

# On the Markov Chain Binomial Model

M. N. Islam<sup>1</sup>, C. D. O'shaughnessy<sup>2</sup>

<sup>1</sup>Department of Mathematics and Statistics, Grant MacEwan University, Edmonton, Canada

<sup>2</sup>Department of Mathematics and Statistics, University of Saskatchewan, Saskatoon, Canada

Email: IslamM@macewan.ca, oshaughn@math.usask.ca

Received May 15, 2013; revised June 15, 2013; accepted June 23, 2013

Copyright © 2013 M. N. Islam, C. D. O'shaughnessy. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## ABSTRACT

Rudolfer [1] studied properties and estimation of a  $\{0,1\}$ -state Markov chain binomial (MCB) model of extra-binomial variation. The variance expression in Lemma 4 is stated without proof but is incorrect, resulting in both Lemma 5 and Theorem 2 also being incorrect. These errors were corrected in Rudolfer [2]. In Sections 2 and 3 of this paper, a new derivation of the variance expression in a setting involving the natural parameters  $(n, p, \delta)$  is presented and the relation of the MCB model to Edwards' [3] probability generating function (pgf) approach is discussed. Section 4 deals with estimation of the model parameters. Estimation by the maximum likelihood method is difficult for a larger number  $n$  of Markov trials due to the complexity of the calculation of probabilities using Equation (3.2) of Rudolfer [1]. In this section, the exact maximum likelihood estimation of model parameters is obtained utilizing a sequence of Markov trials each involving  $n$  observations from a  $\{0,1\}$ -state MCB model and may be used for any value of  $n$ . Two examples in Section 5 illustrate the usefulness of the MCB model. The first example gives corrected results for Skellam's Brassica data while the second applies the "sequence approach" to data from Crouchley and Pickles [4].

**Keywords:** Extrabinomial Variation; Markov Chain Binomial Model; Maximum Likelihood Estimation; Sequence Data

## 1. Introduction

Let  $\{Z_i, i=1,2,\dots\}$  be a stationary two-state Markov chain with states 0 (failure) and 1 (success), an initial probability distribution denoted  $[q, p]$  in which  $p = P(Z_1 = 1)$  and  $q = P(Z_1 = 0) = 1 - p$ , and transition probabilities  $p_{i,j} = P(Z_{m+1} = j | Z_m = i)$  for  $i, j = 0, 1$  and  $m = 1, 2, \dots$  represented in a transition probability matrix

$$P = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix}. \quad (1)$$

If  $\{Z_i, i=1,2,\dots\}$  is a sequence from a  $\{0,1\}$ -state Markov chain, let  $X = \sum_{i=1}^n Z_i$  denotes the number of successes in  $n$  Markov trials. The random variable  $X$  may be said to have a Markov chain binomial (MCB) distribution with parameters  $n$  and  $p$ , and transition matrix  $P$ .

Rudolfer [1] studied properties and estimation for this  $\{0,1\}$ -state Markov chain binomial model. A formula for computing the probabilities  $P(X = k), k = 0, 1, \dots, n$  is given as his Equation (3.2), and an expression for the variance of  $X$  is given as Equation (3.4) in his Lemma 4.

However, the variance expression is incorrect and hence Lemma 5 and Theorem 2 are also incorrect. These results were corrected in Rudolfer (2). Section 3 of this paper provides an alternative derivation of the correct versions of these results.

Even for relatively small values of  $n$ , computation of the probabilities  $P(X = k)$  using Rudolfer's Equation (3.2) involves tedious summations making maximum likelihood estimation for the model difficult. Using the Edwards' [3] formulation and an approach introduced by Devore [5] and results given by Edwards [6], exact maximum likelihood estimation of model parameters is obtained in Section 4 using a sequence of observations assumed to come from a  $\{0,1\}$ -state Markov chain with the transition probability matrix  $P$  given in (1) above.

Two examples are presented in Section 5 to illustrate the effect of the corrections and of parameter estimation utilizing the sequence nature of the data.

