Appl. Statist. (1999)
48, *Part* 4, *pp.* 553–561

On the use of corrections for overdispersion

J. K. Lindsey

Limburgs Universitair Centrum, Diepenbeek, Belgium

[Received May 1998. Final revision February 1999]

Summary. In studying fluctuations in the size of a blackgrouse (*Tetrao tetrix*) population, an autoregressive model using climatic conditions appears to follow the changes quite well. However, the deviance of the model is considerably larger than its number of degrees of freedom. A widely used statistical rule of thumb holds that overdispersion is present in such situations, but model selection based on a direct likelihood approach can produce opposing results. Two further examples, of binomial and of Poisson data, have models with deviances that are almost twice the degrees of freedom and yet various overdispersion models do not fit better than the standard model for independent data. This can arise because the rule of thumb only considers a point estimate of dispersion, without regard for any measure of its precision. A reasonable criterion for detecting overdispersion is that the deviance be at least twice the number of degrees of freedom, the familiar Akaike information criterion, but the actual presence of overdispersion should then be checked by some appropriate modelling procedure.

Keywords: Akaike information criterion; Beta–binomial distribution; Direct likelihood inference; Negative binomial distribution; Overdispersion

1. Introduction

Blackgrouse (*Tetrao tetrix*) in the high eastern region of Belgium form a very small population that is close to extinction. Biologists have followed them closely for many years. The numbers of cocks on their mating grounds are counted each spring, with data available for 30 years.

In modelling this population, we are interested to see whether we can obtain an adequate model for the available data by using only climatic information. If a population is in an ecologically viable equilibrium, it should be able to adjust to a changing habitat; short-term variations in weather should be the only major influence on the size of the population. This can have an immediate effect on the survival of the young, but also a more delayed effect through the numbers of grouse reaching maturation. To account for the latter, population levels in the previous 2 years are included in the model because the cocks take 2 years to reach maturity.

We know that habitat factors, especially variations in the number of foxes as influenced by changes in methods of rabies control, the activity of poachers and evolution of the plant habitat in the region over the observation period, have an effect. However, the question is not whether such variables are missing; we know that they are. The question concerns the adequacy and appropriateness of a climatic model for describing the observed variations. For a further discussion of these assumptions, and details on the models and the conclusions, see Loneux *et al.* (1997).

Address for correspondence: J. K. Lindsey, Department of Biostatistics, Limburgs Universitair Centrum, 3590 Diepenbeek, Belgium.
E-mail: jlindsey@luc.ac.be

 0035–9254/99/48553

With this goal in mind, the climatic conditions over the 30 years, at key periods of the reproduction cycle, have been assembled with considerable effort by Mme Loneux, each variable being chosen for its biological significance. Out of this large collection, the variables providing the best prediction are shown in Table 1 and the population numbers are plotted in Fig. 1. The data in Table 1 can also be obtained from

`http://www.blackwellpublishers.co.uk/rss/`

The best model, using the variables in Table 1 as well as an autoregression on the observed population numbers lagged 1 and 2 years, is also plotted in Fig. 1. It has a deviance of 29.0 with 18 degrees of freedom (the first two years are not used as a response because of the lag). The question arises whether overdispersion is present.

To illustrate better the statistical problem, I shall also consider the simpler model with the least significant variable (x_3, total precipitation the previous September) omitted. This model has a deviance of 39.2 with 19 degrees of freedom, appearing even more clearly to indicate overdispersion. By a frequentist goodness-of-fit test, there is strong evidence of a lack of fit, but, by a model selection criterion such as the Akaike information criterion AIC, there is little

