Journal of the Royal Statistical Society

SERIES B (METHODOLOGICAL)

Vol. XX, No. 2, 1958

THE REGRESSION ANALYSIS OF BINARY SEQUENCES

By D. R. COX

Birkbeck College, University of London

[Read before the RESEARCH SECTION of the ROYAL STATISTICAL SOCIETY, March 5th, 1958, Professor G. A. BARNARD in the Chair]

SUMMARY

A SEQUENCE of 0's and 1's is observed and it is suspected that the chance that a particular trial is a 1 depends on the value of one or more independent variables. Tests and estimates for such situations are considered, dealing first with problems in which the independent variable is preassigned and then with independent variables that are functions of the sequence. There is a considerable amount of earlier work, which is reviewed.

1. INTRODUCTION

Suppose that there are for analysis one or more series of trials, the observation on any one trial taking one of two forms, such as "success" or "failure", "defective" or "not-defective", and so on. Denote the possible observations by 0 and 1, each series of trials therefore giving a sequence of 0's and 1's. Suppose further that corresponding to each trial there are one or more independent variables and that we suspect that the probability that a particular trial gives say the outcome 1 depends on the corresponding values of the independent variables. In this paper we consider methods for estimating and testing such dependencies. If the observations were continuous instead of being (0, 1) variates, the problems would be treated by the standard methods of simple and multiple regression.

The following are some examples. Haldane and Smith (1948) have considered a test for a birth-order effect in the occurrence of hereditary abnormalities. The data here consist of a series for each family such as *N A A N*, meaning that the first child is normal, the second has the particular abnormality under study, and so on. The independent variable is the serial number of the birth. Alternatively, the independent variable may be maternal age and, in an extension of the problem, both maternal age and serial number are available and it is required to examine whether there is a dependence on birth order when any effect of maternal age is eliminated.

In certain learning studies in experimental psychology a task is attempted a number of times in succession and the primary observation is success or failure at each trial. We may wish to examine the dependence of the probability of success on such variables as the number of preceding trials, the number of previously rewarded successes, and so on. The last of these variables is rather difficult to deal with because it is not preassigned, but is a function of the outcomes of the preceding trials in the sequence.

Some standard problems appear as special cases of the situation considered here. Thus, if the independent variable takes on just two levels and if different trials are statistically independent, the numbers of 0's and 1's at the two levels can be written in a 2×2 contingency table and the usual theory is applicable. If the independent variable takes on k values at each of which there are a number of trials, the data can be written as a $2 \times k$ table, and the procedure that we shall obtain is to a first approximation equivalent to the standard method of isolating from χ^2 a single degree of freedom for linear regression on the column variable.

If the independent variable is at several levels corresponding to different doses of a drug and if we are interested in the "position" and "slope" of the dependence, we have the usual situation in assays with quantal response. In this paper, however, we shall be concerned with what, in continuous problems, would be the slope of the regression line, regarding the position of the line as a nuisance parameter. Hence the present paper is not of much direct relevance to assay work.

2. General Formulation

Let $Y_1, \ldots, Y_n$ be mutually independent random variables each taking the values 0 and 1 and let $x_1, \ldots, x_n$ be a set of fixed numbers. In the simplest form of our problem there is suspected to be a relation between $\theta_i = \text{pr}\,(Y_i = 1)$ and x_i. In a more complicated form there are two sets of fixed numbers $\{x_i\}$ and $\{u_i\}$ and we are concerned with the joint dependence between θ_i and x_i, u_i. In further forms of the problem we consider x_i which are functions of $Y_1, \ldots, Y_{i-1}$, such as Y_{i-1} and $Y_1 + \ldots + Y_{i-1}$, the Y_i therefore no longer being independent.

Consider first the case with a single preassigned independent variable. We need a parametric representation of the relation between θ_i and x_i that will have a simple intuitive meaning, that will be a reasonable approximation to relations likely to occur in practice and that will be manageable mathematically. A linear relation is unsuitable, except over a narrow range, because of the restriction that θ_i must lie in [0, 1] and, in the absence of special considerations for a particular problem, the best form seems to be the logistic law

$$\text{logit}\,\theta_i \equiv \log\{\theta_i/(1-\theta_i)\} = \alpha + \beta x_i, \tag{1}$$

i.e.

$$\theta_i \equiv \text{pr}\,(Y_i = 1) = e^{\alpha+\beta x_i}/(1 + e^{\alpha+\beta x_i}), \tag{2}$$

and

$$1 - \theta_i \equiv \text{pr}\,(Y_i = 0) = 1/(1 + e^{\alpha+\beta x_i}). \tag{3}$$

This form has been extensively used in work on bioassays, notably by Berkson.

As stated above, we shall be concerned with inference about β, regarding α as a nuisance parameter. The interpretation of β is that if θ_i is small β is the fractional increase in θ_i per unit increase in x_i, whereas if $1 - \theta_i$ is small, β is the fractional decrease in $1 - \theta_i$ per unit increase in x_i. If all the values of θ_i considered lie near θ^0, (2) represents a linear relation between θ_i and x_i with slope $\beta\theta^0(1 - \theta^0)$.