## 2. The Markov Chain Binomial Model

Rudolfer [1] writes the matrix of transition probabilities

(1) for the  $\{0,1\}$ -state Markov chain as

$$P = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix} \quad (2)$$

in which  $0 < \alpha, \beta < 1$ . For such a chain, the probability of a success at stage  $k \geq 2$ ,  $p_k = P(Z_k = 1)$ , depends on  $k$  and the initial probability distribution  $[q, p]$ . This is seen by noting that, for  $k \geq 2$ ,

$$\begin{aligned} p_k &= P(Z_k = 1) \\ &= P(Z_{k-1} = 1, Z_k = 1) + P(Z_{k-1} = 0, Z_k = 1) \\ &= (1-\beta)p_{k-1} + \alpha(1-p_{k-1}) \\ &= \alpha + (1-\alpha-\beta)p_{k-1}. \end{aligned}$$

Recursive application of this procedure results in

$$p_k = \frac{\alpha}{\alpha + \beta} + (1-\alpha-\beta)^{k-1} \left( p - \frac{\alpha}{\alpha + \beta} \right).$$

Since  $|1-\alpha-\beta| < 1$  it follows that, as  $k \rightarrow \infty$ ,  $p_k$  tends to the limit  $\alpha/(\alpha + \beta)$  independent of the initial state (0 or 1) and of the initial probability distribution  $[q, p]$ . Also, if the initial probability distribution has  $p = \alpha/(\alpha + \beta)$ , the probability of a success  $p_k$  is constantly equal to  $\alpha/(\alpha + \beta)$  throughout all trials.

If the probability of success at the first trial is  $p = P(Z_1 = 1) = \alpha/(\alpha + \beta)$  so that  $P(Z_1 = 0) = q = 1 - p = \beta/(\alpha + \beta)$ , and if  $\delta = 1 - \alpha - \beta$ , then the transition matrix (2) can be written as

$$P = \begin{bmatrix} q + p\delta & p - p\delta \\ q - q\delta & p + q\delta \end{bmatrix}. \quad (3)$$

Replacing Rudolfer's [1] parameter  $\delta$  with  $r$ , the transition matrix (3) is seen to be equivalent to the one for the MCB model in Edwards [3]. The parameter  $\delta$  in Rudolfer [1] has the same interpretation as that of  $r$  in Edwards [3], namely it is the first-order autocorrelation between  $Z_{k-1}$  and  $Z_k$  for  $k \geq 2$ . Thus, Rudolfer's and Edwards' MCB models are exactly the same probability model under re-parameterization with parameters  $n, p$  and  $\delta$ , and transition matrix  $P$  as given in (3). A random variable  $X$  with this distribution may be denoted by  $X \sim MCB(n, p, \delta, P)$ .

Edwards' [3] approach involved the use of the probability generating function (pgf). For given  $n$ , the pgf of the MCB model is

$$G(s) = (qps)P^{n-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad (4)$$

in which  $s$  is a dummy variable. Explicit expressions for the probabilities  $P(X = k), k = 1, 2, \dots, n$  can be obtained from the pgf (4), although the calculations are not straightforward. For example, in the case of  $n = 3$ , the probabilities are

$$P(X = k) = \begin{cases} q(q + \delta p)^2 & (k = 0) \\ pq(1-\delta)\{q(1-\delta) + 2(q + \delta p)\} & (k = 1) \\ pq(1-\delta)\{p(1-\delta) + 2(p + \delta q)\} & (k = 2) \\ p(p + \delta q)^2 & (k = 3) \end{cases}$$

These equations can also be obtained by substituting  $\alpha = p(1-\delta)$  and  $\beta = q(1-\delta)$  into the expressions for the probabilities as given by Rudolfer [1] at the beginning of Section 4.2 of his paper.

### 3. Rudolfer's [1,2] Corrected Results

Rudolfer's [1] Lemma 4 incorrectly expresses the variance of a random variable  $X \sim MCB(n, p, \delta, P)$  as

$$Var[X] = npq + \frac{2\alpha\beta\delta^2}{(\alpha + \beta)^4} \{1 - n\delta^{n-1} + (n-1)\delta^n\}$$

resulting in an incorrect statement of his Lemma 5 and thus Theorem 2. In Rudolfer [2] the variance expression is corrected and proper statements of Lemma 5 and Theorem 2 are given. A corrected version of Lemma 4 is presented with an alternative derivation, and is followed by the corrected Theorem 2.

