

Analysis of the human sex ratio by using overdispersion models

J. K. Lindsey†

Limburgs Universitair Centrum, Diepenbeek, Belgium

and P. M. E. Altham

University of Cambridge, UK

[Received December 1996. Final revision April 1997]

Summary. For study of the human sex ratio, one of the most important data sets was collected in Saxony in the 19th century by Geissler. The data contain the sizes of families, with the sex of all children, at the time of registration of the birth of a child. These data are reanalysed to determine how the probability for each sex changes with family size. Three models for overdispersion are fitted: the beta-binomial model of Skellam, the 'multiplicative' binomial model of Altham and the double-binomial model of Efron. For each distribution, both the probability and the dispersion parameters are allowed to vary simultaneously with family size according to two separate regression equations. A finite mixture model is also fitted. The models are fitted using non-linear Poisson regression. They are compared using direct likelihood methods based on the Akaike information criterion. The multiplicative and beta-binomial models provide similar fits, substantially better than that of the double-binomial model. All models show that both the probability that the child is a boy and the dispersion are greater in larger families. There is also some indication that a point probability mass is needed for families containing children uniquely of one sex.

Keywords: Akaike information criterion; Beta-binomial distribution; Direct likelihood inference; Double-binomial distribution; Finite mixture model; 'Multiplicative' binomial distribution; Non-linear Poisson regression; Overdispersion

1. Introduction

The study of the human sex ratio is greatly indebted to the data collected by Geissler (1889) on the distributions of the sexes of children in families in Saxony during 1876–1885. The data contain the sizes of families, with the sex of all children, at the time of registration of the birth of a child. They, thus, do not necessarily refer to complete families because the parents could have more children in the future. We assume that all children born to the family are included, and not just those still alive at the time of the most recent birth, although this is not clear.

These data involve almost 1 million birth registrations and about 3.7 million births. Many scientists have studied these data; Edwards (1958) provided a critical discussion of the historical references. They present some interesting problems. For example, during the 10-year period, many parents will have had several children and their families will be included more than once, a problem that is essentially insurmountable given the available information. Secondly, the rule determining when parents stop having children may vary with family size,

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

depending more heavily on the sex balance as the size of family grows: parents with an extreme sex imbalance may continue to have children for longer. Thus, the last birth may be particularly suspect as far as providing an unbiased estimate of the biological sex ratio.

Two essential questions for the sex ratio are whether the probability that the child is a boy varies among families and whether it can vary over time within a family. As is well known, for aggregated data such as those that we have available from the Geissler study, these cannot be distinguished. They will both manifest themselves as overdispersion with respect to a binomial model. Thus, we can only ask whether the variability changes with the size of family; if it does, it could be the result of either or both of the factors mentioned. There is common agreement that the overdispersion cannot be explained by multiple births (twins).

Because there is a suspicion of bias for the last child, often analyses are performed on the data without the last recorded birth. This can create some confusion. Thus, for example, Sokal and Rohlf (1969), p. 80, used the data for the first 12 children of families of size 13, whereas Fisher (1958), p. 67, took the complete data for families of size 8.

Edwards (1958) provided a table with the frequencies, without the last recorded birth, for families of size 2–13, for all possible combinations of boys and girls. It is reproduced in Table 1. Families of the same size lie along diagonals. For example, the data used by Sokal and Rohlf (1969), p. 80, can be found on the uppermost diagonal, whereas those of Fisher (1958), p. 67, correspond to the families on the diagonal for seven children. For convenience below, we shall denote by n , varying from 1 to 12, the number of children in each family as included in Table 1; this means that the total number of children in the family, at the time of recording, is $n + 1$.

Edwards (1958) fitted a beta–binomial distribution individually to each family size, i.e. to each diagonal of the table, using the method of moments. He found a very good fit with a χ^2 -test and a rather uniformly increasing variance with family size. With the more powerful methods that are now available, an attempt to fit a global model to the data in the table appears useful. This will allow us to check exactly how the variance is changing with family size and also whether the probability that the child is a boy is changing. We also look at the possibility of there being couples able to have children of only one sex.

