NOVA

University of Newcastle Research Online

nova.newcastle.edu.au

Best Donald John, Rayner John Charles, Thas O, 'Four tests of fit for the beta-binomial distribution'
Originally published in Journal of Applied Statistics, 37 1547-1554 (2010)

Available from: <http://dx.doi.org/10.1080/02664760903089664>

This is an electronic version of an article published in in Journal of Applied Statistics, 37 1547-1554 (2010). Journal of Applied Statistics is available online at:

<http://www.tandfonline.com/openurl?genre=article&issn=0266-4763&volume=37&issue=9&spage=1547>

Accessed from: <http://hdl.handle.net/1959.13/929521>

Four Tests of Fit for the Beta-Binomial Distribution

D. J. Best

School of Mathematical and Physical Sciences, University of Newcastle, NSW 2308, Australia
(John.Best@newcastle.edu.au)

J. C. W. Rayner

School of Mathematical and Physical Sciences, University of Newcastle, NSW 2308, Australia
(John.Rayner@newcastle.edu.au)

O. Thas

Department of Applied Mathematics, Biometrics and Process Control, Ghent University, B-9000 Gent, Belgium (olivier.thas@UGent.be)

Abstract

Tests based on the Anderson-Darling statistic, a third moment statistic and the classical Pearson-Fisher X^2 statistic, along with its third order component, are considered. A small critical value and power study is given. Some examples illustrate important applications.

Key Words: Anderson-Darling statistic; Multinomial distribution; Pearson's X^2 ; Third moment statistic; Third order component.

1. Introduction

In the binomial probability function

$$\phi(x; p) = \binom{n}{x} p^x (1-p)^{n-x}$$

for $x = 0, 1, \dots, n$, where $0 < p < 1$, assume the parameter p follows a beta distribution with probability function

$$p^{\alpha-1} (1-p)^{\beta-1} / B(\alpha, \beta)$$

for $0 < p < 1$, zero elsewhere, where $\alpha > 0$ and $\beta > 0$. The resulting beta-binomial distribution has probability function

$$P(x; \alpha, \beta) = \binom{n}{x} B(\alpha + x, \beta + n - x) / B(\alpha, \beta)$$

for $x = 0, 1, \dots, n$. This distribution can be used to model data which, at first glance, appear to follow a binomial distribution but which are, in fact, overdispersed compared with the binomial. The beta-binomial mean and variance are $\mu = n\pi$ and $\mu_2 = n\pi(1-\pi)\{1 + (n-1)\theta\}$ respectively. Here we have used the reparameterization $\pi = \alpha/(\alpha + \beta)$ and $\theta = 1/(\alpha + \beta + 1)$; $\pi = E[p]$ and θ is a measure of the variation in p . In this paper we will consider the parameter n to be fixed, but we note that in some applications, such as toxicological experiments, n varies. Recent applications with n fixed involve models for (a) taste-test data – see Ennis and Bi

(1998) and (b) market research data – see Danaher and Hardie (2005). Griffiths (1973) modelled medical data using the beta-binomial with n fixed while Hughes and Madden (1993) consider ecological data with n fixed.

In section 2 a third moment statistic, the Pearson-Fisher statistic and its third order component, and the Anderson-Darling statistic are defined. In section 3 a small critical value and power study is presented while section 4 analyses three data sets. Section 5 gives a brief conclusion.

