

Research Article

Estimation of Log-Linear-Binomial Distribution with Applications

Elsayed Ali Habib^{1,2}

¹ Department of Mathematics and Statistics, Faculty of Commerce, Benha University, Al Qalyubiyah, Egypt

² Management and Marketing Department, College of Business, University of Bahrain,
P.O. Box 32038, Bahrain

Correspondence should be addressed to Elsayed Ali Habib, shabib40@gmail.com

Received 14 April 2010; Revised 21 August 2010; Accepted 23 November 2010

Academic Editor: Daniel Zelterman

Copyright © 2010 Elsayed Ali Habib. 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.

Log-linear-binomial distribution was introduced for describing the behavior of the sum of dependent Bernoulli random variables. The distribution is a generalization of binomial distribution that allows construction of a broad class of distributions. In this paper, we consider the problem of estimating the two parameters of log-linear binomial distribution by moment and maximum likelihood methods. The distribution is used to fit genetic data and to obtain the sampling distribution of the sign test under dependence among trials.

1. Introduction

During the last three decades, a growing amount of literature has been observed in generalizing the classical discrete distributions. The main idea was to apply the extended versions of modeling different kinds of dependent count or frequency structure in various fields; see the work of Johnson et al. in [1], and of Bowman and George in [2], and of George and Bowman in [3], Yu and Zelterman [4, 5]. Failure to take account of correlation in the data will cause less precision for binomial-based estimates; see, for example, Kolev and Paiva [6].

As a generalization for the binomial distribution, Lovison [7] has derived the distribution of the sum of dependent Bernoulli random variables as an alternative of Altham's multiplicative-binomial distribution [8] from Cox's log-linear representation [9] for the joint distribution of n binary-dependent responses and it will be called log-linear-binomial distribution. This distribution is characterized by two parameters and provides wider range of distributions than are provided by the binomial distribution where the log-linear binomial distribution includes underdispersion, overdispersion models and the binomial distribution as a special case.

In this paper, the methods of moment and maximum likelihood are used to estimate the two parameters of Lovison's log-linear binomial distribution. The variance-covariance matrix for the estimated parameters is obtained. Log-linear binomial distribution is also used to fit genetic data, and to obtain the sampling distribution of the sign test under dependence among trials.

In Section 2, we define Lovison's log-linear binomial distribution. Estimation of the parameters of the distribution based on moments and maximum likelihood methods are derived in Section 3. Ungrouped data is studied in Section 4 as a special case. Two applications are given in Section 5.

2. Lovison's Log-Linear-Binomial Distribution

Consider the random vector $\mathbf{Z} = [Z_1, \dots, Z_n]'$, Z_i being a binary response which measures whether some event of interest is present, "1", or absent, "0" for a sample of n units and $Y_n = \sum_{i=1}^n Z_i$ denotes the sample frequency of successes. To accommodate the possible dependence among Z_i , and under the assumption that the units are exchangeable Lovison [7] has obtained the distribution of Y_n as

$$P(Y_n = y) = \frac{\binom{n}{y} \psi^y (1 - \psi)^{n-y} \omega^{y(n-y)}}{\sum_{t=0}^n \binom{n}{t} \psi^t (1 - \psi)^{n-t} \omega^{t(n-t)}}, \quad y = 0, 1, \dots, n, \quad (2.1)$$

where $0 < \psi < 1$ and $\omega > 0$ are the parameters; for more details about this distribution; see the work of Lovison in [7]. This distribution provides a wider range of distributions than is provided by the binomial distribution, for example, Figures 1 and 2 show the distribution of Y_n for $n = 10$, $\psi = 0.2, 0.5$, and different values of ω . For the values of $\omega > 1$, the distribution is sharper in the middle than the binomial.

As can be seen from the figures for some values of $\omega < 1$, the distribution can be U , bimodal and unimodal shapes. The expected value and the variance of Y_n are given by

$$\begin{aligned} \mu &= E(Y_n) = n\pi, \\ \sigma^2 &= V(Y_n) = n\pi(1 - \pi) + n(n - 1)(\pi_1 - \pi^2). \end{aligned} \quad (2.2)$$

