Appl. Statist. (1975),
24, *No.* 2, p. 193

The Analysis of Permutations

By R. L. PLACKETT

Newcastle upon Tyne

[Received June 1974. Revised December 1974]

SUMMARY

A probability distribution is defined over the $r!$ permutations of r objects in such a way as to incorporate up to $r!-1$ parameters. Problems of estimation and testing are considered. The results are applied to data on voting at elections and beanstores.

Keywords: CATEGORICAL DATA; CONSUMER PREFERENCES; CONTINGENGY TABLES; LOGISTIC MODELS; LOGIT ANALYSIS; VOTING BIAS; RANDOM PERMUTATIONS

1. INTRODUCTION

IN 1965 I was consulted by a bookmaker with the following problem. When there are 8 or more runners in the field, a punter may bet that a specified horse will be placed, i.e. finish in the first three. The bookmaker then offers place odds. According to a method used at that time, the place odds were taken as one quarter of the odds against. For example, if the odds against A are 6–1 then the place odds are 6–4. However, this system for calculating place odds is often unfavourable to the bookmaker. My client had noticed the possibility, and the argument which follows is his. Consider a race with 8 horses which have exactly the same running ability, so that the finishing order is equally likely to be any of the permutations. The probability that A is placed is 3/8, so the real place odds are 5–3. On the other hand, the odds against A are 7–1, so the bookmaker's place odds are 7–4. Thus, when the profit factor is ignored, the bet is unfair to the bookmaker because if A is placed he will pay out £1·75 on £1 bet when he should pay only £1·67.

From the odds which are published in newspapers, we can calculate the "probability" of a win for each horse in a particular race. The total of the resulting quantities will often exceed 1, but if each of them is scaled down by the same factor, then we obtain estimated probabilities of winning. The question arises as to how the probability that a horse will be placed should be calculated when the probability that each horse wins is assumed to be known. A reasonable procedure is to define the probabilities of winning in a field from which any number of runners has been deleted as conditional probabilities based on those which remain. Suppose that the field consists of horses $1,2,\ldots,r$ with associated winning probabilities $p_1, p_2, \ldots, p_r$ where $\sum p_i = 1$. If horse 1 is withdrawn then the winning probabilities for $2,3,\ldots,r$ are taken to be respectively

$$p_2/(1-p_1),\, p_3/(1-p_1), \ldots, p_r/(1-p_1).$$

This assumption, and its obvious extensions, allow us to calculate the probability of any permutation (Plackett, 1968). Thus the probability for the order ijk is

$$p_i\{p_j/(1-p_i)\}\{p_k/(1-p_i-p_j)\}.$$

Another method of assigning a probability distribution to the permutations of $1,2,\ldots,r$ is as follows. Suppose that $W_1, W_2, \ldots, W_r$ are mutually independent random variables with means $\mu_1, \mu_2, \ldots, \mu_r$ respectively and unit variances. The probability of the permutation $ij\ldots k$ is taken to be

$$p(W_i < W_j < \ldots < W_k).$$

When the variables are normal, this involves the calculation of probabilities for an $(r-1)$-dimensional multivariate normal distribution where all the correlation coefficients immediately above and below the main diagonal of the correlation matrix are $-\frac{1}{2}$ and all others are zero.

A disadvantage of both methods is that $r!-1$ independent probabilities are expressed in terms of only $r-1$ parameters. In what follows, we construct a saturated model with $r!-1$ parameters, consider the problems of inference which arise for unsaturated models in the same class, and apply the results to practical examples.

2. Logistic Model

A typical probability is $p_{ijk\ldots l}$, where $ijk\ldots l$ is a permutation of $1,2,\ldots,r$. The following notation is used:

$$p_i = \sum_{jk\ldots l} p_{ijk\ldots l},$$

where the sum is taken over all permutations of $1,2,\ldots,r$ excluding i. Similarly

$$p_{ij} = \sum_{k\ldots l} p_{ijk\ldots l},$$

where the sum is taken over all permutations which exclude i and j. Thus p_i is the probability that i appears first, p_{ij} is the probability that i appears first and j appears second, and quantities $p_{ijk}, \ldots$ are similarly defined. We require to express the $r!-1$ independent probabilities in terms of $r!-1$ parameters which have a hierarchical structure of the type familiar from analysis of variance.

