.

REFERENCES

[1] H. J. Ader, G. J. Mellenbergh, and D. J. Hand, "Advising on research methods: A consultant's companion". *Huizen, The Netherlands: Johannes van Kessel Publishing* ISBN 978-90-79418-01-5, 2008

[2] J. Carpenter, and J. Bithell, "Bootstrap confidence intervals: when, which, what? A practical guide for medical guide for medical statisticians". *Statistics in Medicine*, 19, 1141-1164, 2000.

[3] A. C. Davison, and D. V. Hinkley, "Bootstrap methods and their application". *Cambridge University Press* ,1996.

[4] T. J. DiCiccio, and B. Efron, "Bootstrap confidence intervals". *Statistical Science*, 11, 189-212, 1996.

[5] B. Efron, and R. J. Tibshirani, " An introduction to the Bootstrap". Chapman and Hall: London, 1993.

[6] A. Gaussiaux, and J. Lemaire, "Methodes d'ajustement de distributions de sinistres" . *Bulletin of the Association of Swiss Actuaries*, 81, 87-95, 1981.

[7] W. L. Goffe, G. Ferrier, and J. John Rogers, "Global optimization of statistical functions with simulated annealing". *Journal of Econometric*, 60 (1/2), 65-100, 1994

[8] C. C. Kokonenji, "On Strict Arcsine Distribution". *Communications in Statistics. Theory Methods*.33(5), 99A3-1006, 2004.

[9] C. C. Kokonendji, and S. Marque, "A strict arcsine regression model". *African Diaspora Journal of Mathematics. Volume 4, Issue 3.* Ed by D. Toka, M. N. Gaston and Z. Said. NOVA Publisher, 2007.

[10] G. Letac, and M. Mora, "Natural real exponential families with cubic variance functions" . *Ann. Statist*., 18, 1-37, 1990.

[11] Y. N. Phang, and E. R. Loh. *Proceedings: IASC 2008:* Joint Meeting of 4th World Conference of the IASC and 6th Conference of the IASC and 6th conference of the Asian Regional Section of the IASC on Computational Statistic and Data Analysis, Yokohama, Japan, 2008.

[12] A. M. Mood, F. A. Graybill, and S. D. Boes, " Introduction to the Theory of Statistic". *McGraw-Hill Series in Probability and Statistics, 1913*