Note that

$$V(Y_n) = V(Y_b) + n(n - 1)\text{Cov}(Z_k, Z_h) \quad (2.3)$$

for $k \neq h$ and Cov stands for covariance. The variance of the binomial is

$$V(Y_b) = n\pi(1 - \pi), \quad (2.4)$$

and the covariance of Z_k and Z_h is

$$\text{Cov}(Z_k, Z_h) = \pi_1 - \pi^2, \quad (2.5)$$

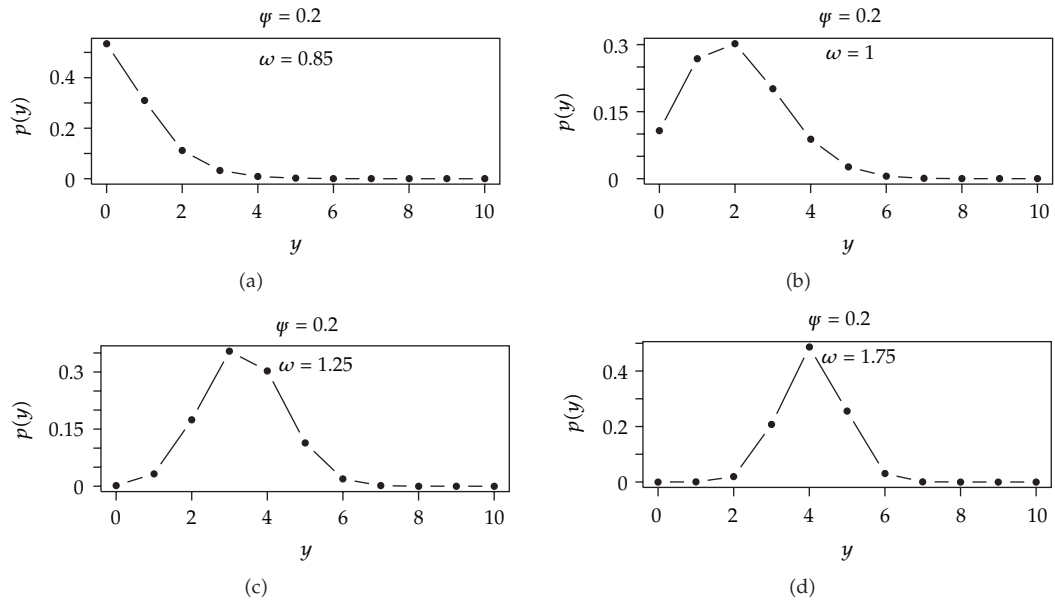


Figure 1: The distribution of Y_n for different values of ω , $\psi = 0.2$, and $n = 10$.

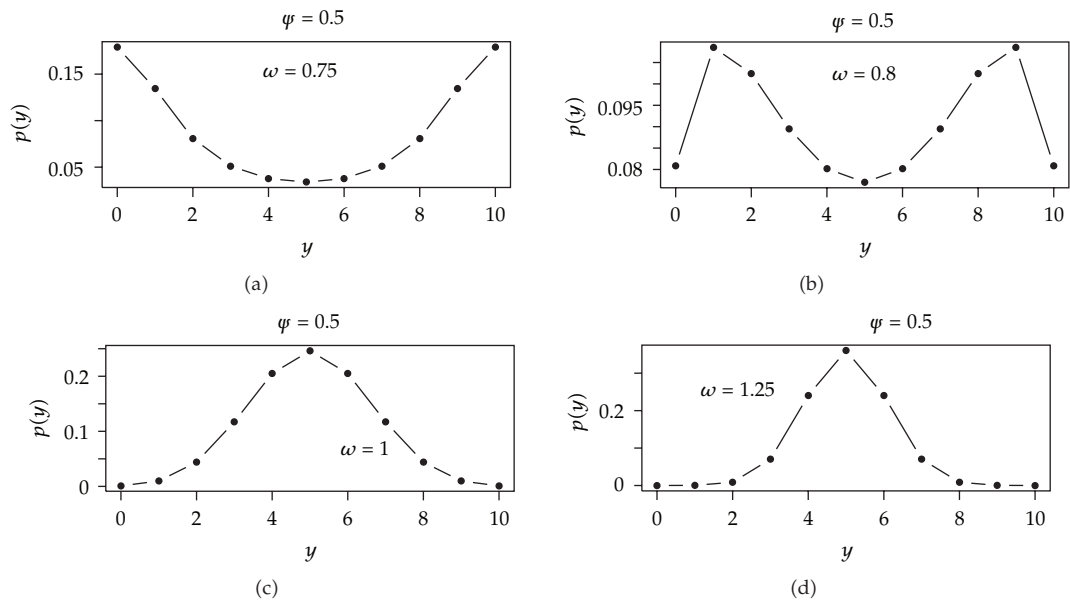


Figure 2: The distribution of Y_n for different values of ω , $\psi = 0.5$, and $n = 10$.

where

$$\begin{aligned} \pi &= \psi \frac{\kappa_{n-1}(\psi, \omega)}{\kappa_n(\psi, \omega)}, \\ \pi_1 &= \psi^2 \frac{\kappa_{n-2}(\psi, \omega)}{\kappa_n(\psi, \omega)}. \end{aligned} \tag{2.6}$$