An appropriate system can be developed as follows. The permutations are first subdivided according to the integer which appears in the leading position. This subdivision is based on the probabilities $p_1, p_2, \ldots, p_r$ or an equivalent system such as

$$\kappa = \log p_r \quad \text{and} \quad \lambda_a = \log(p_a/p_r) \quad (a = 1,2,\ldots,r-1).$$

Thus $\lambda_a = 0$ for all a implies that $p_i = 1/r$ for all i. At the second stage, we subdivide the permutations where the first integer is i by parameters based on the probabilities $\{p_{ij}\}$. A model which uses only parameters from the first stage is given by

$$p_{ij}/p_{ij'} = p_j/p_{j'}.$$

Departures from the simple model can be expressed by quantities which are analogous to the cross-product ratios of contingency tables. Write

$$s = \max(1,2,\ldots,r \text{ excluding } i),$$

and define

$$\lambda_{ib} = \log(p_{ib}\, p_s / p_{is}\, p_b) \quad (b = 1,2,\ldots,r \text{ excluding } i,s).$$

At the third stage, we consider the probabilities $\{p_{ijk}\}$ where i and j are fixed. Write

$$t = \max(1,2,\ldots,r \text{ excluding } i,j),$$

and define

$$\lambda_{ijc} = \log(p_{ijc}\, p_{jt}/p_{ijt}\, p_{jc}) \quad (c = 1, 2, \ldots, r \text{ excluding } i, j, t).$$

This fulfils the requirement of a hierarchical model in that only parameters from the second stage appear when those at the third stage are all zero. The process continues along the same lines and terminates with the $(r-1)$th stage. A count of the number of parameters introduced at each stage is given in Table 1, together with the cumulative

TABLE 1

Parameters in logistic model classified by stage

	Individual		*Cumulative*	
Stage	*Type*	*Number*	*Type*	*Number*
0	κ	1		
1	λ_a	$r-1$	p_i	r
2	λ_{ib}	$r(r-2)$	p_{ij}	$r(r-1)$
3	λ_{ijc}	$r(r-1)(r-3)$	p_{ijk}	$r(r-1)(r-2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$r-1$	$\lambda_{ij\ldots kd}$	$r(r-1)(r-2)\ldots 3$	$p_{ijk\ldots l}$	$r!$

totals. The overall total is $r!$, and $r!-1$ parameters are independent because κ is determined by the λ's and the requirement of total probability. A model which contains parameters up to and including stage q but none thereafter is described as a qth-order model, where $q = 0, 1, 2, \ldots, r-1$. When the permutations consist of only $m\,(<r-1)$ of the integers $1, 2, \ldots, r$, the same system of models is appropriate with $q = 0, 1, 2, \ldots, m$.

Some known distributions $\{p_{ijk\ldots l}\}$ now appear as special cases. A zero-order model is obtained when all the parameters $\{\lambda_a\}, \{\lambda_{ib}\}, \{\lambda_{ijc}\}, \ldots$ are zero, in which case

$$p_{ijk\ldots l} = 1/r!.$$

A first-order model is obtained when all the parameters $\{\lambda_{ib}\}, \{\lambda_{ijc}\}, \ldots$ are zero, in which case

$$p_{ijk\ldots l} = p_i\, p_j\, p_k \ldots p_l/(1-p_i)(1-p_i-p_j)(1-p_i-p_j-p_k)\ldots p_l.$$

A second-order model may be illustrated by taking $r = 4$. We find, for example, that

$$p_{1234} = p_{12}\, p_{23}/(p_{23}+p_{24}).$$

In terms of the parameters which have been introduced, this becomes

$$p_{1234} = \frac{\exp(\kappa+\lambda_1+\lambda_2+\lambda_3+\lambda_{12}+\lambda_{23})}{\{1+\exp(\lambda_2+\lambda_{12})+\exp(\lambda_3+\lambda_{13})\}\{1+\exp(\lambda_3+\lambda_{23})\}}.$$

We note that the parameters $\{\lambda_a\}, \{\lambda_{ib}\}, \ldots$ can be defined starting with the last place of the permutation and working back to the first one, or indeed in terms of any fixed permutation applied to the whole group. While there may be situations in which one of these definitions is more relevant than the definition we have chosen, the latter will usually seem the most natural.