#### Corrected Lemma 4 of Rudolfer [1]

For a random variable  $X \sim MCB(n, p, \delta, P)$  with  $p = \alpha/(\alpha + \beta)$  and  $\delta = 1 - \alpha - \beta$ , the mean and variance are  $E(X) = np$  and

$$Var(X) = npq + 2pq \frac{\delta}{(1-\delta)^2} [n(1-\delta) - 1 + \delta^n] \quad (5)$$

**Proof:** The  $n$ -th step transition probability matrix is equal to

$$P^n = \begin{pmatrix} q + p\delta^n & p - p\delta^n \\ q - q\delta^n & p + q\delta^n \end{pmatrix} \quad (6)$$

as is easily shown by diagonalizing the transition matrix  $P$  given in (3), using its right and left eigenvectors and its eigenvalues. For  $i = 1, 2, \dots, n-1$  and  $h = 1, 2, \dots, n-i$ , it follows from (6) that

$$\begin{aligned} cov(Z_i, Z_{i+h}) &= P(Z_i = 1, Z_{i+h} = 1) - p^2 \\ &= p(p + q\delta^h) - p^2 = pq\delta^h \end{aligned}$$

and thus

$$\begin{aligned} Var(X) &= \sum_{i=1}^n Var(Z_i) + 2 \sum_{i=1}^{n-1} \sum_{h=1}^{n-i} cov(Z_i, Z_{i+h}) \\ &= P(Z_i = 1, Z_{i+h} = 1) - p^2 = npq + 2pq \sum_{i=1}^{n-1} (n-1)\delta^i \\ &= npq + 2pq \frac{\delta}{(1-\delta)^2} [n(1-\delta) - 1 + \delta^n]. \end{aligned}$$

**Corrected Theorem 2 of Rudolfer [1]**

For  $n \geq 2$  and  $X \sim MCB(n, p, \delta, P)$ ,

$$Var(X) \begin{cases} > npq & \text{if } 0 < \delta < 1 \\ < npq & \text{if } -1 < \delta < 0. \end{cases}$$

where  $\sigma_B^2 = np(1-p)$ .

**Proof:** From the variance expression (5), let

$$h(\delta) = \frac{\delta}{(1-\delta)^2} [n(1-\delta) - 1 + \delta^n]$$

$$= \delta [(n-1) + (n-2)\delta + (n-3)\delta^2 + \dots + 2\delta^{n-3} + \delta^{n-2}]. \tag{7}$$

From (7) it follows that

$$\begin{aligned} -(n-1)/2 < h(\delta) < 0 & \text{ if } -1 < \delta < 0, n \text{ odd} \\ -n/2 < h(\delta) < 0 & \text{ if } -1 < \delta < 0, n \text{ even} \\ 0 < h(\delta) < n(n-1)/2 & \text{ if } 0 < \delta < 1 \end{aligned}$$

which immediately gives the result in the theorem.

**4. Estimating the Model**

Consider  $N$  trials from a  $\{0,1\}$ -state Markov chain with parameters  $n, p$  and  $\delta$ . Let  $X_i = \sum_{j=1}^n Z_{ij}$ . Then  $X_i, i = 1, 2, \dots, N$  are independent, identically distributed random variables each having the MCB distribution with parameters  $n, p$  and  $\delta$ .