2. Definitions

Suppose we have a sample of size n^* in which observation x occurs N_x times, with $n^* = N_0 + N_1 + \dots + N_n$. The two parameters of the beta-binomial are commonly estimated using (a) method of moments (MOM) or (b) maximum likelihood (ML). Subsequently we denote MOM estimators using a tilde ($\sim$) and ML estimators using a hat ($\hat{\cdot}$). Similarly statistics using MOM estimators will have a tilde and those using ML will have a hat. The MOM estimators are

$$\begin{aligned}\tilde{\alpha} &= (n - \bar{X} - S^2 / \bar{X})\bar{X} / (nS^2 / \bar{X} + \bar{X} - n) \text{ and} \\ \tilde{\beta} &= (n - \bar{X} - S^2 / \bar{X})(n - \bar{X}) / (nS^2 / \bar{X} + \bar{X} - n),\end{aligned}$$

where, as usual, $\bar{X} = \sum_{x=0}^n xN_x / n^*$ and $S^2 = \sum_{x=0}^n (x - \bar{X})^2 N_x / n^*$. Approach (b) needs an iterative technique such as two dimensional Newton-Raphson. This is outlined, for example, in Kemp and Kemp (1956). In the classical Pearson X^2 statistic if α and β are estimated by maximum likelihood, the test statistic is called the Pearson-Fisher X^2 statistic, and is given by

$$\hat{X}_{PF}^2 = \sum_{x=0}^n \left\{ N_x - n^* P(x; \hat{\alpha}, \hat{\beta}) \right\}^2 / \left\{ n^* P(x; \hat{\alpha}, \hat{\beta}) \right\}.$$

The r th component of $\hat{X}_{PF}^2$ can be defined as $\hat{V}_r = \sum_{x=0}^n N_x g_r(x; \hat{\alpha}, \hat{\beta}) / \sqrt{n}$ for $r = 1, 2, \dots, n$ in

which $\{g_r(x; \alpha, \beta)\}$ are polynomials orthonormal on $P(x; \alpha, \beta)$. Suppose $\mu = \sum_{x=0}^n xP(x; \alpha, \beta)$

and $\mu_r = \sum_{x=0}^n (x - \mu)^r P(x; \alpha, \beta)$ for $r = 2, 3, \dots$. The first four orthonormal polynomials are

$$g_0(x; \alpha, \beta) = 1 \text{ for all } x,$$

$$g_1(x; \alpha, \beta) = (x - \mu) / \sqrt{\mu_2}$$

$$g_2(x; \alpha, \beta) = \{ (x - \mu)^2 - \mu_3(x - \mu) / \mu_2 - \mu_2 \} / \sqrt{\mu_4 - \mu_3^2 / \mu_2 - \mu_2^2} \text{ and}$$

$$g_3(x; \alpha, \beta) = \frac{(x - \mu)^3 - a(x - \mu)^2 - b(x - \mu) - c}{\sqrt{\mu_6 - 2a\mu_5 + (a^2 - 2b)\mu_4 + 2(ab - c)\mu_3 + (b^2 + 2ac)\mu_2 + c^2}}$$

in which

$$\begin{aligned}
a &= (\mu_5 - \mu_3\mu_4 / \mu_2 - \mu_2\mu_3) / d, \\
b &= (\mu_4^2 / \mu_2 - \mu_2\mu_4 - \mu_3\mu_5 / \mu_2 + \mu_3^2) / d, \\
c &= (2\mu_3\mu_4 - \mu_3^3 / \mu_2 - \mu_2\mu_5) / d \text{ and} \\
d &= \mu_4 - \mu_3^2 / \mu_2 - \mu_2^2.
\end{aligned}$$

Here we will only consider $\hat{V}_3$ which uses $g_3(x; \hat{\alpha}, \hat{\beta})$.

The statistic $\hat{X}_{PF}^2$ equals $\sum_{r=1}^n \hat{V}_r^2$ as explained for example in Rayner and Best (1989) or Rayner et al. (2009). These references also show that a component $\hat{V}_r$ can provide a more powerful test of fit than $\hat{X}_{PF}^2$ itself. We will use ML estimates to calculate $\hat{V}_3$.

Let $\tilde{T} = \{m_3 - \tilde{\mu}_3\}$ where $m_3 = \sum_{x=0}^n N_x (x - \bar{X})^3 / n^*$. We use $\tilde{W} = T^2 / \text{var}(T)|_{\alpha=\tilde{\alpha}, \beta=\tilde{\beta}}$ as a third moment test statistic where an approximation to $\text{var}(T)$ is given in Appendix A.