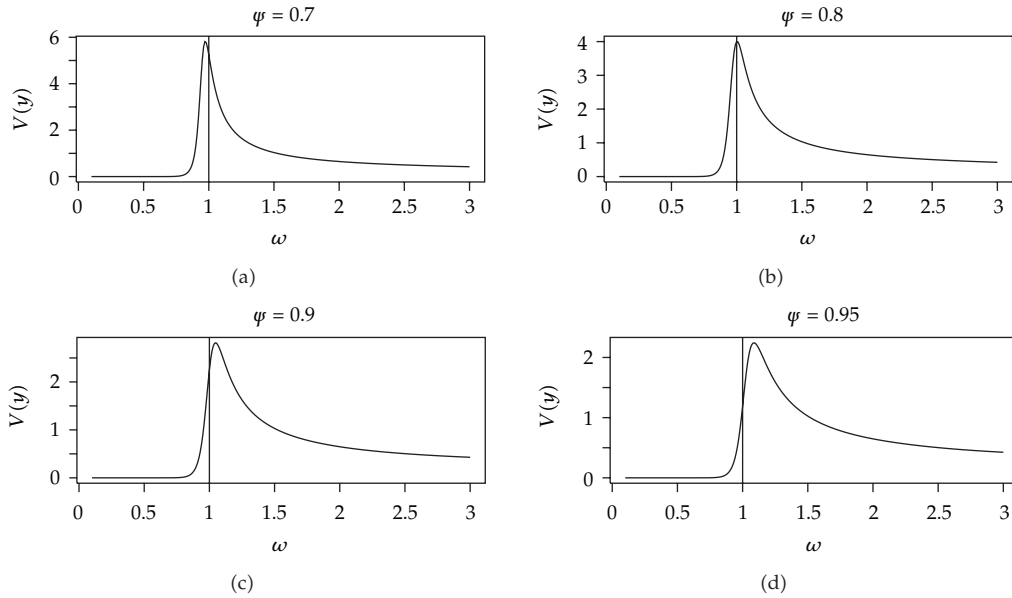


Figure 3: Variance of Y_n for various values of ω at each value of ψ and $n = 25$.

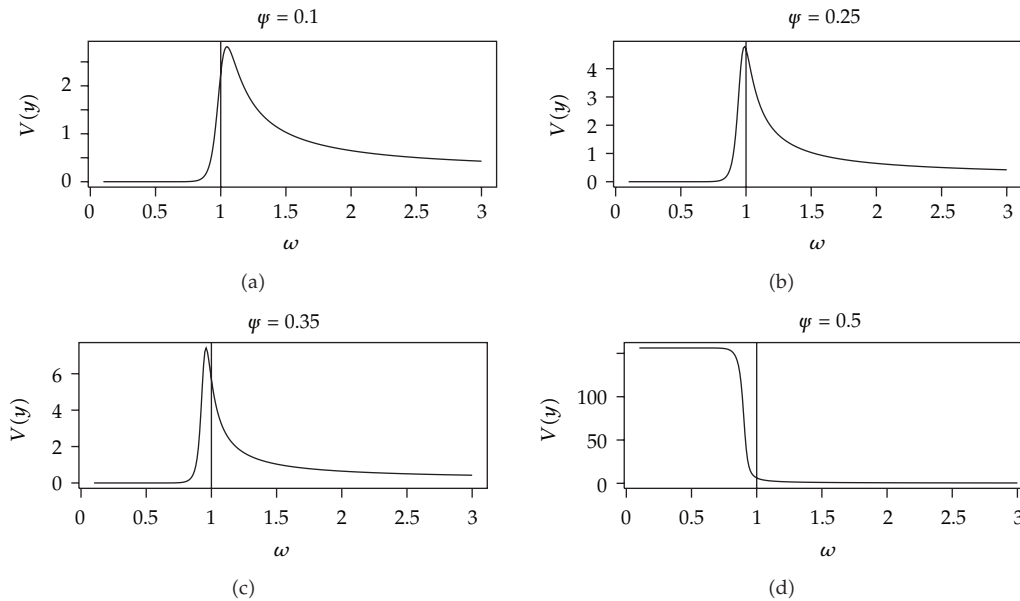


Figure 4: Variance of Y_n for various values of ω at each value of ψ and $n = 25$.

Therefore, the variance of the log-linear binomial is equal to the variance of binomial when $\text{Cov}(Z_k, Z_h) = 0$, greater than the variance of binomial when $\text{Cov}(Z_k, Z_h) > 0$, and less than the variance of binomial when $\text{Cov}(Z_k, Z_h) < 0$.