All the parameters except κ are the logarithms of probability ratios, or the differences of such quantities, and for that reason the models are described as logistic. The interpretation of logistic models can be illustrated by reference to an opinion poll in which people are asked to put in order of priority a specified list of possible actions for the British Government to take. According to a zero-order model, the public are indifferent about the order of priority. With a first-order model, some actions are preferred to others, but the degree of preference remains the same irrespective of the choices which have already been made. In this sense there is no association between the actions with respect to the extent that priorities are assigned to them. When we come to a second-order model, such associations are admitted, but only between pairs of successive choices. In the case of a third-order model, the associations extend over three consecutive choices, and similarly for models of higher order, until ultimately the permutations have no discernible structure.

3. Inference

Denote the frequency of the permutation $ijk \ldots l$ by $n_{ijk\ldots l}$, and the total frequency by n. Partial sums $n_i, n_{ij}, \ldots$ are defined as before. The random variable corresponding to $n_{ijk\ldots l}$ is $N_{ijk\ldots l}$. We suppose that $N_{ijk\ldots l}$ has a Poisson distribution with parameter $\mu_{ijk\ldots l}$, and that the $\{N_{ijk\ldots l}\}$ are mutually independent. Distributions conditional on $N=n$ are multinomial with probabilities

$$p_i = \mu_i/\mu, \quad p_{ij} = \mu_{ij}/\mu, \quad \ldots,$$

where $\mu_i, \mu_{ij}, \ldots$ are partial sums and μ is the sum of $\mu_{ijk\ldots l}$ over all permutations.

Consider the estimation of the parameters $\{\lambda_a\}$ in a first-order model. Denote by m_i the frequency of permutations which end with i. Then the kernel of the likelihood is

$$(\Pi\, p_i)^n / \Pi\, p_i^{m_i}\, \Pi\, (1-p_i)^{n_i}\, \Pi\, (1-p_i-p_j)^{n_{ij}}\, \Pi\, (1-p_i-p_j-p_k)^{n_{ijk}} \ldots,$$

where the last product in the denominator involves $r-2$ different subscripts. The maximum of the likelihood can be determined only by numerical methods. Write T_{xy} for the frequency with which place x of the permutation is occupied by integer y. When $r=3$, the frequencies $\{T_{xy}\}$ are sufficient for $\{p_i\}$. They can be presented as a 3×3 contingency table with marginal totals all equal to n.

Another estimation procedure for the first-order model is as follows. The fact that p_i is the probability that a permutation begins with i suggests the estimation of λ_a by $\log(N_a/N_r)$. However, information about p_i is also available from the second place of a permutation, which provides a sequence of estimators typified by $\log(N_{ia}/N_{ir})$ where $i \neq a$ and $i \neq r$. These estimators, together with $\log(N_a/N_r)$ are mutually independent (Plackett, 1974, p. 38). At the next stage, we obtain a sequence of estimators typified by $\log(N_{ija}/N_{ijr})$, where i, j, a, r are all different. The process terminates at the $(r-1)$th stage, and all the estimators in the sequences are mutually independent because each variable $N_{ijk\ldots l}$ appears exactly once. We now combine the estimators. The asymptotic variance of $\log(N_a/N_r)$ is estimated by $N_a^{-1}+N_r^{-1}$. A refinement is to replace $N_{ijk\ldots}$ throughout by $N_{ijk\ldots}+\frac{1}{2}$, which reduces the bias with respect to both mean and variance (Gart and Zweifel, 1967). The estimator finally obtained is