The analysis of these data has wider interest than simply elucidating the human sex ratio for several reasons. It provides the opportunity to compare three models that have been

Table 1. Geissler's data on the human sex ratio, from Edwards (1958)

<i>Boys</i>	<i>Girls</i>												
	0	1	2	3	4	5	6	7	8	9	10	11	12
12	7	—	—	—	—	—	—	—	—	—	—	—	—
11	24	45	—	—	—	—	—	—	—	—	—	—	—
10	30	93	181	—	—	—	—	—	—	—	—	—	—
9	90	287	492	478	—	—	—	—	—	—	—	—	—
8	264	713	1027	1077	829	—	—	—	—	—	—	—	—
7	631	1655	2418	2309	1801	1112	—	—	—	—	—	—	—
6	1579	3725	4948	4757	3470	2310	1343	—	—	—	—	—	—
5	3666	7908	9547	8498	6436	3878	2161	1033	—	—	—	—	—
4	8628	16340	17332	14479	10263	5917	3072	1540	670	—	—	—	—
3	20540	31611	30175	22221	13972	7603	3895	1783	837	286	—	—	—
2	47819	57179	44793	28630	15700	8171	3951	1776	722	275	104	—	—
1	114609	89213	53789	28101	13740	6233	2719	1152	432	151	72	24	—
0	—	108719	42860	17395	7004	2839	1096	436	161	66	30	8	3

proposed for overdispersion in binomial data. For reasons discussed below, only one of these, the beta–binomial model, has been widely used. A second point is that, in contrast with most regression problems, such as with generalized linear models, here two regression equations, for the probability (location) and for the dispersion, are needed simultaneously, a family of models originally suggested by Lindsey (1974a) and independently by Pregibon (1984). We extend this to fitting a finite mixture model. Finally, methodology based on non-linear Poisson regression is introduced to estimate the models efficiently.

2. Models for overdispersion

The most commonly used model for overdispersion in binomial data is the beta–binomial model (Skellam, 1948):

$$f(y, \pi, \psi) = \binom{n}{y} \frac{B(\kappa + y, \nu + n - y)}{B(\kappa, \nu)}$$

$$= \binom{n}{y} \frac{B\{\pi \exp(\psi) + y, (1 - \pi) \exp(\psi) + n - y\}}{B\{\pi \exp(\psi), (1 - \pi) \exp(\psi)\}}$$

where n will be the family size for the Geissler data (in fact, the number of children, less 1) and y the number of boys, and $B(\cdot)$ is the beta function. This parameterization is convenient for comparison with the two to follow. The correlation is $\rho = 1/\{\exp(\psi) + 1\}$, with positive values (i.e. $\psi < \infty$) indicating overdispersion (Lindsey (1993), pages 159–160). This distribution can be derived from the binomial distribution if the Bernoulli probability is thought to vary in the population according to a beta distribution, and the marginal distribution taken. Although the binomial distribution is a member of the exponential family, the beta–binomial distribution is not.

Two members of the exponential family have, however, been proposed to handle binomial overdispersion. Unfortunately, both have intractable normalizing constants and thus have not yet been widely used. They may nevertheless be fitted by the method of Lindsey and Mersch (1992), as explained by Lindsey (1995), pages 129–132.

Altham (1978) introduced two generalizations of the binomial distribution. That which she called the ‘multiplicative’ generalization is a member of the exponential family. It can be written

$$f(y; \pi, \psi) = c_2(\pi, \psi) \binom{n}{y} \pi^y (1 - \pi)^{n-y} \exp\{\psi y(n - y)\}$$

where $c_2(\pi, \psi)$ is the intractable normalizing constant. The distribution will be overdispersed for $\psi < 0$, with $\psi = 0$ yielding the usual binomial distribution. This parameter, as -2ψ , is the logarithmic conditional cross-product for the sexes of any pair of births given all the others. A parameter in Cox’s (1972) quadratic binomial model has a similar interpretation (see Cox and Wermuth (1996), p. 80).