Table 1. Yearly numbers of blackgrouse cocks with the climatic variables†

Year	*Count*	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1967	80	−1.4	0.0	51.3	64.9	133.4	101.2	9.5
1968	115	−2.1	−1.4	93.3	143.3	114.1	46.9	9.3
1969	140	−2.6	−2.1	217.5	73.0	94.9	55.8	10.2
1970	165	−3.7	−2.6	42.3	94.2	106.5	72.7	8.5
1971	198	−2.2	−3.7	96.4	62.8	52.9	30.0	10.0
1972	160	−1.6	−2.2	50.2	97.2	149.8	133.8	7.3
1973	118	−1.6	−1.6	115.0	93.3	92.3	25.4	8.1
1974	100	−1.1	−1.6	75.2	102.8	89.7	56.4	11.6
1975	52	−0.2	−1.1	160.1	84.7	85.0	67.9	8.8
1976	42	−2.1	−0.2	108.3	35.9	61.1	57.3	8.4
1977	81	−1.5	−2.1	72.3	80.4	60.1	34.4	14.3
1978	58	−1.6	−1.5	58.2	103.0	91.5	91.5	10.0
1979	55	−3.6	−1.6	131.8	59.5	52.2	51.3	6.9
1980	54	0.7	−3.6	28.4	165.2	141.0	81.8	7.6
1981	46	−2.7	0.7	53.6	62.8	78.3	52.4	6.6
1982	34	−2.4	−2.7	83.6	104.7	122.0	94.7	7.3
1983	35	−1.4	−2.4	67.2	165.4	142.3	140.9	12.7
1984	66	−2.3	−1.4	138.0	184.4	126.1	22.9	8.5
1985	40	−3.4	−2.3	310.1	192.5	119.0	40.9	6.5
1986	50	−4.0	−3.4	77.7	118.7	148.1	139.5	7.9
1987	74	−3.5	−4.0	95.4	144.2	141.0	136.5	13.3
1988	57	−0.8	−3.5	121.8	136.6	146.5	114.0	8.6
1989	45	0.1	−0.8	148.2	126.3	115.2	39.7	8.9
1990	53	0.2	0.1	71.9	59.2	59.3	59.3	9.8
1991	42	−1.5	0.2	119.0	56.7	70.5	70.5	8.8
1992	43	−1.1	−1.5	58.4	68.5	103.8	103.5	9.8
1993	55	−1.4	−1.1	69.2	44.2	97.8	88.4	9.9
1994	47	−1.4	−1.4	176.5	73.6	80.5	59.1	8.7
1995	36	0.1	−1.4	167.0	139.5	114.5	99.1	11.5
1996	26	−3.1	0.1	104.3	143.7	157.6	127.9	10.3

†x_1, average minimum temperature the previous winter; x_2, average minimum temperature two winters before; x_3, total precipitation the previous September; x_4, precipitation during the four weeks of the previous year starting May 19th; x_5, precipitation during the four weeks of the previous year starting May 25th; x_6, precipitation during the three weeks of the previous year starting June 1st; x_7, minimum temperature during the three weeks of the previous year starting June 16th.

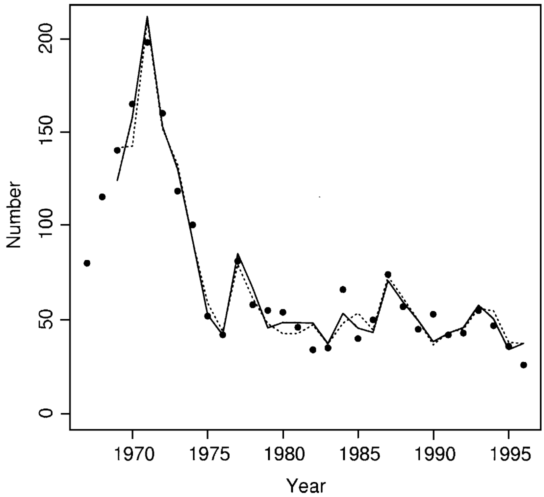

Fig. 1. Observed numbers of blackgrouse (•) with the best fitted model (———) and the model with one variable removed (··········)

evidence, if x_3 were not available for inclusion. Omitting this variable is not a relevant biological exercise because, if it had been missing, one of the many other available weather variables would have taken its place almost as well. The question is rather statistical for at least two reasons. A scientific problem generally requires models incorporating not all available significant variables, but only those relevant to that problem. As well, the full model contains nine variables with only 28 usable observations so that overfitting may be a serious consideration and a simplification reasonable. What is the effect of such omissions on overdispersion?

