Zero Truncated Strict Arcsine Model
Y. N. Phang and E. F. Loh

Abstract—The zero truncated model is usually used in modeling count data without zero. It is the opposite of zero inflated model. Zero truncated Poisson and zero truncated negative binomial models are discussed and used by some researchers in analyzing the abundance of rare species and hospital stay. Zero truncated models are used as the base in developing hurdle models. In this study, we developed a new model, the zero truncated strict arcsine model, which can be used as an alternative model in modeling data without zero and with extra variation. Two simulated data sets are used and fitted into this developed model. The results show that the model provides a good fit to the data. Maximum likelihood estimation method is used in estimating the parameters.

Keywords—Hurdle models, maximum likelihood estimation method, positive count data.

I. INTRODUCTION

Zero truncated model is commonly used in modeling count data without zero. Examples of these data are number of hospital stay, number of times a voter has voted among the people who has voted during the general election. The most popular ones are zero truncated Poisson (ZIP) [1]-[3] and zero truncated negative binomial models (ZTNB). ZTNB is recommended for count data without zero but with extra variations. It has been used to analyze the abundance of rare species [4], [5]. Reference [6] analyzed the overdispersed positive count data of ischemic stroke hospitalizations using truncated negative binomial mixed regression model. Reference [7] applied the truncated model to a number of recreational fishing trips taken from a sample of Alaskan fishermen. Reference [8] applied it to the [9] data set on contract strikes. Reference [10] discussed zero-truncated Poisson-Lidley distribution and its applications. Reference [11] gave modeling Sage data with a truncated gamma-Poisson model. Another application of zero truncated models is it can be used as a building block for hurdle models. In this study, we developed a new model named zero truncated strict arcsine model which can be used as an alternative model in modeling data without zero and with extra variations. We fit the developed model into one real life and two simulated data sets. The chi-square value shows that it provides a good fit to the three data sets. Maximum likelihood estimation method is used in estimating the parameters in this model. The properties and characteristics of strict arcsine and zero truncated strict arcsine models are provided in Section II. Section III studies the parameter estimation method used. Section IV discusses the application and the results of this developed model to three data sets. A short concluding remark is given in Section V.

II. PROPERTIES OF THE DISTRIBUTIONS

A. The Strict Arcsine Distribution

The SA distribution is introduced by [12]. Reference [13] studied the properties of the strict arcsine distribution and found that the SA distribution is overdispersed, skewed to the right and leptokurtic.

The probability mass function of SA is given by

\[ \Pr(SA(x)) = \frac{A(x; \alpha)}{x!} \alpha^x \exp\left(-\alpha \arcsin(p)\right), \quad x = 0, 1, 2, \ldots \]  

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

\[ A(x; \alpha) = \begin{cases} \alpha^x / (x! \alpha^x) & \text{if } x = 2z; \text{ and } A(0, \alpha) = 1 \\ \frac{1}{\alpha} \left( \sum_{i=0}^{z-1} (\alpha^2 + 4i^2) \right) & \text{if } x = 2z + 1; \text{ and } A(1, \alpha) = \alpha \end{cases} \]  

The recurrence formula of SA is

\[ \Pr(x + 1) = \frac{A(x + 1; \alpha)}{A(x; \alpha)} \frac{x}{x + 1} \Pr(x), \quad x = 0, 1, 2, \ldots \]  

with

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

The likelihood \( L \) is given by

\[ L_{SA} = \prod_{k=0}^{x} \Pr_{SA}(k)^{F_k}, \quad x = 0, 1, 2, \ldots \]

and the log-likelihood is

\[ \ln L_{SA} = \sum_{k=0}^{x} F_k \ln \Pr_{SA}(k). \]

The likelihood score functions are given below.

---

Y. N. Phang is with the 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 the Academy of Language Studies, Universiti Teknologi MARA, Kampus Bandaraya Melaka, Malaysia (e-mail: david_loh@melaka.uitm.edu.my).
\[
\frac{\ell_{ZTSA}}{\ell_{\alpha}} = \sum_{k=0}^{x} F_k \frac{\partial \log A(k, \alpha)}{\partial \alpha} \arcsin(p), \quad x = 0, 1, 2, \ldots \quad (6)
\]

where

\[
\frac{\partial \log A(k, \alpha)}{\partial \alpha} = \begin{cases} 
\sum_{i=0}^{\alpha} \frac{2 \alpha}{(\alpha^2 + 4 k^2)}, & \text{if } x = 2 \alpha; \\
\sum_{i=0}^{\alpha} \frac{2(2k+1)^2 - 2 \alpha^2}{(\alpha^2 + (2k+1)^2)^2}, & \text{if } x = 2 \alpha + 1; \\
0, & \text{otherwise}
\end{cases}
\]

\[
\frac{\partial \log A(0, \alpha)}{\partial \alpha} = 0
\]

\[
\frac{\ell_{ZTSA}}{\ell_{p}} = \sum_{k=0}^{x} F_k \left[ \frac{k \alpha}{p \sqrt{1 - p^2}} \right], \quad x = 0, 1, 2, \ldots \quad (7)
\]

\[
\mu = ap(1 - p^2)^{-1/2} = E(X)
\]

\[
\alpha = \mu(1 - p^2)^{-1/2}
\]

\[
\sigma^2 = \alpha p(1 - p^2)^{-3/2} = VAR[X]
\]

B. The Zero Truncated Strict Arcsine Model

The zero truncated strict arcsine model (ZTSA) is developed by dividing the probability mass function of strict arcsine by 1-P(0).