The expected value and the variance of Y_n are nonlinear on both ψ and ω . For example, the nonlinearity in the variance of Y_n is depicted in Figures 3 and 4.

3. Estimation of the Parameters

Let a random sample of R sets of n trials each be available, the number of given y successes being $f_y (y = 0, 1, 2, \dots, n)$, and $R = \sum_{y=0}^n f_y$.

3.1. Method of Moments

We can use the first two sample moments to find moment estimates for ψ and ω as follows. The first sample moment is

$$m_1 = \sum_{y=0}^n \frac{y f_y}{R},$$

$$M_2 = \sum_{y=0}^n \frac{f_y (y - m_1)^2}{R}. \quad (3.1)$$

Equating these sample moments to the corresponding population moments we obtain the estimates of ψ and ω by solving the two equations

$$m_1 = n\psi \frac{\kappa_{n-1}(\psi, \omega)}{\kappa_n(\psi, \omega)},$$

$$M_2 = n(n-1)\psi^2 \frac{\kappa_{n-2}(\psi, \omega)}{\kappa_n(\psi, \omega)} + n\psi \frac{\kappa_{n-1}(\psi, \omega)}{\kappa_n(\psi, \omega)} - n^2\psi^2 \left[\frac{\kappa_{n-1}(\psi, \omega)}{\kappa_n(\psi, \omega)} \right]^2. \quad (3.2)$$

The solution of these two equations needs numerical methods. The numerical solution of these two equations may be found by using nonlinear equation solver (*nleqslv*) in statistical *R*-software. The author has a program written in *R* for finding the moment estimates upon request.

3.2. Method of Maximum Likelihood

The method of maximum likelihood provides estimators that have a reasonable intuitive basis and many desirable statistical properties. The likelihood of the sample can be written as

$$L(\psi, \omega \mid n, y, f) = R! \prod_{y=0}^n \frac{[P(y; \psi, \omega)]^{f_y}}{f_y!}. \quad (3.3)$$

Take the logarithm of this function

$$l = \log R! + \sum_{y=0}^n f_y \log P(y; \psi, \omega) - \sum_{y=0}^n \log f_y! \quad (3.4)$$

The first partial derivatives with respect to ψ and ω are

$$\begin{aligned}\frac{\partial l}{\partial \psi} &= \sum_{y=0}^n f_y \frac{(\partial/\partial \psi)P(y; \psi, \omega)}{P(y; \psi, \omega)}, \\ \frac{\partial l}{\partial \omega} &= \sum_{y=0}^n f_y \frac{(\partial/\partial \omega)P(y; \psi, \omega)}{P(y; \psi, \omega)}.\end{aligned}\tag{3.5}$$

Let

$$\begin{aligned}q_0 &= \binom{n}{y} \psi^y (1-\psi)^{(n-y)} \omega^{y(n-y)}, \\ q_1 &= \sum_{t=0}^n \binom{n}{t} \psi^t (1-\psi)^{(n-t)} \omega^{t(n-t)}.\end{aligned}\tag{3.6}$$

Hence,

$$P(y; \psi, \omega) = \frac{q_0}{q_1}.\tag{3.7}$$

3.2.1. Estimation of ψ and ω

After simplification the first partial derivative for ψ and ω can be written as

$$\begin{aligned}\frac{\partial l}{\partial \psi} &= \frac{\sum_{y=0}^n y f_y}{\psi} - \frac{\sum_{y=0}^n (n-y) f_y}{1-\psi} - \frac{R q'_{1\psi}}{q_1}, \\ \frac{\partial l}{\partial \omega} &= \sum_{y=0}^n \frac{y(n-y) f_y}{\omega} - \frac{q'_{1\omega} R}{q_1}.\end{aligned}\tag{3.8}$$

where

$$\begin{aligned}q'_{1\psi} &= \sum_{t=0}^n \binom{n}{t} \psi^t (1-\psi)^{(n-t)} \omega^{t(n-t)} \left[\frac{t}{\psi} - \frac{n-t}{1-\psi} \right], \\ q'_{1\omega} &= \sum_{t=0}^n \binom{n}{t} t(n-t) \psi^t (1-\psi)^{(n-t)} \omega^{t(n-t)-1}.\end{aligned}\tag{3.9}$$

By solving $\partial l/\partial \psi = 0$ and $\partial l/\partial \omega = 0$ simultaneously we obtain the maximum likelihood estimates $\hat{\psi}_{ml}$ and $\hat{\omega}_{ml}$ under the sufficient condition $\partial^2 l/\partial \psi^2 < 0$ and $\partial^2 l/\partial \omega^2 < 0$.