$$\lambda_a^* = \frac{\Sigma\{(N_{ij\ldots ka}+\frac{1}{2})^{-1}+(N_{ij\ldots kr}+\frac{1}{2})^{-1}\}^{-1} \log\{(N_{ij\ldots ka}+\frac{1}{2})/(N_{ij\ldots kr}+\frac{1}{2})\}}{\Sigma\{(N_{ij\ldots ka}+\frac{1}{2})^{-1}+(N_{ijk\ldots r}+\frac{1}{2})^{-1}\}^{-1}},$$

where the summations are taken over all permutations $ij \ldots k$ consisting of $0, 1, 2, \ldots$ integers selected from $1, 2, \ldots, r-1$ excluding a.

Three statistics have been proposed to test a simple hypothesis against a composite alternative in large samples (Rao, 1965, p. 349). One of them is based on derivatives of the log likelihood function, and has the advantage that the calculation of the maximum likelihood estimate is not required. The corresponding statistic for testing that $\lambda_a = 0$ for all a against the alternative that $\lambda_a \neq 0$ for some a is derived by Downton (1972).

Consider next the estimation of the parameters $\{\lambda_a\}$ and $\{\lambda_{ib}\}$ in a second-order model. These parameters are together equivalent to $\{p_{ij}\}$. Information about p_{ij} is obtainable from $\log(N_{ij}/N_{r,r-1})$, $\log(N_{gij}/N_{gil})$, ... and the estimators in the various sequences are combined in the same way as before. Similar remarks apply to models of higher order. The logical conclusion of this process is to consider a saturated model with as many parameters as permutations. Each parameter $\lambda_{ij\ldots a}$ is estimated by $\hat{\lambda}_{ij\ldots a}$, obtained from the substitution of $N_{ijk\ldots l}$ for $p_{ijk\ldots l}$ in the appropriate expression. The results can be set out in a form similar to an analysis of variance table by evaluating the statistic

$$Z^2 = (\hat{\lambda}_{ij\ldots a})^2/(\text{est. var. } \hat{\lambda}_{ij\ldots a})$$

for each parameter $\lambda_{ij\ldots a}$. Inspection of the table will suggest which parameters can be omitted from the saturated model. We can then calculate revised estimates of the surviving parameters along the lines described.

4. Applications

Both the examples which follow have $r = 3$. A convenient notation, for this section only, is

$$p = p_1, \quad q = p_2, \quad r = p_3, \quad \alpha = p_1/p_3, \quad \beta = p_2/p_3,$$

$$\lambda = p_{13}\, p_2/p_{12}\, p_3, \quad \mu = p_{21}\, p_3/p_{23}\, p_1, \quad \nu = p_{32}\, p_1/p_{31}\, p_2.$$

Example 1. The 1973 Local Government Elections in England and Wales are studied by Upton and Brook (1974) and Brook and Upton (1974). Some of their data refer to parties with 3 candidates in the Tyne–Wear area. The observed frequencies of the 6 possible within-party alphabetical orders are shown in Table 2.

Table 2

Orders of three candidates in elections

Order	123	132	213	231	312	321	*Total*
Observed frequency	232	136	174	151	114	141	948
Expected frequency (N)	199	159	193	151	125	121	948
Expected frequency (L)	201	143	204	149	124	127	948

N = normal model. L = logistic model.

They fit a distribution to the permutations based on normal random variables with different means, and the corresponding expected frequencies are labelled N in the table. The model is justified by a normal approximation for the total number of votes cast.

An analysis of the saturated logistic model is given in Table 3, where each value of Z^2 should be referred to a table of χ_1^2. There are five values, so we compare them with the 1 per cent point, namely 6·63. This comparison suggests a first-order model.

TABLE 3

Analysis of saturated model for election data

Parameter	Z^2
α	20·23
β	8·39
λ	4·53
μ	2·67
ν	5·22

The likelihood of a first-order model has a maximum when

$$\hat{p} = 0{\cdot}362, \quad \hat{q} = 0{\cdot}373, \quad \hat{r} = 0{\cdot}265.$$

On the other hand, the alternative method of estimation described in the previous section gives

$$p^* = 0{\cdot}355, \quad q^* = 0{\cdot}379, \quad r^* = 0{\cdot}266.$$

The expected frequencies calculated from the maximum likelihood estimates are labelled L in Table 2. Both sets of expected frequencies are thus based on models with two parameters.