Corrections for overdispersion are widely used when modelling count data. Some statisticians, following McCullagh and Nelder (1989), pages 90 and 125, but ignoring their qualification ('unless the data or prior information indicate otherwise'), suggest that it is wise to assume that overdispersion is always present. Cox (1983) showed that the heterogeneity factor usually provides an adequate correction, without further modelling, unless the data are highly unbalanced. However, these arguments are based on point estimates of a dispersion parameter, generally not involving an actual probabilistic model for the overdispersion. In contrast, Finney (1971), p. 72, and Finney (1978), pages 96 and 373, took a more cautious approach, warning against corrections as a global remedy. The present paper is in the spirit of Finney.

Lindsey (1993) pointed out that overdispersion should only be suspected when *counts*, i.e. repeated events on the same units, are being observed. Overdispersion will not occur when *frequencies*, i.e. independent events on different units, are being recorded. Allowing for overdispersion in the latter case instead of using the available covariates may just be an excuse for careless modelling, the point made by Finney.

For count data, the standard overdispersion model is the negative binomial distribution, obtained by assuming that the mean varies randomly in the population, following a gamma distribution. The corresponding model for binomial overdispersion is the beta–binomial distribution, derived by assuming that the probability has a beta distribution. Both have simple analytical forms because they arise from conjugate distributions in the exponential family.

Other mixing distributions can be used, such as the normal distribution (Hinde, 1982). Overdispersed distributions without a mixture interpretation are also available. Altham

(1978) has suggested a ‘multiplicative’ binomial model; an analogous model can be derived as a generalization of the Poisson distribution. Efron (1986) introduced the family of double-exponential distributions, with appropriate members for binomial and Poisson data. This set of four possible overdispersion models each for binomial and for Poisson counts provides a considerable degree of robustness to various forms of overdispersion. When some form of overdispersion is present, these various models can yield very different results.

One advantage of using true distributions instead of quasi-likelihood or pseudolikelihood methods is that the relative goodness of fit of various models can be directly compared through the likelihood function. In this way, the need to correct for overdispersion can be determined. This paper shows that the commonly used criterion, that the deviance be greater than its number of degrees of freedom, can yield misleading inferences from a direct likelihood point of view. Indeed, as we shall also see, this criterion is little used in that one area, the social sciences, where contingency tables most frequently occur.

2. Inferences

Overdispersion is an area where the differences in conclusions between the probability-based frequentist and Bayesian approaches and the direct likelihood model selection procedures (Lindsey, 1996, 1999a) can most clearly be seen.

For example, in the social sciences, contingency tables with large frequencies often occur so that inferences based on a χ^2-distribution are known not to provide reasonable answers. An alternative procedure proposed by Raftery (1986a, b) is the Bayesian information criterion BIC (Schwarz, 1978). It has a direct likelihood interpretation as does the Akaike (1973) information criterion AIC. The former penalizes the deviance (twice the negative log-likelihood) by adding $p \log(n)$, where p is the number of estimated parameters in the model and n is the number of events in the contingency table. In a saturated model, this is equivalent to assuming that overdispersion is only present if the deviance is greater than $(N-p) \log(n)$, where N is the number of cells in the table so that $N-p$ is the number of degrees of freedom.

Similarly, if AIC is used for model selection in contingency tables (Bai *et al.* (1992); see also Burnham *et al.* (1995) and Buckland *et al.* (1997)), overdispersion will only be considered present if the deviance is more than twice the number of degrees of freedom.

Even the traditional rule of thumb is a form of direct likelihood procedure because it does not take into account the degrees of freedom through a sampling distribution as is usually done in frequentist methods. It is equivalent to an AIC with the penalty equal to the number of estimated parameters instead of twice that number. If it is interpreted inappropriately as a frequentist ‘test’ on the deviance, the criterion that overdispersion is present when the deviance is larger than the degrees of freedom would reject the null hypothesis of no overdispersion, for example, at a 42% level with 5 degrees of freedom and at a 48% level with 100 degrees of freedom. If the average deviance is used as an estimate of overdispersion, a frequentist criterion at the 5% level would indicate overdispersion when the deviance was greater than 11 with 5 degrees of freedom and greater than 124 with 100 degrees of freedom. Thus the ratio changes from about 2 to 1.25 as the number of degrees of freedom increases, but it does not reach 1.