Recently Lockhart et al. (2007) gave a general discussion of empirical distribution function (EDF) tests of fit for discrete distributions. The Anderson-Darling statistic often provides the most powerful EDF test and for the beta-binomial it may be defined by

$$\hat{A}^2 = \sum_{x=0}^n Z_x^2 P(x; \hat{\alpha}, \hat{\beta}) / \{h_x(1 - h_x)\} \text{ in which } Z_x = \sum_{y=0}^x N_y - n^* h_x \text{ and } h_x = \sum_{y=0}^x P(y; \hat{\alpha}, \hat{\beta}).$$

3. Critical Value and Power Comparisons

In Table 1 below 95% critical values of several test statistics were estimated based on 100,000 simulations for the n^* values given. The values of the test statistic were ordered and the 95,000th value taken as the critical value. The test statistics were $\tilde{W}$, the Pearson-Fisher statistic $\hat{X}_{PF}^2$ and its third order component $\hat{V}_3$, and $\hat{A}^2$. For Table 1 (a) the null beta-binomial $\alpha = 1.0, \beta = 2.0, n = 5$ distribution gives multinomial probabilities $\{0.38, 0.27, 0.18, 0.11, 0.05, 0.02\}$; for Table 1 (b) the null beta-binomial $\alpha = 2.0, \beta = 0.6, n = 13$ distribution gives multinomial probabilities $\{0.005, 0.010, 0.015, 0.021, 0.028, 0.035, 0.043, 0.052, 0.063, 0.076, 0.092, 0.117, 0.158, 0.284\}$; for Table 1(c) the null beta-binomial $\alpha = \beta = 2.0, n = 6$ gives multinomial probabilities $\{0.083, 0.143, 0.179, 0.190, 0.179, 0.143, 0.083\}$. Thus in (a) the probabilities decrease, in (b) they increase and in (c) they increase then decrease and are symmetric.

It appears that $\hat{A}^2$ reaches an asymptotic critical value quickly in Tables 1 (a), (b) and (c). Also the $\hat{X}_{PF}^2$ and $\hat{V}_3^2$ critical values approach their asymptotic χ^2 critical values at a reasonable rate.

Table 2 shows some powers, all based bootstrap simulations with 1000 simulations in both the inner and outer loops. Appendix B gives further details of the parametric bootstrap. In Table 2, $B(n, p)$ are binomial alternatives, $P(\lambda)$ are Poisson alternatives truncated at n and $BB(n, \pi, \theta^*)$ are beta-binomial alternatives where $\theta^* = 1/(\alpha + \beta)$. The negative binomial alternatives are also truncated at n . The multinomial alternatives were chosen so that as n, α and β varied the probabilities were generally increasing or decreasing. The Poisson and negative binomial alternatives were chosen as these are well known distributions where the variance, like that of the beta-binomial, is equal to or greater than the mean.

For the (admittedly few) alternatives considered here, no one test stands out. Each of the four statistics has the best power in at least one case. We note that $\hat{A}^2$ has relatively good power for all the alternatives and has the advantage of providing a consistent test. Within each subsection of Table 2, n and n^* are constant. As the beta-binomial distribution has a variety of shapes depending on n , n^* , α and β we think it is sensible to have a number of subsections in Table 2. We do not give powers for a binomial alternative as this is a beta-binomial distribution with $\theta = 0$ if we use the (π, θ) parameterization. Thus for a binomial alternative the power would be the test size. The Table 2(c) powers can be compared with those in Lin and Chou (2007, Table 2). The only statistic in common is $\hat{X}_{PF}^2$ and our powers for this test are essentially the same as theirs. The powers we give for $\hat{A}^2$ seem slightly better than the powers of the tests they consider.