Fig. 1 displays an isometric contour plot of the function

$$f(p,q,r) = -1/\log \quad (\text{kernel of likelihood}),$$

which is positive or zero according as the likelihood is positive or zero respectively. Here p,q,r are the distances of a point from the sides of an equilateral triangle. The figure shows where the function $f(.)$ falls to specified percentages of the maximum value, from which we see that the likelihood surface is very flat. A computer program for plotting a function of two variables is described by Evans (1973), and has been extended by him to plot in a triangular co-ordinate system. Maximum likelihood estimates may be calculated by using a simplex procedure of function minimization (O'Neill, 1971). First approximations can be obtained from isometric contour plots by inspection.

Example 2. A consumer preference study at Mississippi State College involving three varieties of snap beans is reported by Anderson (1959). One lot of each variety was displayed in retail stores, and customers were asked to rank the beans in order of choice. The data from one store on one day are shown in Table 4.

Suppose that p_{xy} is the probability that place x of a permutation is occupied by integer y. According to the zero-order model,

$$p_{xy} = 1/3 \quad (x, y = 1, 2, 3).$$

Anderson tests this hypothesis, which is decisively rejected. We therefore proceed to fit a first-order model. The observed frequencies $\{t_{xy}\}$ are sufficient for p,q,r and Fig. 2 shows an isometric contour plot of the function $f(p,q,r)$. Maximum likelihood estimates obtained by function minimization are

$$\hat{p} = 0{\cdot}436, \quad \hat{q} = 0{\cdot}168, \quad \hat{r} = 0{\cdot}396.$$