Efron (1986) proposed a family, which he called double exponential, that is a member of the exponential family. For overdispersed binomial data, the double-binomial distribution in this family may be appropriate:

$$f(y; \pi, \psi) = c_3(\pi, \psi) \binom{n}{y} \frac{n^{n\psi} \pi^{y(\psi+1)} (1 - \pi)^{(n-y)(\psi+1)}}{y^{y\psi} (n - y)^{(n-y)\psi}}$$

where $c_3(\pi, \psi)$ is again an intractable normalizing constant. Again, the distribution will be overdispersed for $-1 < \psi < 0$. Here, $1/(\psi + 1)$ has an approximate interpretation as the variance inflation factor.

In each of these models, we let the probability that the child is a boy vary with family size in the following way:

$$\begin{aligned}\theta_n &= \log\left(\frac{\pi_n}{1 - \pi_n}\right) \\ &= \alpha_0 + n\alpha_1,\end{aligned}\tag{1}$$

a logistic regression. Similarly, the dispersion parameter is allowed to depend on family size as follows:

$$\psi_n = \beta_0 + n\beta_1.\tag{2}$$

This completes the description of our models. Two problems remain: how can the models be fitted and how can they be compared?

Procedures for fitting the beta-binomial distribution are reasonably well known, at least for simple data sets. As already mentioned, fitting the other two distributions is considerably more difficult because of the normalizing constant $c_j(\pi, \psi)$. For one fixed n , i.e. one diagonal of our data table, they may be fitted by the method of Lindsey and Mersch (1992) using a standard linear Poisson regression model for the frequencies; see Lindsey (1995), pages 131–132, for fitting the double-binomial distribution to the diagonal with $n = 12$. The frequencies are regressed on the sufficient statistics, y and $y(n - y)$ for the multiplicative binomial distribution and y and $y \log(y) + (n - y) \log(n - y)$ for the double-binomial distribution. (That given by Lindsey (1995), p. 131, contains an error.) The normalizing constant is estimated by the intercept of the regression equation, the constant $(1 - \pi)^n$ also being absorbed by it. (In what follows, this fitting procedure, using Poisson regression, should not be confused with the regression equations (1) and (2), used in the models being fitted.)

Using this method, Lindsey and Laurent (1996) fitted the Efron double-Poisson regression model to overdispersed Poisson count data. They allowed both the location and the dispersion parameters to vary over time according to regression equations similar to those given above. For the Geissler data, the problem is more complex because the normalizing factor $(1 - \pi)^n$ is not constant.

The Poisson regression required to fit these models may be written

$$\log(\nu_{yn}) = \delta_n - \log\{f(y_n; \pi_n, \psi_n)\}\tag{3}$$

where ν_{yn} is the (Poisson) mean frequency of families of size n with y_n boys, $f(\cdot)$ is one of the three distributions described above and δ_n corresponds to a factor variable with a different level for each family size. δ_n fixes the total number of families of each size at the observed value, ensuring that each distribution (for each diagonal) is properly normalized (Lindsey (1995), pages 44 and 127).

For simple models in the exponential family, equation (3) is linear in the unknown parameters (except for the normalizing constant that does not involve the observations and is estimated by the intercept). However, for our models, it is non-linear, even for the binomial distribution, because of the term $n \log(1 - \pi)$ mentioned above, where n varies from 1 to 12. Thus, a non-linear Poisson regression function is required here to fit the models by using equation (3).

To compare the models, we use direct likelihood methods (Edwards, 1972; Lindsey, 1974b, 1996). For each model fitted, we give the -2 log-likelihood based on the Poisson distribution, so that smaller values are preferable. Deviances for nested models can be obtained by subtraction.