Table 1. 95% critical values for various tests of beta-binomial fit for various samples of size n^*
(a) $\alpha = 1.0$, $\beta = 2.0$, $n = 5$

n^*	$\tilde{W}$	$\hat{X}_{PF}^2$	$\hat{V}_3^2$	$\hat{A}^2$
25	3.74	7.56	3.72	0.54
50	3.78	7.48	3.80	0.57
100	3.80	7.70	3.80	0.57
500	3.83	7.80	3.80	0.57
1,000	3.84	7.71	3.80	0.57
10,000	3.84	7.79	3.81	0.57
∞	3.84	7.81	3.84	-

(b) $\alpha = 2.0$, $\beta = 0.6$, $n = 13$

n^*	$\tilde{W}$	$\hat{X}_{PF}^2$	$\hat{V}_3^2$	$\hat{A}^2$
25	3.60	18.82	3.17	0.44
50	3.77	19.33	3.38	0.44
100	3.77	19.40	3.52	0.44
500	3.81	19.68	3.60	0.44
1,000	3.81	19.68	3.62	0.44
10,000	3.84	19.68	3.69	0.44
∞	3.84	19.68	3.84	-

(c) $\alpha = 2.0$, $\beta = 2.0$, $n = 6$

n^*	$\tilde{W}$	$\hat{X}_{PF}^2$	$\hat{V}_3^2$	$\hat{A}^2$
25	3.93	9.23	3.90	0.64
50	3.86	9.47	3.84	0.64
100	3.85	9.47	3.84	0.63
500	3.85	9.49	3.83	0.63
1,000	3.84	9.49	3.82	0.63
10,000	3.84	9.49	3.83	0.64
∞	3.84	9.49	3.84	-

We also considered powers for V_3 , X^2 and A^2 when the statistic was calculated using MOM estimators rather than ML estimators. The powers were very similar to those resulting when using ML estimators.

As noted in the Introduction we do not consider the case where n varies. Garren (2004) summarizes approaches for this case while Lin and Chou (2007) suggest the use of EDF tests.

Table 2. Powers of 5% level tests for beta-binomial fit

(a) (i) $n = 5$ and $n^* = 50$

Alternative	$\tilde{W}$	$\hat{X}_{PF}^2$	$\hat{V}_3^2$	$\hat{A}^2$
Multinomial (0.4, 2*0.2, 2*0.05, 0.1)	0.31	0.37	0.34	0.32
Multinomial (2*0.3, 2*0.15, 2*0.05)	0.13	0.18	0.13	0.14
Truncated Poisson (2)	0.05	0.22	0.05	0.12
Truncated Poisson (1.5)	0.10	0.09	0.09	0.07

(a) (i) $n = 5$ and $n^* = 100$

Alternative	$\tilde{W}$	$\hat{X}_{PF}^2$	$\hat{V}_3^2$	$\hat{A}^2$
Multinomial (0.4, 2*0.2, 2*0.05, 0.1)	0.58	0.72	0.63	0.66
Multinomial (2*0.3, 2*0.15, 2*0.05)	0.23	0.35	0.24	0.26
Truncated Poisson (2)	0.08	0.32	0.08	0.20
Truncated Poisson (1.5)	0.14	0.12	0.16	0.09

(b) (i) $n = 13$ and $n^* = 50$

Alternative	$\tilde{W}$	$\hat{X}_{PF}^2$	$\hat{V}_3^2$	$\hat{A}^2$
Multinomial (3*0.1, 10*0.05, 0.2)	0.21	0.20	0.21	0.30
Multinomial (6*0.07, 6*0.06, 0.08, 0.2)	0.29	0.15	0.30	0.26
Multinomial (3*0.01, 0.02, 7*0.05, 0.1, 0.2, 0.3)	0.12	0.18	0.15	0.21
Truncated Negative Binomial(4, 0.5)	0.38	0.25	0.34	0.30