**4.1. Moment Estimation of  $p$  and  $\delta$**

Let  $X_1, X_2, \dots, X_N$  be a random sample of size  $N$  from the  $X \sim MCB(n, p, \delta, P)$  distribution, and let  $\bar{x} = \sum f_k k / \sum f_k$  and  $m_2 = \sum f_k k^2 / \sum f_k$  be the sample mean and sample second raw moment respectively. Using the method of moments, estimators  $\hat{p}$  and  $\hat{\delta}$  are determined as solutions to the equations

$$np - \bar{x} = 0 \tag{8}$$

and

$$npq + 2pq \frac{\delta}{(1-\delta)^2} [n(1-\delta) - 1 + \delta^n] + n^2 p^2 - m_2 = 0. \tag{9}$$

Equation (8) gives  $\hat{p} = \bar{x}/n$ . Replacing  $p$  by  $\hat{p}$  in Equation (9) and rearranging gives

$$\begin{aligned} 2\hat{p}(1-\hat{p})\delta^{n+1} - \delta^2 [n\hat{p}(1-\hat{p}) - n^2\hat{p}^2 + m_2] \\ + 2\delta [m_2 - n^2\hat{p}^2 - \hat{p}(1-\hat{p})] + n\hat{p}(1-\hat{p}) + n^2\hat{p}^2 - m_2 = 0 \end{aligned}$$

This nonlinear equation can be solved for  $\hat{\delta}$  by iteration using MAPLE or other mathematical software.

**4.2. Maximum Likelihood Estimation of  $p$  and  $\delta$  for Sequence Data**

Even for a relatively small value of  $n$ , maximum likeli-

hood estimation of the model parameters  $p$  and  $\delta$  may prove to be difficult using either Rudolfer's [1] Equation (3.2) (in this case, the summation in the expression for the probabilities is complex) or Edwards' [3] pgf approach. However, if  $N$  sequences  $\{Z_j, j = 1, 2, \dots, n\}$  each involving  $n$  trials from a  $\{0,1\}$ -state Markov chain with transition matrix (1) are observed, an alternative approach to obtaining maximum likelihood estimates of the parameters  $p$  and  $\delta$  is as follows.

The likelihood function for such sequence data can be written in the form

$$L = p^{N_1} (1-p)^{N_0} \prod_{i,j=0}^1 p_{ij}^{N_{ij}} \tag{10}$$

in which  $N_1$  is the number of the  $N$  sequences that begin with the occurrence of state 1 (that is, success),  $N_0 = N - N_1$  is the number of sequences that begin with the occurrence of state 0 (that is, failure), and  $N_{ij}$  denotes the total number of transitions from state  $i$  to state  $j$  observed within the  $N$  sequences. Since each of the  $N$  sequences involves  $n$  trials, a total of  $nN$  transitions will be observed. If the transition probabilities are written in terms of  $p, q = 1-p$  and  $\delta$  as in the version of the transition matrix  $P$  in expression (3), the likelihood function (10) becomes

$$\begin{aligned} L(p, \delta) = p^{N_1+N_{01}} q^{N_0+N_{10}} (p+q\delta)^{N_{11}} \\ \times (q+p\delta)^{N_{00}} (1-\delta)^{N_{01}+N_{10}}. \end{aligned} \tag{11}$$

Devore [5] obtained maximum likelihood solutions by ignoring the contribution of the initial state  $Z_1$ , and considering only the modified likelihood function

$$L^*(p, \delta) = p^{N_{01}} q^{N_{10}} (p+q\delta)^{N_{11}} (q+p\delta)^{N_{00}} (1-\delta)^{N_{01}+N_{10}} \tag{12}$$

The solutions obtained restricting attention to this modified likelihood function are

$$\hat{p} = \frac{1 - \hat{p}_{00}}{2 - \hat{p}_{00} - \hat{p}_{11}} \tag{13}$$

$$\hat{\delta} = \hat{p}_{00} + \hat{p}_{11} - 1 \tag{14}$$

in which  $\hat{p}_{00} = N_{00}/(N_{00} + N_{01})$  and

$$\hat{p}_{11} = N_{11}/(N_{10} + N_{11}).$$

Edwards [6] provides a solution using the full likelihood function (11). Equating each of the partial derivatives  $\partial \log L / \partial p$  and  $\partial \log L / \partial \delta$  of the log-likelihood function to zero and solving for  $\delta$  provides the two expressions

$$\hat{\delta} = 1 - \frac{N_{10} + N_0 p - N_1 q}{q(N_{10} + N_{11} + N_0 q - N_1 p)} \tag{15}$$

and

$$\hat{\delta} = 1 - \frac{N_{01} + N_1q - N_0p}{p(N_{00} + N_{01} + N_1q - N_0p)} \tag{16}$$