In more complicated problems, for example with several independent variables, we shall use natural generalizations of (1).

All the significance tests to be developed below for null hypotheses expressing the absence of regression, are non-parametric, in that the logistic law enters only into the derivation of the test criterion and not into the development of its sampling distribution under the null hypothesis.

3. Remarks on Sampling Without Replacement

It will appear in §4 that the distributional problems to be solved in connection with the present problems depend on sampling without replacement from a finite population. It is therefore convenient first to recall some results about such sampling.

Let $x_1, \ldots, x_n$ denote the finite population and let $X_{0,y}$ be a random variable defined as the sum of y numbers selected randomly without replacement.* The moments of $X_{0,y}$ are

$$m_{0,y} = E(X_{0,y}) = ym_1, \tag{4}$$

$$\sigma^2_{0,y} = V(X_{0,y}) = y(n-y)\, m_2/(n-1), \tag{5}$$

$$E[(X_{0,y} - m_{0,y})^3] = y(n-y)(n-2y)\, m_3/[(n-1)(n-2)], \tag{6}$$

$$E[(X_{0,y} - m_{0,y})^4] = \frac{y(n-y)}{(n-1)(n-2)(n-3)}\{(n^2 - 6ny + n + 6y_2)\, m_4 + 3(y-1)\, n(n-y-1)\, m_2^2\}, \tag{7}$$

where

$$m_1 = \Sigma\, x_i/n, \tag{8}$$

$$m_r = \Sigma\, (x_i - m_1)^r/n, \qquad r > 1. \tag{9}$$

These formulae are given, for example, by Irwin and Kendall (1944).

In numerical work it is preferable to calculate the m_r from the power sums about the origin:

$$m_1 = s_1/n, \tag{10}$$

$$m_2 = (ns_2 - s_1^2)/n^2, \tag{11}$$

$$m_3 = (n^2 s_3 - 3ns_2 s_1 + 2s_1^3)/n^3, \tag{12}$$

$$m_4 = m(n^3 s_4 - 4n^2 s_1 s_3 + 6ns_2 s_1^2 - 3s_1^4)/n^4, \tag{13}$$

where

$$s_r = \Sigma\, x_i^r.$$

For our theoretical purposes it is more useful, however, to write $f = y/n$ and to make an expansion with f fixed and n large. This gives

$$m_{0,y} = ym_1, \tag{14}$$

$$\sigma^2_{0,y} \sim ym_2(1 - f + 1/n), \tag{15}$$

$$\gamma_{1(0,y)} \sim \frac{(1-2f)\,\gamma_1}{\sqrt{\{f(1-f)\,n\}}}, \tag{16}$$

$$\gamma_{2(0,y)} \sim -\frac{6}{n} + \frac{[1 - 6f(1-f)]\,\gamma_2}{nf(1-f)}, \tag{17}$$

* The notation is chosen to fit in with the application considered in §4.

where γ_1 and γ_2 refer to cumulant ratios of the finite population. The term independent of γ_2 in (17) could be removed by redefinition of the γ_2's in terms of the K's of finite sampling theory instead of the m's.

Equations (16) and (17) suggest that in samples for which y and $n - y$ are both appreciable, $X_{0,y}$ will be nearly normally distributed and that the "central limit effect" will be relatively more rapid than in samples of size Min $(y, n - y)$ drawn with replacement. Rigorous proofs that as $n \to \infty$ with f fixed, $(X_{0,y} - m_{0,y})/\sigma_{0,y}$ converges in distribution function, under suitable restrictions, to the unit normal law are given by Madow (1948), Hoeffding (1951) and Motoo (1957); the first two authors, but not the last, require all moments of the population to behave suitably as $n \to \infty$.

For special forms of finite population, explicit expressions may be attained. Thus, if r of the x_i are equal to one and the remaining $(n - r)$ values are zero, $X_{0,y}$ has the hypergeometric distribution with probability generating function

$$\sum_s \frac{\binom{r}{s}\binom{n-r}{y-s}}{\binom{n}{y}} \zeta^s. \tag{18}$$

If $x_i = i$ for $i = 1, \quad \ldots \quad, n$, Haldane and Smith (1948) have shown that the probability generating function is

$$\frac{1}{\binom{n}{y}} \prod_{r=1}^{y} \left(\frac{\zeta^r - \zeta^{n+1}}{1 - \zeta^r}\right) \tag{19}$$

and have expressed the cumulants of $X_{0,y}$ in terms of Bernouilli's numbers.