(b) (ii) $n = 13$ and $n^* = 100$

Alternative	$\tilde{W}$	$\hat{X}_{PF}^2$	$\hat{V}_3^2$	$\hat{A}^2$
Multinomial (3*0.1, 10*0.05, 0.2)	0.39	0.40	0.41	0.54
Multinomial (6*0.07, 6*0.06, 0.08, 0.2)	0.57	0.24	0.53	0.49
Multinomial (3*0.01, 0.02, 7*0.05, 0.1, 0.2, 0.3)	0.18	0.32	0.18	0.37
Truncated Negative Binomial(4, 0.5)	0.58	0.49	0.61	0.50

(c) (i) $n = 20$ and $n^* = 25$

Alternative	$\tilde{W}$	$\hat{X}_{PF}^2$	$\hat{V}_3^2$	$\hat{A}^2$
0.5*B(20, 0.1) + 0.5*B(20, 0.5)	0.29	0.20	0.31	0.54
0.95*BB(20, 0.1, 0.1) + 0.05*B(20, 0.8)	0.39	0.29	0.35	0.22
0.85*BB(20, 0.1, 0.15) + 0.15*B(20, 0.9)	0.37	0.27	0.35	0.20
0.95*BB(20, 0.15, 0.1) + 0.05*B(20, 0.8)	0.49	0.31	0.36	0.34
0.85*BB(20, 0.15, 0.15) + 0.15*B(20, 0.9)	0.34	0.18	0.28	0.39

(c) (ii) $n = 20$ and $n^* = 50$

Alternative	$\tilde{W}$	$\hat{X}_{PF}^2$	$\hat{V}_3^2$	$\hat{A}^2$
0.5*B(20, 0.1) + 0.5*B(20, 0.5)	0.52	0.52	0.56	0.86
0.95*BB(20, 0.1, 0.1) + 0.05*B(20, 0.8)	0.69	0.59	0.77	0.54
0.85*BB(20, 0.1, 0.15) + 0.15*B(20, 0.9)	0.71	0.53	0.70	0.60
0.95*BB(20, 0.15, 0.1) + 0.05*B(20, 0.8)	0.69	0.42	0.65	0.84
0.85*BB(20, 0.15, 0.15) + 0.15*B(20, 0.9)	0.82	0.54	0.81	0.85

Table 3. Comparison of counts and expected counts using beta-binomial models for the number of alcohol drinking days

<i>Days</i>	N_x	BB	GBB	f
0	47	54.6	47.9	46.7
1	54	42.0	50.1	47.9
2	43	38.9	46.5	46.2
3	40	38.5	42.1	42.0
4	40	40.1	38.6	37.6
5	41	44.0	37.3	36.4
6	39	53.1	41.8	45.3
7	95	87.8	94.8	97.2

4. Examples

We now give three real data examples. The first of these has $\hat{X}_{PF}^2$ non-significant at the 5% significance level whereas $\tilde{W}$ and $\hat{A}^2$ are highly significant. The second example illustrates that using a conventional binomial model leads to a different conclusion than that using the beta-binomial model, while in the third example we show the beta-binomial model fits well and so can be used to derive sample sizes for future similar experiments.

Alcohol Consumption Example

Recently Rodríguez-Avi et al. (2007) discussed fitting a beta-binomial model to the number of days per week (0, 1, ..., 7) alcohol was consumed by a random sample of 399 Dutch respondents. Thus $n = 7$ and $n^* = 399$. The counts for this sociological study were 47, 54, 43, 40, 40, 41, 39 and 95. We find $\hat{X}_{PF}^2 = 9.51$, $\tilde{W} = 7.10$, $\hat{A}^2 = 0.629$ and $\hat{V}_3^2 = 7.31$ with parametric bootstrap p-values of 0.09, 0.007, 0.009 and 0.007 respectively. Although $\hat{X}_{PF}^2$ is not significant at the 5% level, $\tilde{W}$ and $\hat{A}^2$ are each highly significant and so the beta-binomial model is not satisfactory. Rodríguez-Avi et al. (2007) proceed to fit a three parameter generalised beta-binomial (GBB) model. See Table 3. Using $\hat{\pi} = \hat{\alpha} / (\hat{\alpha} + \hat{\beta})$ we can say that an estimate of the probability of drinking alcohol on a given single day is $\hat{\pi} = 0.5541$. Rodríguez-Avi et al. (2007) used their GBB model to better describe the alcohol consumption data. Another improved model is