The numerical solution for these two equations may be obtained using nonlinear equation solver (*nleqslv*) in *R*-software. The author has a program written in *R* for finding the maximum likelihood estimates upon request.

3.2.2. The Asymptotic Variances

Suppose that $\theta = [\psi \ \omega]'$ and under certain regularity conditions the information matrix is

$$I_{ij}(\theta) = -E \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log L(\theta) \right]. \quad (3.10)$$

The asymptotic variance-covariance matrix can be obtained as

$$V(\hat{\theta}) \approx [I_{ij}(\theta)]^{-1}. \quad (3.11)$$

To find the information matrix for the log-linear binomial distribution, we note that

$$\begin{aligned} \frac{\partial^2 l}{\partial \psi^2} &= -\frac{\sum_{y=0}^n y f_y}{\psi^2} - \frac{\sum_{y=0}^n (n-y) f_y}{(1-\psi)^2} - \left[\frac{q_1 q_{1\psi}'' R - q_{1\psi}'^2 R}{q_1^2} \right], \\ \frac{\partial^2 l}{\partial \omega^2} &= -\sum_{y=0}^n \frac{y(n-y) f_y}{\omega^2} - \left[\frac{q_1 q_{1\omega}'' R - q_{1\omega}'^2 R}{q_1^2} \right], \\ \frac{\partial}{\partial \psi} \frac{\partial l}{\partial \omega} &= -\frac{R q_1 q_{1\omega\psi}'' - R q_{1\omega}' q_{1\psi}'}{q_1^2}, \end{aligned} \quad (3.12)$$

where

$$\begin{aligned} q_{1\psi}'' &= \sum_{t=0}^n \binom{n}{t} \psi^t (1-\psi)^{(n-t)} \omega^{t(n-t)} \left[\frac{t(t-1)}{\psi^2} - \frac{2t(n-t)}{\psi(1-\psi)} + \frac{(n-t)(n-t-1)}{(1-\psi)^2} \right], \\ q_{1\omega}'' &= \sum_{t=0}^n \binom{n}{t} t(n-t)[t(n-t)-1] \psi^t (1-\psi)^{(n-t)} \omega^{t(n-t)-2}, \\ q_{1\omega\psi}'' &= \sum_{t=0}^n \binom{n}{t} t(n-t) \psi^t (1-\psi)^{(n-t)} \omega^{t(n-t)-1} \left[\frac{t}{\psi} - \frac{n-t}{1-\psi} \right]. \end{aligned} \quad (3.13)$$

Taking the expectation, the information matrix is obtained as

$$I(\theta) = \begin{bmatrix} \frac{R\mu}{\psi^2} + \frac{R(n-\mu)}{(1-\psi)^2} + \left[\frac{q_1 q_{1\psi}'' R - q_{1\psi}'^2 R}{q_1^2} \right] & \frac{R q_1 q_{1\omega\psi}'' - R q_{1\omega}' q_{1\psi}'}{q_1^2} \\ \frac{R q_1 q_{1\omega\psi}'' - R q_{1\omega}' q_{1\psi}'}{q_1^2} & \frac{R[\mu(n-\mu) - \sigma^2]}{\omega^2} + \left[\frac{q_1 q_{1\omega}'' R - q_{1\omega}'^2 R}{q_1^2} \right] \end{bmatrix}, \quad (3.14)$$

where $E[\sum_{y=0}^n y(n-y)f_y] = nE(\sum_{y=0}^n yf_y) - E(\sum_{y=0}^n y^2f_y) = nR\mu - R(\mu^2 + \sigma^2)$. The variance-covariance matrix will be

$$V(\hat{\theta}) \approx [I(\theta)]^{-1}. \quad (3.15)$$

The estimated variance-covariance matrix is

$$\hat{V}(\hat{\theta}) \approx [I(\hat{\theta})]^{-1}. \quad (3.16)$$

The author has a program written in *R* for finding the estimated variance-covariance matrix upon request.