Because the models contain differing numbers of parameters, we use a standard model selection procedure, penalizing complexity by using the Akaike information criterion (AIC) (Akaike, 1973). This means that twice the number of parameters estimated will be added to each -2 log-likelihood in comparing them. Smaller values are still to be preferred. Such a method is generally appropriate only for the reasonably small sample sizes that are usually encountered, i.e. those obtained by a sample size calculation appropriate to detect an effect of scientific interest. Although in other cases it will generally indicate too complex a model, we nevertheless use it here with nearly 1 million observations.

3. Modelling the human sex ratio

The saturated Poisson regression (log-linear model) for the table, with 90 parameters (the number of frequencies in the table), has -2 log-likelihood = 842.4 (including the factorial of the Poisson distribution here and in all following likelihoods). The binomial distribution, ignoring overdispersion, has a common estimate for all family sizes, for the log-odds, $\hat{\theta} = 0.0596$, or, for the probability, $\hat{\pi} = 0.5149$, with -2 log-likelihood = 1866.5. As a non-linear Poisson regression model, it has 13 parameters, 12 of which (δ_n) fix the family sizes, leaving 77 degrees of freedom, as expected. These are the two extreme models providing points of comparison.

We first fit the beta-binomial model and compare our global model with the results obtained by Edwards (1958) for separate family sizes. The results are in the first part of Table 2. (As indicated there, some of the parameter estimates are multiplied by 100 for clarity.)

Table 2. Parameter estimates†, likelihoods and AIC values for Geissler’s data

<i>Beta-binomial model</i>				
$\hat{\alpha}_0$	0.060 (0.001)	0.049 (0.002)	0.059 (0.001)	0.050 (0.002)
$100\hat{\alpha}_1$	—	0.186 (0.038)	—	0.173 (0.039)
$\hat{\beta}_0$	4.580 (0.036)	4.580 (0.036)	5.400 (0.125)	5.387 (0.124)
$100\hat{\beta}_1$	—	—	-11.000 (1.431)	-10.866 (1.424)
-2 log-likelihood	1033.1	1011.0	970.7	950.4
AIC	1061.1	1041.0	1000.7	982.4
<i>Multiplicative binomial model</i>				
$\hat{\alpha}_0$	0.057 (0.001)	0.051 (0.002)	0.057 (0.001)	0.051 (0.002)
$100\hat{\alpha}_1$	—	0.115 (0.036)	—	0.104 (0.036)
$100\hat{\beta}_0$	-1.950 (0.066)	-2.006 (0.066)	-0.523 (0.199)	-0.527 (0.199)
$100\hat{\beta}_1$	—	—	-0.193 (0.025)	-0.197 (0.025)
-2 log-likelihood	1009.7	1001.4	953.6	948.5
AIC	1037.7	1031.4	983.4	980.5
<i>Double-binomial model</i>				
$\hat{\alpha}_0$	0.060 (0.001)	0.051 (0.002)	0.059 (0.001)	0.052 (0.002)
$100\hat{\alpha}_1$	—	0.067 (0.038)	—	0.142 (0.038)
$100\hat{\beta}_0$	-3.150 (0.140)	-3.131 (0.140)	0.527 (0.316)	0.528 (0.316)
$100\hat{\beta}_1$	—	—	-0.857 (0.060)	-0.862 (0.060)
-2 log-likelihood	1375.9	1356.4	1028.6	1009.1
AIC	1403.9	1386.4	1058.6	1041.4

†Standard errors are given in parentheses. Each model has 12 parameters (δ_j) fixing the family sizes, in addition to those shown.

From the likelihoods, we see that regressions are required for both the probability and the dispersion; both parameters change with family size. The four models are all much better than the binomial distribution. The last two, with a regression for the dispersion parameter, are considerably superior to the saturated model with 90 parameters according to the AIC: 1000.7 and 982.4, compared with 1022.4. The dispersion parameter is estimated to be less than 1, decreasing with increasing family size, indicating increasing variability.