$$f(x; \hat{\alpha}, \hat{\beta}) = n^* P(x; \hat{\alpha}, \hat{\beta}) \{1 + \hat{V}_3 g_3(x; \hat{\alpha}, \hat{\beta}) / \sqrt{n^*}\}$$

as can be seen in Table 3. Rayner et al. (2009) discuss this type of improved model.

Sensory Evaluation Example

Ennis and Bi (1998) gave taste test data where $n^* = 100$ consumers participated in a three replication ($n = 3$) preference test for two beverages. Table 4 shows the data (number of times there was a preference for beverage A) and the fit for a binomial model and a beta-binomial model. Before using either the binomial or the beta-binomial model it is clearly important to check the goodness of fit of both models. Table 4 indicates the binomial is not a suitable model while the beta-binomial is. Also the tests based on $\hat{A}^2$, $\tilde{W}$ and $\hat{V}_3^2$ all give a parametric bootstrap p-value of 0.19.

In the binomial model we can estimate the probability of preference A as $\hat{p} = \sum_{x=0}^n xN_x / (nn^*)$, giving $\hat{p} = 0.57$ for the binomial parameter p . An approximate 95% confidence interval for p is $\hat{p} \pm 1.96 \sqrt{\hat{p}(1-\hat{p})/300}$ or (0.51, 0.63) which does not include 0.5 and so we might conclude there is a preference for beverage A.

On the other hand we find $\tilde{\pi} = 0.57$ and $\tilde{\theta} = 0.53$ for the beta-binomial model with 95% confidence interval for π of $\tilde{\pi} \pm 1.96 \sqrt{\tilde{\pi}(1-\tilde{\pi})\{1+(n-1)\tilde{\theta}\}/(nn^*)}$ or (0.49, 0.65), which does include 0.5 and so we might now conclude differently, that there is not a preference for beverage A.

Table 4. Comparison of counts and expected counts using binomial and beta-binomial models for some taste-test preference data

x	N_x	Binomial	BB (MOM)
0	21	8.0	19.5
1	18	31.6	22.4
2	30	41.9	25.6
3	31	18.5	32.5
$\tilde{\chi}^2$		39.1	1.76
df		2	1
p-value		0.00	0.18

Tomato Virus Example

Hughes and Madden (1993) show that the classic Cochran (1936) data for spotted wilt disease on tomato plants is exceptionally well fitted by the beta-binomial: see Table 5. We have $n^* = 160$ and $n = 9$, and $\hat{X}_{PF}^2 = 2.248$, $\tilde{W} = 0.296$, $\hat{A}^2 = 0.027$ and $\hat{V}_3^2 = 0.304$ with parametric bootstrap p-values of 0.81, 0.60, 0.99 and 0.62 respectively. On the basis of these nonsignificant p-values we conclude the beta-binomial model fits well and this agrees with the conclusion of Hughes and Madden (1993). Our calculation of $\hat{X}_{PF}^2$ does not involve any pooling of classes. Knowledge that the beta-binomial applies would help an ecologist use an appropriate sample size in future similar investigations.