4. Special Case: Ungrouped Data

If the values of Bernoulli random variables Z_i are known, the parameters ψ and ω can be estimated as follows. By noticing that in a vector of binary responses z there are $n(n-1)/2$ pairs of responses, and if the order is irrelevant three types of pairs are distinguishable: $(n-y)(n-y-1)/2$ pairs of $(z_k = 0, z_h = 0)$, $y(y-1)/2$ pairs of $(z_k = 1, z_h = 1)$, and $(n-y)y$ pairs of $(z_k = 0, z_h = 1)$, or $(z_k = 1, z_h = 0)$ where $y = \sum_{k=1}^n z_k$; see the work of Lovison in [7]. Therefore, the estimate of ω is

$$\hat{\omega} = \frac{1}{\sqrt{\widehat{\text{CPR}}}}. \quad (4.1)$$

The estimated cross-product ratio (CPR) is

$$\widehat{\text{CPR}} = \frac{0.25y(y-1)(n-y)(n-y-1)}{\#(z_k = 0, z_h = 1) \#(z_k = 1, z_h = 0)}. \quad (4.2)$$

To obtain the maximum likelihood estimate of ψ , we need to solve

$$\frac{dl}{d\psi} = \frac{y}{\psi} - \frac{(n-y)}{1-\psi} - \frac{q'_{1\psi}}{q_1} = 0. \quad (4.3)$$

The estimated variance of $\hat{\psi}$ can be obtained when $R = 1$.

5. Applications

5.1. Sampling Distribution of the Sign Test for Comparing Paired Sample

The sign test is a nonparametric test which makes very few assumptions about the nature of the distributions under test. It is for use with two repeated (or correlated) measures, and measurement is assumed to be at least ordinal. The usual null hypothesis for this test is that there is no difference between the two treatments (groups, g_k , $k = 1, 2$). Formally, let

$\tau = P(X > Y)$, and then test the null hypothesis $H_0: \tau = 0.5$ for no differences against $H_1: \tau \neq 0.5$ for differences. The sign test can be written as

$$s_d = \sum_{i=1}^n I(g_{1i} - g_{2i}) = \sum_{i=1}^n Z_i = Y_n, \quad (5.1)$$

where

$$Z_i = \begin{cases} 1, & g_{1i} - g_{2i} > 0 (+), \\ 0, & g_{1i} - g_{2i} < 0 (-). \end{cases} \quad (5.2)$$

Under the assumptions of two outcomes, fixed probability of success, and independent trials it is assumed that the sampling distribution of s_d is binomial distribution. The rejection region is

$$P_{\text{value}} = P(\text{reject } H_0 \mid H_0 \text{ is true}, n, y_d) = \sum_{j=0}^{y_d} \binom{n}{j} 0.5^n \quad (5.3)$$

and $y_d = \min(\#1, \#0)$. For a two-tailed test we reject H_0 if $2P_{\text{value}} \leq \alpha$, else we do not reject H_0 , where α is prespecified value.

5.1.1. Sign Test Under Dependence of the Trials

Suppose the assumption of mutual independence in the data is violated and the trials are dependent; see, for example, the work of Tallis in [10] and Luceño [11]. In this case, we suggest the log-linear binomial distribution as a sampling distribution of $s_d = \sum_{i=1}^n Z_i$ rather than the binomial distribution. Let $\tau = P(X > Y) = P(Z = 1) = \varphi(\kappa_{n-1}(\varphi, \omega) / \kappa_n(\varphi, \omega))$ represent the probability of success. Then the null hypothesis $H_0: \tau = 0.5$ is equivalent to $H_0: \varphi = 0.5$. Therefore, under the null hypothesis, the rejection region can be obtained as

$$P_{\text{value.lb}} = P(\text{reject } H_0 \mid H_0 \text{ is true}, n, y_d, \omega) = \frac{\sum_{j=0}^{y_d} \binom{n}{j} 0.5^n \omega^{j(n-j)}}{\sum_{t=0}^n \binom{n}{t} 0.5^n \omega^{t(n-t)}}, \quad (5.4)$$

where $y_d = \min(\#1, \#0)$ and $\omega = 1/\sqrt{\text{CPR}}$. For a two-tailed test, we reject H_0 if $2P_{\text{value.lb}} \leq \alpha$, else do not reject H_0 and α is a prespecified value (H_0 : there is no difference against H_1 : there is difference.)

Example 5.1. A physiologist wants to know if monkeys prefer stimulation of brain area A to stimulation of brain area B. In the experiment, 14 rhesus monkeys from the same family are taught to press two bars. When a light comes on, presses on Bar 1 result in stimulation of area A and presses on Bar 2 result in stimulation of area B. After learning to press the bars, the monkeys are tested for 15 minutes, during which time the frequencies for the two bars are recorded. The data are shown in Table 1.

Table 1: Number of bar presses in brain stimulation experiment.

Subject	bar 1	bar 2	Difference	Z_i (sign)
1	20	40	-20	0 (-)
2	18	25	-7	0 (-)
3	24	38	-14	0 (-)
4	14	27	-13	0 (-)
5	5	31	-26	0 (-)
6	26	21	5	1 (+)
7	15	32	-17	0 (-)
8	29	38	-9	0 (-)
9	15	25	-10	0 (-)
10	9	18	-9	0 (-)
11	25	32	-7	0 (-)
12	31	28	3	1 (+)
13	35	33	2	1 (+)
14	12	29	-17	0 (-)

The data are obtained from Weaver [12].