The variance of the Bernoulli probability π , that is assumed to have a beta distribution in the construction of the beta-binomial distribution, is

$$\sigma_n^2 = \frac{\pi_n(1 - \pi_n)}{\exp(\psi_n) + 1}$$

where π_n and ψ_n vary according to family size following respectively equations (1) and (2). This provides an indication of how the dispersion is changing. The values for the different family sizes are in Table 3, along with those calculated separately for each family size, as given by Edwards (1958). The probabilities from the global model, and the separate probabilities given by Edwards, are also shown, as are the correlations estimated from the global model. We see that all these parameters increase with family size, with our estimates of the variance changing more rapidly than those of Edwards.

Table 2 also provides the fits and parameter estimates for the models based on the other two distributions, the double- and multiplicative binomial models; the latter gives the best fit of the three. With 16 parameters, it, like the beta-binomial model, fits exceptionally well for such a large sample size: the AIC value is here 980.5 compared with 1022.4 for the saturated model. In contrast, the standard asymptotic significance test for goodness of fit of this model is extremely misleading: $\chi^2_{74} = 106.2$, from the difference in log-likelihoods (for the reason, see Lindsey (1996), pages 104–110, 302–303 and 310–311).

As with the beta-binomial model, both the probability and the dispersion increase with family size in both of these models. Unfortunately, because of the intractable normalizing constant, the variance is difficult to calculate.

For all three distributions, the log-odds parameters without the regression have virtually identical estimates whereas those with the regression are very similar. In contrast, the

Table 3. Probabilities, variances and correlations from the beta-binomial distribution, according to family size, for Geissler's data†

Family size	$\hat{\pi}_n$	$\hat{\pi}_n$	$\hat{\sigma}_n^2$	$\hat{\sigma}_n^2$	$\hat{\rho}_n$
2	0.5134	0.5138	0.00141	0.00185	0.00565
3	0.5138	0.5144	0.00157	0.00139	0.00630
4	0.5143	0.5141	0.00175	0.00127	0.00702
5	0.5147	0.5141	0.00195	0.00180	0.00781
6	0.5151	0.5149	0.00217	0.00236	0.00870
7	0.5156	0.5147	0.00242	0.00260	0.00969
8	0.5160	0.5156	0.00270	0.00290	0.01079
9	0.5164	0.5178	0.00300	0.00298	0.01202
10	0.5168	0.5174	0.00334	0.00320	0.01338
11	0.5173	0.5171	0.00372	0.00384	0.01489
12	0.5177	0.5192	0.00414	0.00378	0.01657

†For each of π_n and σ_n^2 , the first column is the estimate given by the global model fitted here and the second column is the individual value calculated by Edwards (1958).

estimates of the parameters in the dispersion regression are unfortunately not comparable, making interpretation difficult.

We can, at least, conclude that the probability that the child is a boy increases on average with family size. In addition, within a given family size, this probability varies in unknown ways among families, more so for larger families. Combined, these two results may be indications either that the probability that the child is a boy is increasing with the order of birth for individual families or that there is a difference among families such that those with higher probability of giving birth to a son tend to have larger families.

One additional hypothesis of interest is that some couples are only capable of having children of one sex. This can be modelled by using a finite mixture model with equal point masses for families with either all girls or all boys:

$$(1 - \xi)z + \xi f(y; \pi, \psi)$$

where ξ is the mixing probability and $z = 1$ if the family has all boys or all girls and $z = 0$ otherwise. We add this mixture to the models fitted above for the beta-binomial and the multiplicative binomial models with two regression equations. The -2 log-likelihood values are respectively 949.3 and 944.6, both slightly better than for the corresponding models without the mixture. The parameters for the latter model are now $\hat{\alpha}_0 = 0.052$, $100\hat{\alpha}_1 = 0.0969$, $100\hat{\beta}_0 = -0.555$ and $100\hat{\beta}_1 = -0.173$ (with the same standard errors as in Table 2). The estimate of the logit of the mixing probability is -8.310 (standard error 0.254), yielding an estimated probability of $1 - \hat{\xi} = 0.000246$ for one-sex-children only families. If we add a third regression equation to the model, allowing the mixing probability to change with family size, similar to equation (1) but for ξ instead of π , the fit is not improved at all. Thus, we have some evidence for the existence of such families, although they are very rare. (This probability, none-the-less, refers to almost 250 families in the data set.)