Table 5. Observed and expected frequencies on the beta-binomial model for tomato plant data

Diseased plants per quadrant	N_x	BB (MOM)
0	36	36.3
1	48	47.8
2	38	37.7
3	23	22.2
4	10	10.4
5	3	4.0
6	1	1.2
7	1	0.3
8	0	0.04
9	0	0.004

5. Conclusion

In this note we have given small sample power comparisons for four tests of fit for the beta-binomial model. We have applied the four tests of fit to three real data sets. The Table 2 powers in section 3 showed that no one statistic was always most powerful. However, $\hat{A}^2$ was either most powerful or close to most powerful and has the advantage of being a consistent test statistic. The tests based on $\tilde{W}$ and $\hat{V}_3^2$ may have poor power even for large n^* if the alternative has a similar third moment to a beta-binomial model. The third moment test based on $\tilde{W}$ was almost as powerful as that based on $\hat{A}^2$, the test statistic has a convenient asymptotic distribution which can be used to give approximate p-values for $n^* > 50$ say, and the test uses easily calculated MOM estimators. A significant $\tilde{W}$ also suggests whether the data differ from the null in the third moment. The test based on $\hat{V}_3^2$ performs similarly to that based on $\tilde{W}$ but uses ML estimators which need iterative methods to calculate. The Pearson-Fisher $\hat{X}_{PF}^2$ test is generally the least powerful of those considered here.

Thus we suggest use of $\hat{A}^2$ to test the fit of the beta-binomial model but also note that the test based on $\tilde{W}$ has some advantages. We have given three examples illustrating important applications of the beta-binomial model with n fixed and we have illustrated use of $\hat{V}_3$ to give an improved model.

Acknowledgement

The authors wish to thank two anonymous referees for their constructive comments. This research was supported by the IAP research network, Grant number P6/03 of the Belgium Government (Belgium Science Policy).

References

- Cochran, W.G. (1936). The statistical analysis of field counts of diseased plants. *Supplement of the Journal of the Royal Statistical Society*, 3, 49-67.
- Danaher, P.J. and Hardie, B.G.S. (2005). Bacon with your eggs? Applications of a new bivariate beta-binomial distribution. *American Statistician*, 59, 282-286.
- Devroye, L. (1986). *Non-Uniform Random Variate Generation*. New York: Springer-Verlag.
- Ennis, D.M. and Bi, J. (1998). The beta-binomial model: accounting for inter-trial variation in replicated difference and preference tests. *Journal of Sensory Studies*, 13, 389-412.
- Garren, S.T. (2004). Goodness-of-fit testing for the beta-binomial Model. Chapter 9 in *Handbook of Beta Distribution and its Applications* eds. Gupta, A.K. and Nadarajah, S, New York: Dekker, 237-253.
- Griffiths, D.A. (1973). Maximum likelihood estimators for the beta-binomial distribution and an application to household distribution of the total number of cases of a disease. *Biometrics*, 29, 637-648.
- Gürtler, N. and Henze, N. (2000). Recent and classical goodness-of-fit tests for the Poisson distribution. *Journal of Statistical Planning and Inference*, 90, 207-225.
- Hughes, G. and Madden, L.V. (1993). Using the beta-binomial distribution to describe aggregated patterns of disease incidence. *Phytopathology*, 83, 759-763.
- IMSL (1995). *Users' Manual*. Houston: IMSL.

- Kemp, C.D. and Kemp, A.W. (1956). The analysis of point quadrat data. *Australian Journal of Botany*, 4, 167-174.
- Lin, C. T. and Chou, C.C. (2007). Empirical-distribution-function tests for the beta-binomial model. *Journal of Applied Statistics*, 34(6), 715-724.
- Lockhart, R.A., Spinelli, J.J. and Stephens, M.A. (2007). Cramér-von Mises statistics for discrete distributions with unknown parameters. *Canadian Journal of Statistics*, 35, 125-133.
- Rayner, J.C.W. and Best, D.J. (1989). *Smooth Tests of Goodness of Fit*. New York: Oxford University Press.
- Rayner, J.C.W., Thas, O. and Best, D.J. (2009). *Smooth Tests of Goodness of Fit: Using R* (2nd ed.). Singapore: Wiley.
- Rodríguez-Avi, J., Conde-Sánchez, A., Sáez-Castillo, A.J. and Olmo-Jiménez, M.J. (2007). A generalization of the beta-binomial distribution. *Applied Statistics*, 56, 51-61.