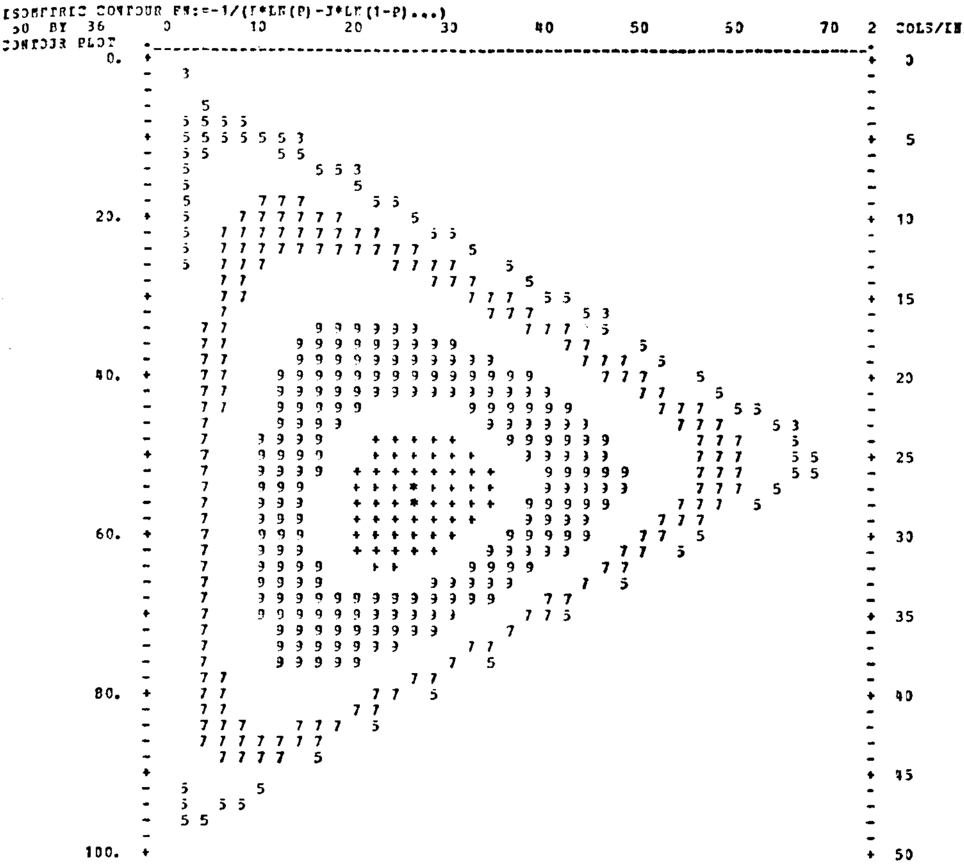

FIG. 1. Isometric contour plot for election data.

The ratio $(f - f_{\min})/(f_{\max} - f_{\min})$ is indicated by contour zone codes as follows.

3	4	5	6	7	8	9		+	*
25%	35%	45%	55%	65%	75%	85%	95%	97·5%	99·9%	100%

The expected frequencies corresponding to $\{t_{xy}\}$ are shown in brackets in Table 4.

We cannot make an analysis on the lines of Table 3 because the frequencies of the individual permutations are unknown. The parameters of a second-order model can be estimated efficiently only by direct maximization of the likelihood function. When available, the first method of analysis is preferable. A close inspection of the data is involved and the calculations are straightforward.

TABLE 4

Ranks of three variates of beans

	Rank			
Variety	1	2	3	*Total*
1	42 (54)	64 (46)	17 (23)	123
2	31 (21)	16 (30)	76 (73)	123
3	50 (49)	43 (47)	30 (27)	123
Total	123	123	123	

Expected frequencies in brackets.

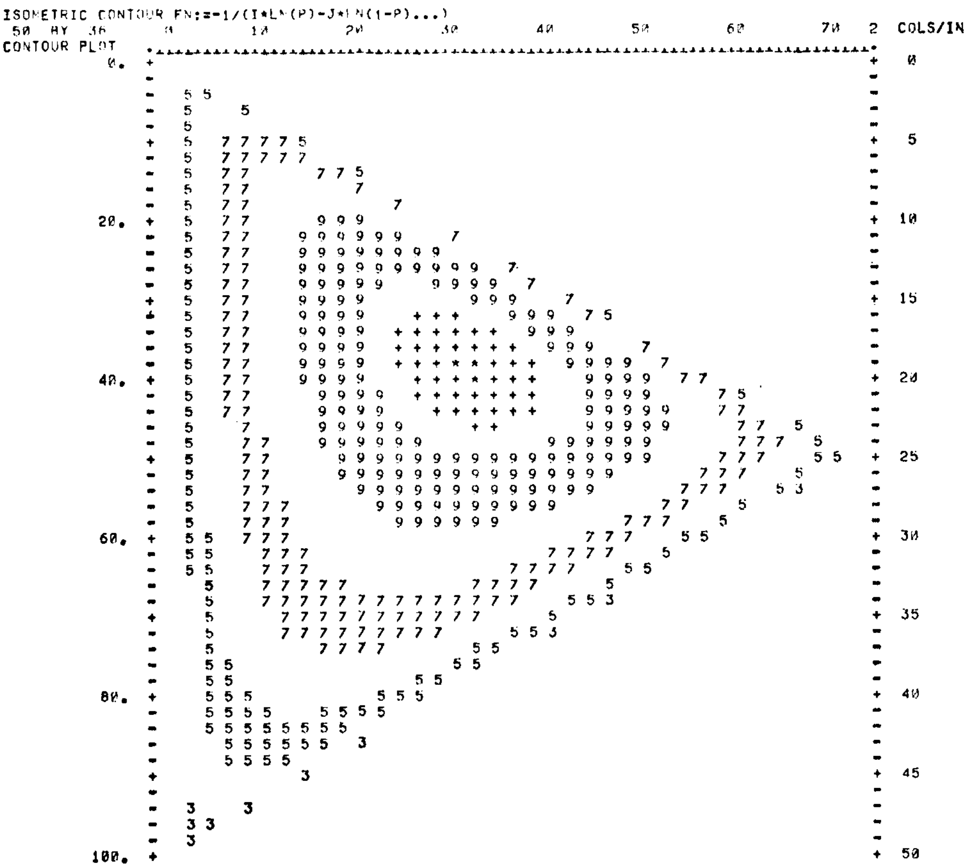

FIG. 2. Isometric contour plot for beanstore data.