It is intriguing to see what happens when we fit a mixture model to the data set for complete families of size 8 (Fisher (1958), p. 67), the same families for which the first seven children are shown in Table 1. In contrast with the results above, the beta-binomial and multiplicative binomial models here have identical log-likelihoods, 52.7 without and 41.0 with a mixture, as differences from the saturated model. (For the first seven children, the log-likelihoods are 6.3 and 5.7 for the beta-binomial model.) With 6 and 5 degrees of freedom, the models do not fit well to the complete families. The estimated mixing probability is not the same as above, both being considerably larger, respectively 0.001540 and 0.000699. This stronger evidence of a mixture in complete families of size 8 seems to indicate that a sub-population of single-sex families might be a social, not a genetic, phenomenon, although this conclusion is questionable because of the poor fit.

In social terms, we might argue that some of the parents who originally had small families, those with very unequal sex ratios, might continue to have children so that the distribution of the remaining small families would be underdispersed. For the same reason, larger families would be overdispersed because equally divided families would be under-represented. Without rather strong assumptions (Edwards, 1958), these data provide little evidence of this phenomenon as, with all models, they are overdispersed for all family sizes.

4. Discussion

The appearance of some families several times in the data set obviously makes our models incorrect, at least for comparisons among families of different sizes. (A family cannot be represented twice on the same diagonal.) Without further information, however, we cannot

judge to what extent this is true, or how it might affect the excellent fit of the beta-binomial and multiplicative binomial models. If it has an influence, it should primarily be on the slope parameters for family size.

The exceptionally good fit of the multiplicative binomial model for these data raises the question of its interpretation. Does it have some biological significance similar to the idea for the beta-binomial model of the Bernoulli probability varying across the population according to a beta distribution? Can more be said about the parameters than the empirical statement that the probability and dispersion both increase with family size? Given that ψ in the multiplicative binomial model measures the logarithmic conditional cross-product between pairs of births, it is not surprising that it changes with family size nor that the model provides a similar fit to the beta-binomial model with its correlation between pairs of births. But does the way in which these dependence parameters change have biological, or social, meaning?

In these data, the last birth has been removed because of possible social 'biases' in continuing to have children, as has usually been done in the literature. However, Geissler (1889) also provided the comparable data for all births, including the last, for all family sizes. Fitting the same sequence of models to the complete families of sizes 2–13 changes the results very little, including the estimate of the mixing probability. This indicates that inclusion of the last birth makes little difference and contradicts the results for families of size 8 mentioned earlier. It may be a clue that the role of biological factors is relatively more important than that of social factors.

For these data, the multiplicative binomial model fitted somewhat better than did the beta-binomial model. For other data sets, the same might occur with the double-binomial model. This suggests that these two models should be much more widely used. In spite of the intractable normalizing constant, in standard problems, these two models are much easier to fit than the beta-binomial model because this can be done with linear Poisson regression in any statistical package by using the method of Lindsey and Mersch (1992).

For the problems involving a series of different distributions, as we have here for different family sizes, the fitting procedure is still fairly straightforward, despite the non-linear Poisson regression.