Appendix A: Var(T)

First write $t = 1/(\tilde{\alpha} + \tilde{\beta} + 1)$ and $p = \tilde{\alpha}/(\tilde{\alpha} + \tilde{\beta})$. Then an approximation to the variance of T is given by

$$\begin{aligned} \text{var}(T) = & 2(n-1)(n-2)(1-p)(t-1)(pt-p+1)(pt-p-t)(1-t+nt) (230np^2t^4 - 6np^2t + \\ & 37np^2t^2 - 230npt^4 + 28n^2p^2t^3 - 242np^2t^5 + 124npt^3 - 3n^2p^2t^2 + 80n^2p^2t^5 + 3n^2pt^2 - \\ & 124np^2t^3 + 242npt^5 - 80n^2pt^5 - 70n^2p^2t^4 + 6npt + 70n^2pt^4 - 28n^2pt^3 - 37npt^2 + 6t + \\ & 141np^2t^6 - 47n^2p^2t^6 - 9pt - 183p^2t^4 + 9p^2t - 56p^2t^2 + 56pt^2 + 16nt^2 + 62nt^4 - 14n^2t^4 - \\ & 22nt^3 + 10n^2t^3 + 183pt^4 - 142pt^3 + 142p^2t^3 + 3p - 8t^2 - 3p^2 - 44t^4 + 40t^3 + 161p^2t^5 - \\ & 161pt^5 + 24n^2t^5 + 46t^5 - 68nt^5 - 94p^2t^6 + 94pt^6 - 12n^2t^6 + 36nt^6 - 141npt^6 + 47n^2pt^6 - \\ & 24t^6 + 24p^2t^7 - 24pt^7 - 36np^2t^7 + 12n^2p^2t^7 + 36npt^7 - 12n^2pt^7)np/\{(1+4t)(1+3t)(1+ \\ & 2t)(1+t)^5n^*\} \end{aligned}$$

It follows that $\tilde{W}$ has an approximate χ_1^2 distribution. For $n^* \geq 50$ this approximation seems to give reasonable p-values.

Appendix B: Parametric Bootstrap p -values

Gürtler and Henze (2000, p. 223) suggest p-values can be obtained using an analogue of the parametric bootstrap. If W_n denotes a test statistic calculate $w_n := W_n(x_1, x_2, \dots, x_n)$ where $x_1, x_2, \dots, x_n$ denotes, as usual, the data. Find estimates $\tilde{\alpha}$ and $\tilde{\beta}$ (not necessarily the maximum likelihood estimates) from the data and conditional on those estimates, generate $B = 10,000$ say pseudo-random samples of size n , each having the beta-binomial $(\tilde{\alpha}, \tilde{\beta})$ distribution. For $j = 1, \dots, B$ compute the value $W_{n,j}^*$ on each random sample. Then the parametric bootstrap p-value is the proportion of $W_{n,j}^*$ that is at least the observed w_n , namely $\sum_{j=1}^B I(W_{n,j}^* \geq w_n) / B$.

The above requires random beta-binomial (α, β) values. These can easily be generated using an algorithm for random multinomial values where the multinomial probabilities are specified to be $P(x; \alpha, \beta)$. Devroye (1986, p.559) gives such an algorithm. Alternatively the IMSL (1995) routine RNMTN can be used. To obtain p-values for two-tailed tests proceed as above and find the p-value, P say. If $P \leq 0.5$ the two-tailed p-value is $2P$; if $P > 0.5$ the two-tailed p-value is $2(1 - P)$.