Models in the examples to follow will be compared by using a direct likelihood approach with the negative log-likelihood penalized by adding the number of parameters estimated (a form of AIC). Smaller values indicate relatively better models. However, the deviances quoted are the standard *two times* the difference in negative log-likelihood with respect to the saturated model.

AIC is used because the author believes that an appropriate sample size should be chosen, *given* some inference criterion, in such a way that a scientifically interesting effect can be detected; this is not possible with BIC.

3. Models

In all examples below, all of the distributions mentioned above (conjugate, normal mixture, Altham and Efron) were fitted and all led to the same conclusions about lack of overdispersion. Only details of the results from the binomial and beta–binomial, Poisson and negative binomial models are presented.

With y_{ij} events, on the individual indexed by i within the group having covariate value j, the negative binomial distribution is

$$\Pr(y_{ij}) = \frac{\Gamma(y_{ij} + \kappa_j)}{y_{ij}!\,\Gamma(\kappa_j)} \left(\frac{1}{1+\upsilon_j}\right)^{\kappa_j} \left(\frac{\upsilon_j}{1+\upsilon_j}\right)^{y_{ij}}$$

with mean $\mu_j = \kappa_j \upsilon_j$ and correlation $\rho_j = 1/\kappa_j$, where $\Gamma(\cdot)$ is the gamma function. For a constant correlation, we can use

$$\mathrm{var}(Y_{ij}) = \mu_j(1 + \rho\mu_j).$$

However, to obtain the variance proportional to the mean, we can set υ_j constant to obtain

$$\mathrm{var}(Y_{ij}) = \mu_j(1 + \upsilon)$$

so that the heterogeneity factor is $\phi = 1 + \upsilon$.

With y_{ij} successes out of n_{ij} trials, and using the same indexing as above, the beta–binomial distribution is given by

$$\Pr(y_{ij}) = \binom{n_{ij}}{y_{ij}} \frac{B(\kappa_j + y_{ij},\, \upsilon_j + n_{ij} - y_{ij})}{B(\kappa_j,\, \upsilon_j)}$$

with $\kappa_j = (1-\rho)\pi_j/\rho$ and $\upsilon_j = (1-\rho)(1-\pi_j)/\rho$ where π_j is the probability of success, ρ is the constant correlation among events on a unit and $B(\cdot)$ is the beta function. The variance is

$$\mathrm{var}(Y_{ij}) = n_{ij}\pi_j(1-\pi_j)\{1 + (n_{ij} - 1)\rho\}$$

and the heterogeneity factor is approximately $\phi = 1 + (n-1)\rho$ where n is the average number of trials.

4. Blackgrouse and the climate

The biological problem and the final model were described in Section 1 and the latter plotted in Fig. 1. Let us now consider the possibility of overdispersion more closely. The parameter estimates from the Poisson distribution, with and without the least significant variable x_3, are shown in Table 2. The full model has a deviance of 29.0 with 18 degrees of freedom whereas the simplified model has deviance 39.2 with 19 degrees of freedom. The increased lack of fit arises primarily from years 1970, 1980, 1984 and 1985, whereas the fit for 1969 improves. The parameter estimates and standard errors change little when x_3 is removed.