Using the binomial distribution, we have $n = 14$, $y_d = 3$, and P_{value} is

$$2P_{\text{value}} = \sum_{j=0}^3 \binom{14}{j} 0.5^{14} = 2 \times 0.0287 = 0.0574. \quad (5.5)$$

Since $0.0574 > 0.05$, we do not reject H_0 .

Using the log-linear binomial distribution, we have $n = 14$, $y_d = 3$, and

$$\omega = \frac{1}{\sqrt{(55)(3)/(25)(8)}} = 1.1, \quad (5.6)$$

and the P_{value} is

$$2P_{\text{value.lb}} = \frac{\sum_{j=0}^3 \binom{14}{j} 0.5^{14} 1.1^{j(14-j)}}{\sum_{t=0}^{14} \binom{14}{t} 0.5^{14} 1.1^{t(14-t)}} = 2(0.00687) = 0.0137, \quad (5.7)$$

where $0.0137 < 0.05$, so we reject H_0 . That is, we would conclude that monkeys prefer stimulation in brain area B to stimulation in area A. Note that the rejection of H_0 agrees with Wilcoxon Signed-Ranks test for the same data; see the work of Weaver in [12] and Siegel and Castellan in [13].

Power of the Test

Following the method given by Groebner et al. in [14], we may use the normal approximation to study the power of the sign test at $n = 14$, $\omega = 1.1$ and using one-tail test for simplicity as follows. The power of the test is

$$\begin{aligned} \text{power} &= 1 - \beta, \\ \beta &= P(\text{accept } H_0 \mid H_0 \text{ is false}). \end{aligned} \quad (5.8)$$

For one-tailed test $H_0: \tau \leq 0.5$ and $H_1: \tau > 0.5$ using binomial distribution with $n = 14$, $\tau = 0.5$ and $\sigma = \sqrt{14(0.5)(0.5)} = 1.8708$. The critical value with $\alpha = 0.05$ is

$$y_{0.05,b} = 14(0.5) + 1.645(1.8708) = 10.077. \quad (5.9)$$

The power of the test is

$$\text{power}_b \approx 1 - P\left(Z < \frac{10.077 - 14\tau_1}{1.8708}\right), \quad (5.10)$$

for $\tau_1 > 0.5$.

Using log-linear binomial distribution with $n = 14$, $\psi = 0.5$, and $\omega = 1.1$, we obtain $\sigma = 1.4616$, then under H_0 and $\alpha = 0.05$, the critical value is

$$y_{0.05,lb} = 14(0.5) + 1.645(1.4616) = 9.404. \quad (5.11)$$

The power of the test under log-linear binomial distribution is

$$\text{power}_{lb} \approx 1 - P\left(Z < \frac{9.404 - 14\tau_1}{1.4616}\right), \quad (5.12)$$

$\tau_1 > 0.5$.

Power of the sign test is given in Table 2. In this case, the log-linear binomial distribution shows improvement in the power of the sign test over the sign test under binomial; for example, when $\tau_1 = 0.70$, the power increases from 0.44 to 0.61 about 1.40 times.

5.2. Fitting Genetic Data

The data are taken from Salmaniyaa hospital records in Bahrain for a genetic study on the gender ratio. Table 3 shows the number of male children in 3475 families with 7 children.

The first two sample moments are

$$m_1 = 3.12057, \quad M_2 = 1.21567. \quad (5.13)$$

Table 2: Power of the sign test under binomial and log-linear binomial distribution for one-tail test with $n = 14$, $\omega = 1.1$, and $\alpha = 0.05$ based on normal approximation.

	τ_1					
	0.5	0.60	0.70	0.80	0.90	0.95
Binomial distribution						
Z	1.64	0.89	0.15	-0.60	-1.35	-1.72
Power	0.05	0.18	0.44	0.73	0.91	0.96
Log-linear binomial distribution						
Z	1.64	0.68	-0.27	-1.23	-2.18	-2.66
Power	0.05	0.25	0.61	0.89	0.98	0.996

Table 3: Numbers of male children in 3475 families of size 7

Male	Families	Binomial	Log-linear binomial
0	15	55.80	17.53
1	200	314.22	203.20
2	778	758.24	769.21
3	1240	1016.51	1233.02
4	897	817.65	903.93
5	295	394.62	303.07
6	45	105.81	43
7	5	12.16	2.02
Total	3475	3475	3475
Chi-square		193.01	1.374