The PMF for zero truncated strict arcsine model is given by

\[
P_{ZTSA}(Y = y) = Pr_{SA}(y)/(1 - Pr_{SA}(0)), \quad y = 1, 2, 3, \ldots \quad (9)
\]

The likelihood \(L\) is given by

\[
L_{ZTSA} = \prod_{k=1}^{x} Pr_{ZTSA}(k)^{x_k}, \quad x = 1, 2, \ldots \quad (10)
\]

and the log-likelihood is

\[
\ln L_{ZTSA} = \sum_{k=1}^{x} F_k \ln Pr_{ZTSA}(k) \quad (11)
\]

III. PARAMETER ESTIMATION

The maximum likelihood method together with simulated annealing [14], a global optimization routine, is used to obtain the unknown parameters for this model. Simulated annealing has been used in various combinatorial optimization problems. It is particularly good in solving problems involving circuit design [15]. The applied concept is based on the manner in which liquid freeze or metals recrystalize in the process of annealing. 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.

We find the parameter estimates for a real life and two simulated data sets using the above-mentioned method. Table I shows the real life data set from [6]. Table II shows simulated data set with \( p = 0.40, \alpha = 8.00, \) and sample size=500. Table III shows the data set simulated with \( p = 0.70, \alpha = 5.00 \) and sample size=1000.

<table>
<thead>
<tr>
<th>Observed frequency</th>
<th>Expected frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>554</td>
</tr>
<tr>
<td>2</td>
<td>99</td>
</tr>
<tr>
<td>3</td>
<td>18</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
</tr>
</tbody>
</table>

\(-\log\text{likelihood} = 407.29\)

\(\chi^2 = 2.10\)

\(\hat{p}_{ZTSA} = 0.3361, \quad \hat{\alpha}_{ZTSA} = 1.0347\)

<table>
<thead>
<tr>
<th>Simulated frequency</th>
<th>Expected frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>62</td>
</tr>
<tr>
<td>2</td>
<td>102</td>
</tr>
<tr>
<td>3</td>
<td>107</td>
</tr>
<tr>
<td>4</td>
<td>89</td>
</tr>
<tr>
<td>5</td>
<td>62</td>
</tr>
<tr>
<td>6</td>
<td>38</td>
</tr>
<tr>
<td>7</td>
<td>21</td>
</tr>
<tr>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
</tr>
</tbody>
</table>

\(-\log\text{likelihood} = 1497.57\)

\(\chi^2 = 0.40\)

\(\hat{p}_{ZTSA} = 0.3605, \quad \hat{\alpha}_{ZTSA} = 8.9318\)

IV. RESULTS

The newly developed zero truncated strict arcsine model provides very good fit for the three data sets. The chi-values \(\chi^2\)
= 2.10 for the first real life data set (Table I), $\chi^2 = 0.40$ for the second simulated data set (Table II) and $\chi^2 = 2.47$ for the third simulated data set (Table III) indicate that zero truncated strict arcsine model can be used as an alternative model in modeling overdispersed positive count data. The results also show that the estimation method used provides very good estimates which are very close to the set parameters.

Table III

<table>
<thead>
<tr>
<th></th>
<th>Simulated frequency</th>
<th>Expected frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>74</td>
<td>72.63</td>
</tr>
<tr>
<td>2</td>
<td>132</td>
<td>129.31</td>
</tr>
<tr>
<td>3</td>
<td>157</td>
<td>159.04</td>
</tr>
<tr>
<td>4</td>
<td>154</td>
<td>156.43</td>
</tr>
<tr>
<td>5</td>
<td>131</td>
<td>133.71</td>
</tr>
<tr>
<td>6</td>
<td>103</td>
<td>104.44</td>
</tr>
<tr>
<td>7</td>
<td>77</td>
<td>76.92</td>
</tr>
<tr>
<td>8</td>
<td>55</td>
<td>54.49</td>
</tr>
<tr>
<td>9</td>
<td>39</td>
<td>37.59</td>
</tr>
<tr>
<td>10</td>
<td>27</td>
<td>25.48</td>
</tr>
<tr>
<td>11</td>
<td>18</td>
<td>17.05</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td>11.31</td>
</tr>
<tr>
<td>13</td>
<td>8</td>
<td>7.46</td>
</tr>
<tr>
<td>14</td>
<td>6</td>
<td>4.90</td>
</tr>
<tr>
<td>15</td>
<td>4</td>
<td>3.20</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>6.03</td>
</tr>
</tbody>
</table>

$log$likelihood 2355.75  
$\chi^2$ 2.47

$\hat{p}_{ZTS}\alpha = 0.6778$, $\hat{\alpha}_{ZTS\alpha} = 5.2538$

V. CONCLUDING REMARKS

Zero truncated models are commonly used in modeling positive count data. It is used as a base in developing hurdle model. Zero truncated strict arcsine model is found to be suitable in modeling positive count data with extra variations. The results indicate that this model fits well into two simulated and one real life data sets. This study also shows that maximum likelihood estimation method together with simulated annealing is a good choice in estimating the parameters in this model.

ACKNOWLEDGMENT

This research is funded by the Fundamental Research Grant Scheme (FRGS), Ministry of Higher Education Malaysia, that is managed by the Research Management Institute, Universiti Teknologi MARA (600-RMI/ST/FRGS 5/3/Fst (217/2010).

REFERENCES