Table 2. Parameter estimates and standard errors for the two models fitted to the blackgrouse data by using a Poisson distribution

Parameter	*Full model*		*Simplified model*	
	Estimate	*Standard error*	*Estimate*	*Standard error*
Intercept	2.9999	0.2035	2.7746	0.1912
Lag1	0.0113	0.0011	0.0106	0.0011
Lag2	−0.0039	0.0011	−0.0028	0.0010
x_1	−0.0761	0.0213	−0.0707	0.0215
x_2	−0.0889	0.0261	−0.0994	0.0263
x_3	−0.0015	0.0005	—	—
x_4	−0.0048	0.0013	−0.0053	0.0013
x_5	0.0080	0.0022	0.0075	0.0023
x_6	−0.0085	0.0014	−0.0070	0.0014
x_7	0.0822	0.0144	0.0838	0.0143

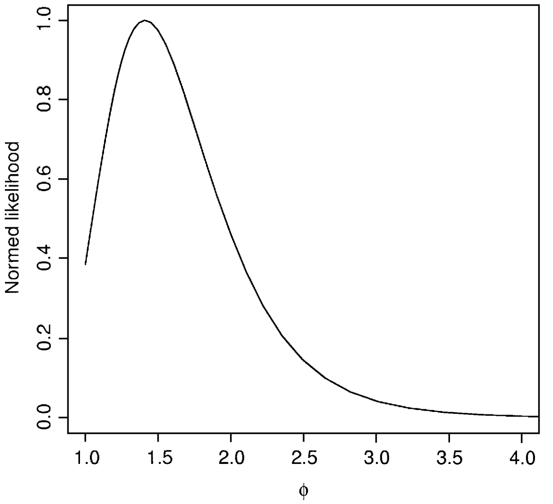

Fig. 2. Normed profile likelihood for the heterogeneity factor in the negative binomial model for the blackgrouse data

Allowing for overdispersion, using any of the distributions described above, hardly changes the likelihood for either model. For the simplified model, the AICs are 111.8 for the Poisson, 112.4 and 111.9 for the negative binomial respectively for constant correlation and for variance proportional to the mean, 112.0 for the double-Poisson, 112.8 for the multiplicative Poisson and 112.4 for the normal–Poisson distributions. In all cases, the estimates and standard errors are close to those given in Table 2 for the Poisson distribution.

For the negative binomial distribution, in the simplified model, with the variance proportional to the mean, the heterogeneity factor is estimated to be $\hat{\phi} = 1 + \exp(-0.8932) = 1.41$ compared with $39.2/19 = 2.06$ from the deviance. The normed profile likelihood for this parameter of the negative binomial model is plotted in Fig. 2. The profile is very wide, giving no reason to exclude $\phi = 1$; this value is about as likely as the estimate of 2.06 derived from the average deviance. This confirms the conclusions from the AICs for the various models. Even with one of the essential available variables eliminated, there is no indication of overdispersion. We can conclude that climatic factors provide a reasonable description of the data.

5. Further examples

5.1. Binomial counts

Crowder (1978) gave two data sets on the germination of seeds in vegetable extracts. In his second data set, a 2×2 factorial lay-out, the two factor variables are type of seed and type of extract. The deviance for the full model is 33.28 and the Pearson χ^2-statistic is 31.65 with 17 degrees of freedom, yielding a correction factor of $\surd(31.65/17) = 1.4$ for the standard errors.

The ratio of the deviance to the degrees of freedom is almost 2, close to indicating overdispersion by my criterion. In agreement with this, the AIC for the binomial model is 58.9 compared with 58.8 for the beta–binomial model, indicating that this overdispersed model does not fit better. (A frequentist test can also be constructed; the P-value for rejecting the null hypothesis of a binomial distribution in favour of the beta–binomial distribution is 0.28.) The estimate of the correlation is $\hat{\rho} = 0.0124$. There is considerable variability in the n_{ij} but their average is about 40 so the maximum likelihood estimate of the heterogeneity factor is about $1 + 39 \times 0.0124 = 1.48$ compared with 1.96 from the average deviance. Thus, the corrections to the standard error are respectively about 1.2 and 1.4.

The parameter estimates and standard errors for the two models are given in Table 3. We see that the variance inflation factor calculated from the goodness-of-fit statistics overcorrects the standard errors of the binomial distribution, compared with those from the beta–binomial distribution. But, in addition, even the larger standard errors of the beta–binomial distribution are not supported by the data because the simpler binomial distribution fits as well. In such situations, one generally chooses the simplest model.