Note: frequencies marked with bracket were combined for purpose of calculating χ^2

If we use binomial distribution, the estimated value of τ from the sample is

$$\hat{\tau} = \frac{3.12057}{7} = 0.4458. \tag{5.14}$$

This value of $\hat{\tau}$ was used to obtain the expected frequencies shown in Table 3. The value of $\chi^2 = 193.01$ with 7 degrees of freedom gives $p = 0 < 0.05$, the simple binomial model has to be rejected. If we use log-linear binomial distribution, and fit the data by maximum likelihood, we find that

$$\hat{\psi} = 0.423, \quad \hat{\omega} = 1.14543. \tag{5.15}$$

The estimated variance-covariance matrix is

$$\hat{V}(\hat{\theta}) \approx \begin{bmatrix} 70974.89 & 1071625 \\ 10716.25 & 8927.83 \end{bmatrix}^{-1} = \begin{bmatrix} 0.000017 & -0.000021 \\ -0.000021 & 0.000137 \end{bmatrix}. \tag{5.16}$$

The expected frequencies based on these estimates are shown in Table 3. The value of $\chi^2 = 1.288$ with 6 degrees of freedom gives $p = 0.97 > 0.05$. Thus, the log-linear binomial distribution provides a good fit to the observed data.

6. Conclusion

The parameters of the log-linear binomial distribution were estimated by the moment and maximum likelihood methods. Both methods needed solving nonlinear equations to obtain the estimators of the parameters. We used nonlinear equation solver (*nleqslv*) package in statistical R-software to find the estimates of the parameters. The variance-covariance matrix for the maximum likelihood estimates was obtained. Moreover, the sampling distribution of the sign test was studied when trials are dependent. A set of genetic data from Salmaniyaa hospital in Bahrain has been fitted using log-linear binomial distribution. The fit is found preferable over fitting the binomial distribution.

Acknowledgment

The author wishes to thank the two referees for their valuable comments that enhanced the contents and presentation of the paper.

References

- [1] N. L. Johnson, S. Kotz, and A. W. Kemp, *Univariate Discrete Distributions*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons, New York, NY, USA, 2nd edition, 1992.
- [2] D. Bowman and E. O. George, "A saturated model for analyzing exchangeable binary data: application to clinical and development toxicity studies," *Journal of American Statistical Association*, vol. 90, pp. 871–879, 1995.
- [3] E. O. George and D. Bowman, "A full likelihood procedure for analysing exchangeable binary data," *Biometrics*, vol. 51, no. 2, pp. 512–523, 1995.
- [4] C. Yu and D. Zelterman, "Sums of dependent Bernoulli random variables and disease clustering," *Statistics & Probability Letters*, vol. 57, no. 4, pp. 363–373, 2002.
- [5] C. Yu and D. Zelterman, "Sums of exchangeable Bernoulli random variables for family and litter frequency data," *Computational Statistics & Data Analysis*, vol. 52, no. 3, pp. 1636–1649, 2008.
- [6] N. Kolev and D. Paiva, "Random sums of exchangeable variables and actuarial applications," *Insurance*, vol. 42, no. 1, pp. 147–153, 2008.
- [7] G. Lovison, "An alternative representation of Altham's multiplicative-binomial distribution," *Statistics & Probability Letters*, vol. 36, no. 4, pp. 415–420, 1998.
- [8] P. M. E. Altham, "Two generalizations of the binomial distribution," *Journal of the Royal Statistical Society Series C*, vol. 27, no. 2, pp. 162–167, 1978.
- [9] D. R. Cox, "The analysis of multivariate binary data," *Applied Statistics*, vol. 21, pp. 113–120, 1972.
- [10] G. M. Tallis, "The use of a generalized multinomial distribution in the estimation of correlation in discrete data," *Journal of the Royal Statistical Society Series B*, vol. 24, pp. 530–534, 1962.
- [11] A. Luceño, "A family of partially correlated Poisson models for overdispersion," *Computational Statistics and Data Analysis*, vol. 20, no. 5, pp. 511–520, 1995.
- [12] B. Weaver, *Nonparametric Tests*, Lecture Notes, chapter 3, 2002.
- [13] S. Siegel and N. J. Castellan, *Nonparametric Statistics for the Behavioral Sciences*, McGraw-Hill, New York, NY, USA, 2nd edition, 1988.
- [14] D. F. Groebner, P. W. Shannon, P. C. Fry, and K. D. Smith, *Business Statistics: A Decision Making Approach*, Prentice Hall, Upper Saddle River, NJ, USA, 7th edition, 2007.