The ratio $(f-f_{\min})/(f_{\max}-f_{\min})$ is indicated by contour zone codes as follows.

3	4	5	6	7	8	9	+	*
25%	35%	45%	55%	65%	75%	85%	95%	97·5%	99·9%	100%

Denote the frequency of permutation 123 by x. Then the frequencies of other permutations are as follows.

123	132	213	231	312	321
x	$42-x$	$30-x$	$1+x$	$34+x$	$16-x$

Consequently, $0 \leqslant x \leqslant 16$, and the kernel of the likelihood function is

$$l = \sum_{x=0}^{16} p_{12}^{x}\, p_{13}^{43-x}\, p_{21}^{30-x}\, p_{23}^{1+x}\, p_{31}^{34+x}\, p_{32}^{16-x}.$$

Now substitute

$$p_{12} = pq/(q+\lambda r), \quad p_{23} = qr/(r+\mu p), \quad p_{31} = rp/(p+\nu q),$$

$$p_{21} = \mu pq/(r+\mu p), \quad p_{32} = \nu qr/(p+\nu q), \quad p_{13} = \lambda pr/(q+\lambda r).$$

We finally obtain

$$l = \frac{p^{106} q^{47} r^{93} \lambda^{42} \mu^{30} \nu^{16}}{(q+\lambda r)^{42}(r+\mu p)^{31}(p+\nu q)^{50}} \sum_{x=0}^{16} (\lambda\mu\nu)^{-x}.$$

The maximum values of $\log l$ for various models are shown in Table 5, together with the corresponding values of the parameters. Thus the likelihood-ratio statistic for comparing the expected frequencies of a first-order model with those observed has a value of 38 (3 d.f.).

TABLE 5

Maximum values of log likelihood for various models

Order of model	p	q	r	λ	μ	ν	*log* l
Second	0·378	0·198	0·424	2·495	4·055	0·827	−181·282
First	0·436	0·168	0·396	1	1	1	−200·183
Zero	1/3	1/3	1/3	1	1	1	−217·553

On the basis of these two examples, a first-order logistic model should be viewed with some reserve as a satisfactory description of voting patterns. The analysis of further sets of data may suggest what order of model is generally required.

ACKNOWLEDGEMENTS

I am grateful to Dennis Evans for computer plots and to Sheila Boyd for calculations. Comments made by referees are also appreciated.

REFERENCES

ANDERSON, R. L. (1959). Use of contingency tables in the analysis of consumer preference studies. *Biometrics*, **15**, 582–590.

BROOK, D. and UPTON, G. J. G. (1974). Biases in Local Government elections due to position on the ballot paper. *Appl. Statist.*, **23**, 414–419.

DOWNTON, F. (1972). Discussion on Professor Cox's paper. *J. R. Statist. Soc.* B, **34**, 202–205.

EVANS, D. A. (1973). The influence of computers on the teaching of statistics. *J. R. Statist. Soc.* A, **136**, 153–190.

GART, J. J. and ZWEIFEL, J. R. (1967). On the bias of various estimators of the logit and its variance with applications to quantal bioassay. *Biometrika*, **54**, 181–187.

O'NEILL, R. (1971). Function minimization using a simplex procedure. *Appl. Statist.*, **20**, 338–345.

PLACKETT, R. L. (1968). Random permutations. *J. R. Statist. Soc.* B, **30**, 517–534.

—— (1974). *The Analysis of Categorical Data.* London: Griffin.

RAO, C. R. (1965). *Linear Statistical Inference and Its Applications.* New York: Wiley.

UPTON, G. J. G. and BROOK, D. (1974). The importance of positional voting bias in British Elections. *J. Polit. Studies*, **22**, 178–190.