The normed profile likelihood for the correlation in the beta–binomial model has a maximum at 0.0124. Although this point estimate is different from 0, a zero correlation is very plausible, having a normed likelihood of about 0.5. The correlation estimated from the goodness-of-fit statistic is about $1/39 = 0.0256$, a value that is also plausible according to the normed profile likelihood (0.65), but a value that is rather far from the maximum likelihood estimate.

If the corrected standard errors were used for inference, we would conclude that the interaction, and then the seed effect, could be removed from the model. However, the AIC is 61.7 for the binomial model without seed and its interaction, compared with 58.9 for the full model, given above, and respectively 60.2 and 58.8 for the corresponding beta–binomial models. From both comparisons, we conclude that the seed effect is necessary in the model. The same conclusions are obtained if the normal–binomial, double-binomial or multiplicative binomial distributions are fitted.

The conclusions to be drawn from these data are ambiguous. The ratio of deviance to degrees of freedom is close to 2 and the AICs from the binomial and overdispersed distri-

Table 3. Parameter estimates and their standard errors from the binomial and beta–binomial distributions†

Parameter	*Binomial model*			*Beta–binomial model*	
	Estimate	*Standard error*	*Corrected standard error*	*Estimate*	*Standard error*
Intercept	−0.558	0.126	0.176	−0.542	0.164
Seed	0.146	0.223	0.312	0.097	0.274
Vegetable	1.318	0.178	0.249	1.320	0.234
Interaction	−0.778	0.306	0.428	−0.798	0.378

†The corrected standard errors have been inflated by using the heterogeneity factor.

butions are similar, indicating that there might be some lack of fit. None of these four overdispersion models may adequately be capturing the variability in these data. For example, a more heavy-tailed mixing distribution may be necessary because of the great variability among counts or, preferably, the adequacy of control of the experimental conditions should be verified.

In contrast, in the first of Crowder's examples, the factor variable is the dilution of the extract, with three levels. Here the deviance is 48.22 with 13 degrees of freedom. The binomial and beta–binomial AICs are respectively 51.8 and 49.2, indicating an improved fit with the latter model.

5.2. Poisson counts

Lindsey (1993), pages 166–173, analysed data on the numbers of species in the Galapagos Islands. For the model with a non-linear response surface for area, elevation and distance from Santa Cruz, the deviance is 21.4, and the Pearson χ^2-statistic is 21.3, with 11 degrees of freedom, again indicating substantial overdispersion. However, the AICs for the Poisson and the negative binomial models are respectively 80.7 and 81.7 indicating a poorer fit for the latter. The former is better than the saturated model, with a different mean for each island: it has an AIC of 81.0. (Again, a frequentist test can be constructed for rejecting the Poisson distribution in favour of the negative binomial, here with a *P*-value of 0.99.) Here, as for the blackgrouse, the parameter estimates and standard errors are virtually identical in the two models. And yet the usual correction to the standard errors would have inflated them by a factor of 1.4.

For the negative binomial model with the variance proportional to the mean, the heterogeneity factor is estimated to be virtually 1: $\hat{\phi} = 1 + \exp(-9.36)$. The normed profile likelihood for this parameter of the negative binomial model decreases monotonously from its maximum at this value. The value of ϕ (1.94) estimated from the goodness-of-fit statistic is relatively implausible, with a normed likelihood of about 0.05.

The same conclusion about the lack of overdispersion is drawn if the normal–Poisson, double-Poisson or multiplicative Poisson distributions are fitted.

6. Discussion

The above examples indicate that the 'rule of thumb' that overdispersion is present if the deviance is greater than its number of degrees of freedom should be revised so that one only looks for overdispersion if the deviance is at least twice the number of degrees of freedom. Such a criterion has a theoretical justification: it is equivalent to comparing the fitted model with the saturated model by using AIC. However, once the *possibility* of overdispersion is determined, more formal methods of model building should be used to check whether it actually is present and, if so, to account for it. As we have seen in the examples, the usual point estimate obtained from the deviance may be misleading: the average deviance generally overestimates the heterogeneity factor, often greatly, compared with the maximum likelihood estimate from the negative binomial or beta–binomial distribution.