Equating expressions (15) and (16), and letting  $a = N_{01} + N_{11} - N_1$ ,  $b = N_1 + N_{01}$ ,  $c = N_1 + N_{00} + N_{01}$  and  $d = N_{10} - N_1$  results in the cubic equation in  $p$

$$2N^2p^3 - N(N - a + b + c - d)p^2 - [ab + cd + N(a - b)]p + ab = 0 \tag{17}$$

which can be solved by standard methods. Substituting the solution  $\hat{p}$  from (17) into either (15) or (16) provides the solution  $\hat{\delta}$ .

### 5. Examples

#### 5.1. Application to Skellam's [7] Brassica Data

Rudolfer [1,2] used the "Brassica" data of Skellam [7] to illustrate the model and parameter estimation. Using this data and the method of moments with result (5) of the corrected Lemma 4, the obtained estimate  $\hat{p} = 0.5806$  is essentially the same as the maximum likelihood estimate  $\hat{p} = 0.5807$ . The corrected moment Equation (9) for the variance in the case of  $n = 3$  is

$$2p(1-p)\delta^2 + 4p(1-p)\delta + 3p(1-p) + 9p^2 - m_2 = 0 \tag{18}$$

On substitution of  $\hat{p} = 0.5806$ , the quadratic Equation (18) gives the possible root  $\hat{\delta} = 0.1218$ , rejecting the other root as it is less than  $-1$ .

Comparisons of the  $p$ -values and of the values of the Chi-square statistics given in **Table 1** suggest that the Markov chain binomial model fitted by the method of moments using the corrected variance Formula (5) gives a fit very similar to those provided by the maximum likelihood method in Rudolfer [1] and the corrected method of moments method in Rudolfer [2].

#### 5.2. Application to Human Birth-Order Data

Using the solutions provided by Devore's [5] modified likelihood function and by Edwards' [6] full likelihood function, the MCB model is fitted to a case in which a sequence of observations is given. **Table 2** gives birth-order data for families with four children from Finland and from the United States. This data appears in Crouchley and Pickles [4]. It is assumed that the data are generated by a Markovian dependent sequence as defined in (1). The parameters  $p$  and  $\delta$  are the quantities of interest in the analysis.

Let state 0 refer to females and state 1 to males. For the Finish data, the full likelihood function, as given by (10) and then (11), is

**Table 1. Fits of the MCB model to Skellam's [7] Brassica data.**

$k$	$f_k$	$MCB^{(a)}$	$MCB^{(b)}$	$MCB^{(c)}$	$MCB^{(d)}$
0	32	135.08	33.98	33.94	34.13
1	103	93.41	97.13	97.19	97.05
2	122	66.73	127.73	127.81	127.51
3	80	41.78	78.16	78.06	78.31
Total	337	337.00	337.00	337.00	337.00
$\chi^2$		35.59	0.77	0.77	0.77
D.F.		1	1	1	1
$p$ -value		0.00	0.38	0.38	0.38
$\hat{p}$		0.5806	0.5807	0.5806	0.5806
$\hat{\delta}$		0.3824	0.1223	0.1218	0.1240

<sup>(a)</sup> $MCB$  estimated by method of moments (Rudolfer [1]—with error), <sup>(b)</sup> $MCB$  estimated by maximum likelihood method (Rudolfer [1]), <sup>(c)</sup> $MCB$  estimated by method of moments (Rudolfer [2]—error corrected), <sup>(d)</sup> $MCB$  estimated by method of moments with variance Formula (5) in natural parameter setting.