The non-linear Poisson regression function was programmed in R (Ihaka and Gentleman, 1996), which uses the non-linear optimization algorithm of Dennis and Schnabel (1983). (GAUSS, GENSTAT or S-PLUS would also be appropriate languages for this problem.) The optimization algorithm calculates derivatives numerically so that only the minus log-likelihood function must be supplied. In this way, models with from 13 to 18 parameters, all treated as being non-linear, were fitted. As a check, the beta-binomial models were also fitted directly, in the usual way without the Poisson regression, giving the same results. Because of the high dimensionality of the problem, we have no absolute guarantee of having reached the optimum for the models based on the other two distributions that cannot be checked in this way, but the similarity of results is a good indication.

As in most non-linear optimization problems, the initial values are crucial. Estimates from an approximate linear Poisson regression were used. In all three models, the regression parameters for the probability are almost information orthogonal to those for the dispersion, whereas those within the same regression equation are highly correlated. This can create problems for the optimization routine which varies only one parameter at a time, not knowing that two need to be changed simultaneously to improve the likelihood. Orthogonalizing the parameters (Ross (1990), p. 17, and Lindsey (1996), pages 109 and 234–240) will usually aid in obtaining convergence.

Acknowledgements

We thank Robert Gentleman and Ross Ihaka for developing the R software and Anthony Edwards for discussion of the Geissler data.

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.
- Cox, D. R. (1972) The analysis of multivariate binary data. *Appl. Statist.*, **21**, 113–120.
- Cox, D. R. and Wermuth, N. (1996) *Multivariate Dependencies: Models, Analysis, and Interpretation*. London: Chapman and Hall.
- Dennis, J. E. and Schnabel, R. B. (1983) *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. New York: Prentice Hall.
- Edwards, A. W. F. (1958) An analysis of Geissler's data on the human sex ration. *Ann. Hum. Genet.*, **23**, 6–15.
- (1972) *Likelihood: an Account of the Statistical Concept of Likelihood and Its Application to Scientific Inference*. Cambridge: Cambridge University Press.
- Efron, B. (1986) Double exponential families and their use in generalized linear regression. *J. Am. Statist. Ass.*, **81**, 709–721.
- Fisher, R. A. (1958) *Statistical Methods for Research Workers*, 13th edn. Edinburgh: Oliver and Boyd.
- Geissler, A. (1889) Beiträge zur Frage des Geschlechtsverhältnisses der Geborenen. *Z. Königl. Sächs. Statist. Bur.*, **35**, 1–24.
- Ihaka, R. and Gentleman, R. (1996) R: a language for data analysis and graphics. *J. Comput. Graph. Statist.*, **5**, 299–314.
- Lindsey, J. K. (1974a) Construction and comparison of statistical models. *J. R. Statist. Soc. B*, **36**, 418–425.
- (1974b) Comparison of probability distributions. *J. R. Statist. Soc. B*, **36**, 38–47.
- (1993) *Models for Repeated Measurements*. Oxford: Oxford University Press.
- (1995) *Modelling Frequency and Count Data*. Oxford: Oxford University Press.
- (1996) *Parametric Statistical Inference*. Oxford: Oxford University Press.
- Lindsey, J. K. and Laurent, C. (1996) Estimating the proportion of lymphoblastoid cells affected by exposure to ethylene oxide through micronuclei counts. *Statistician*, **45**, 223–229.
- Lindsey, J. K. and Mersch, G. (1992) Fitting and comparing probability distributions with log linear models. *Comput. Statist. Data Anal.*, **13**, 373–384.
- Pregibon, D. (1984) Review of 'Generalized Linear Models' by McCullagh and Nelder. *Ann. Statist.*, **12**, 1589–1596.
- Ross, G. J. S. (1990) *Non-linear Estimation*. Berlin: Springer.
- Skellam, J. G. (1948) A probability distribution derived from the binomial distribution by regarding the probability of success as variable between sets of trials. *J. R. Statist. Soc. B*, **10**, 257–261.
- Sokal, R. R. and Rohlf, F. J. (1969) *Biometry: the Principles and Practice of Statistics in Biological Research*. San Francisco: Freeman.