Evidence has accumulated indicating that standard errors are generally not trustworthy for making inferences from models that are not based on the normal distribution. We see here that corrections for overdispersion can also make inferences based on them even more misleading. Inflating standard errors when this is not supported by the data can lead to the non-detection of real treatment effects. It can also waste resources because experiments are conducted with larger sample sizes than necessary because the precision that will be obtained has been underestimated.

Ad hoc corrections are not acceptable when exact models for overdispersion can be fitted in a few seconds and various such models compared to check the robustness of results. Correction using a heterogeneity factor should be replaced by modern modelling techniques, although situations are not always as simple as in the examples presented here. The dispersion may vary systematically with the covariates so this will have to be modelled in its own right. Lindsey (1999b) gives examples where overdispersion only occurs in biologically extreme conditions, requiring response surface models both for the probability of an event and for the correlation among events.

Acknowledgements

I thank Robert Gentleman and Ross Ihaka for developing the R software (Ihaka and Gentleman, 1996) in which two functions were developed to fit the models to the data in the examples. These functions (`gnlr` and `glmm`, the latter for the normal mixtures) are available in my public R libraries respectively called `gnlm` and `repeated` on the complete R archive (available via `ftp.stat.math.ethz.ch`).

References

Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. In *Proc. 2nd. Int. Symp. Inference Theory* (eds B. N. Petrov and F. Csàki), pp. 267–281. Budapest: Akadémiai Kiadó.
Altham, P. M. E. (1978) Two generalizations of the binomial distribution. *Appl. Statist.*, **27**, 162–167.
Bai, Z. D., Krishnaiah, P. R., Sambamoorthi, N. and Zhao, L. C. (1992) Model selection for log-linear models. *Sankhya* B, **54**, 200–219.
Buckland, S. T., Burnham, K. P. and Augustin, N. H. (1997) Model selection: an integral part of inference. *Biometrics*, **53**, 603–618.
Burnham, K. P., White, G. C. and Anderson, D. R. (1995) Model selection strategy in the analysis of capture-recapture data. *Biometrics*, **51**, 888–898.
Cox, D. R. (1983) Some remarks on overdispersion. *Biometrika*, **70**, 269–274.
Crowder, M. J. (1978) Beta-binomial anova for proportions. *Appl. Statist.*, **27**, 34–37.
Efron, B. (1986) Double exponential families and their use in generalized linear regression. *J. Am. Statist. Ass.*, **81**, 709–721.
Finney, D. J. (1971) *Probit Analysis*. Cambridge: Cambridge University Press.
——— (1978) *Statistical Method in Biological Assay*. London: Griffin.
Hinde, J. (1982) Compound Poisson regression models. In *GLIM82* (ed. R. Gilchrist), pp. 109–121. Berlin: Springer.
Ihaka, R. and Gentleman, R. (1996) R: a language for data analysis and graphics. *J. Comput. Graph. Statist.*, **5**, 299–314.
Lindsey, J. K. (1993) *Models for Repeated Measurements*. Oxford: Oxford University Press.
——— (1996) *Parametric Statistical Inference*. Oxford: Oxford University Press.
——— (1999a) Some statistical heresies (with discussion). *Statistician*, **48**, 1–40.
——— (1999b) Response surfaces for overdispersion in the study of the conditions for fish eggs hatching. *Biometrics*, **55**, in the press.
Loneux, M., Lindsey, J. K. and Ruwet, J. C. (1997) Influence du climat sur l'évolution de la population de tétras lyres *Tetrao tetrix* dans les Hautes-Fagnes de Belgique de 1967 à 1996. *Cah. Ethol.*, **17**, 345–386.
McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*, 2nd edn. London: Chapman and Hall.
Raftery, A. E. (1986a) A note on Bayes factors for log-linear contingency table models with vague prior information. *J. R. Statist. Soc.* B, **48**, 249–250.
——— (1986b) Choosing models for cross-classifications. *Am. Sociol. Rev.*, **51**, 145–146.
Schwarz, G. (1978) Estimating the dimension of a model. *Ann. Statist.*, **6**, 461–464.