**Table 2. Frequencies of birth-order for families of four children from Finland and the United States.**

Order	Finland	USA	Order	Finland	USA
MMMM	469	1133	FFFF	481	913
MMMF	484	1140	FFFM	442	952
MMFM	466	1106	FFMF	428	1010
MMFF	448	1046	FFMM	424	1028
MFMM	398	1105	FMFF	342	935
MFMF	448	1049	FMFM	406	1071
MFFM	420	1094	FMMF	406	1019
MFFF	425	982	FMMM	419	1085

$$L(p, \delta) = p^{3558} (1-p)^{3348} p_{00}^{5239} p_{01}^{5005} p_{10}^{5119} p_{11}^{5355} = p^{8563} (1-p)^{8467} (1-\delta)^{10124} \times (p + (1-p)\delta)^{5355} (1-p + p\delta)^{5239}.$$

From the latter form, the maximum likelihood estimates are determined to be  $\hat{p} = 0.5039$  and  $\hat{\delta} = 0.02266$ , and the estimated matrix of transition probabilities is

$$\hat{P} = \begin{bmatrix} 0.5078 & 0.4922 \\ 0.4848 & 0.5152 \end{bmatrix}$$

Using Devore's [5] modified likelihood function (12), the estimated matrix of transition probabilities is

$$\hat{P} = \begin{bmatrix} 0.5114 & 0.4886 \\ 0.4887 & 0.5113 \end{bmatrix}$$

from which, applying Equations (13) and (14),  
 $\hat{p} = 0.4999$  and  $\hat{\delta} = 0.0227$ .

Similarly, for the case of the United States population the estimated parameters from the full likelihood function are  $\hat{p} = 0.5150$  giving as the estimated matrix of transition probabilities

$$\hat{P} = \begin{bmatrix} 0.4820 & 0.5180 \\ 0.4879 & 0.5121 \end{bmatrix}$$

Using Devore's method, the estimated matrix of transition probabilities is

$$\hat{P} = \begin{bmatrix} 0.4834 & 0.5166 \\ 0.4898 & 0.5106 \end{bmatrix}$$

so that  $\hat{p} = 0.5135$  and  $\hat{\delta} = -0.0060$ .

## 6. Discussion

In the application of the Markov chain binomial model to the Brassica data of Skellum [7] using the method of moments, **Table 2** illustrates that the variance expression (5) results in a fit little different from that provided by the method of moments using Rudolfer's [2] corrected values.

The second example dealing with birth-order data presented in Crouchley and Pickles [4] illustrates the usefulness of the "sequence data approach" in finding maximum likelihood estimates for the MCB model. Very little difference is seen in the estimates in this approach between using the full likelihood equation solution discussed by Edwards [6] and using the modified likelihood

equation as discussed by Devore [5].

## REFERENCES

- [1] S. M. Rudolfer, "A Markov Chain Model of Extrabinomial Variation," *Biometrika*, Vol. 77, No. 2, 1990, pp. 255-264. <http://dx.doi.org/10.1093/biomet/77.2.255>
- [2] S. M. Rudolfer, "Correction to a Markov Chain Model of Extrabinomial Variation," *Biometrika*, Vol. 78, No. 4, 1991, p. 935. <http://dx.doi.org/10.2307/2336950>
- [3] A. W. F. Edwards, "The Meaning of Binomial Distribution," *Nature (London)*, Vol. 186, 1960, p. 1074. <http://dx.doi.org/10.1038/1861074a0>
- [4] R. Crouchley and A. R. Pickles, "Methods for the Identification of Lexian, Poisson and Markovian Variations in the Secondary Sex Ratio," *Biometrics*, Vol. 40, No. 1, 1984, pp. 165-175. <http://dx.doi.org/10.2307/2530755>
- [5] J. L. Devore, "A Note on the Estimation of Parameters in a Bernoulli Model with Dependence," *Annals of Statistics*, Vol. 4, No. 5, 1976, pp. 990-992. <http://dx.doi.org/10.1214/aos/1176343597>
- [6] A. W. F. Edwards, "Estimation of the Parameters in Short Markov Sequences," *Journal of the Royal Statistical Society, Series B*, Vol. 25, No. 1, 1963, pp. 206-208.
- [7] J. G. Skellam, "A Probability Distribution Derived from the Binomial Distribution by Regarding the Probability of Success as Variable between the Sets of Trials," *Journal of the Royal Statistical Society, Series B*, Vol. 10, No. 2, 1948, pp. 257-261.