A generalization, that will be needed in §5, concerns sampling from a bivariate finite population. We suppose that there are n pairs $(x_1, u_1), \quad \ldots \quad, (x_n, u_n)$, that a random sample of y pairs is drawn without replacement and that we consider the joint distribution of the random variables defined as before. The joint sampling cumulants of $X_{0,y}$ and $U_{0,y}$ do not seem to have been considered in the literature. The cumulants up to the third order are given by

$$\kappa_{11(0,y)} = E[(X_{0,y} - m_{0,y}{}^{(x)})(U_{0,y} - m_{0,y}{}^{(u)})] = y(n-y)\, m_{11}/(n-1), \tag{20}$$

$$\kappa_{21(0,y)} = E[(X_{0,y} - m_{0,y}{}^{(x)})^2(U_{0,y} - m_{0,y}{}^{(u)})] = \frac{y(n-y)(n-2y)\, m_{21}}{(n-1)(n-2)}, \tag{21}$$

where

$$m_{\alpha\beta} = \frac{1}{n} \Sigma\, (x_i - m_1{}^{(x)})^\alpha\, (u_i - m_1{}^{(u)})^\beta. \tag{22}$$

The simplest way to prove say (21) is to note first that the expectation must be symmetric in i, of degree (2, 1) and invariant under separate translations of the x_i and the u_i and hence must be of the form $A(y, n)\, m_{21}$. To determine A, suppose that $x_i = u_i$, when m_{21} becomes m_{30}, and the expectation is the third moment of $X_{0,y}$. The form of A follows immediately from (6). Mixed fourth moments can be calculated similarly.

4. SINGLE PREDETERMINED INDEPENDENT VARIABLE

4.1. *General theory.*—Let $x_1, \ldots, x_n$ be, as in §2, fixed quantities. The likelihood of observations $Y_1, \ldots, Y_n$ is

$$\exp\{\Sigma(\alpha+\beta x_i)\,Y_i\}/\Pi\,(1+e^{\alpha+\beta x_i}) = e^{\alpha Y+\beta X}/\Pi\,(1+e^{\alpha+\beta x_i}), \tag{23}$$

where $Y = \Sigma\, Y_i$, $X = \Sigma\, Y_i x_i$. Thus X, Y are jointly sufficient for the parameters α, β, as was noted by Garwood (1941).

Further, it is clear that Y, which is the total number of 1's, can tell us nothing about β, which is concerned with the way the distribution of 1's changes with x_i. Hence, to make an inference about β, we argue conditionally on the observed value of Y. The arguments for doing this as a matter of principle raise interesting general issues analogous to the corresponding ones for the 2×2 contingency table. The point will not be discussed here.

Consider the distribution of X conditionally on the observed value of Y, which will be denoted by y. Now

$$\operatorname{pr}(X=x,\ Y=y) = \frac{N_{x,y}\, e^{\alpha y+\beta x}}{\Pi\,(1+e^{\alpha+\beta x_i})} \tag{24}$$

where $N_{x,y}$ is the number of distinct ordered sets of y numbers, taken from $x_1, \ldots, x_n$, whose sum is x. Therefore

$$\operatorname{pr}(Y=y) = \frac{\Sigma_x\, N_{x,y}\, e^{\alpha y+\beta x}}{\Pi\,(1+e^{\alpha+\beta x_i})}, \tag{25}$$

where Σ_x denotes summation over all values taken by the random variable X. Finally, if we take the ratio of (24) to (25), we have that

$$\operatorname{pr}(X=x \mid Y=y) = \frac{N_{x,y}\, e^{\beta x}}{\Sigma_x\, N_{x,y}\, e^{\beta x}}, \tag{26}$$

and the nuisance parameter α has been eliminated. When $\beta = 0$, (26) is the distribution of the random variable $X_{0,y}$ of §3. If $X_{\beta,y}$ denotes the random variable X considered conditionally on y when β is the true parameter value, we have by (26) that

$$\operatorname{pr}(X_{\beta,y}=x) = \frac{N_{x,y}}{\binom{n}{y}}\,\frac{e^{\beta x}}{M_{0,y}(\beta)}, \tag{27}$$

where $M_{0,y}(s)$ is the moment generating function of $X_{0,y}$.

Thus the moment generating function of $X_{\beta,y}$ is given by

$$M_{\beta,y}(s) = \frac{M_{0,y}(\beta+s)}{M_{0,y}(\beta)}, \tag{28}$$

and the corresponding relation for cumulant generating functions is

$$K_{\beta,y}(s) = K_{0,y}(\beta+s) - K_{0,y}(\beta). \tag{29}$$

In particular it follows from (29) that the cumulants of $X_{\beta,y}$ for general values of β are given by the derivatives of the cumulant generating function for $X_{0,y}$ evaluated at β instead of at the origin.

If a manageable explicit expression were available for the distribution of $X_{0,y}$ or for its moment generating function, direct calculation of, or approximation to, the properties of $X_{\beta,y}$ would be possible. But in general there seems little that can be done except to obtain from (29) a series expansion for the cumulants of $X_{\beta,y}$ in ascending powers of β; thus

$$\kappa_{\beta,y}^{(1)} = E(X_{\beta,y}) = m_{0,y} + \beta\sigma^2_{0,y} + \frac{1}{2}\beta^2\sigma^3_{0,y}\,\gamma_{1(0,y)} + \ldots, \tag{30}$$

$$\kappa_{\beta,y}^{(2)} = V(X_{\beta,y}) = \sigma^2_{0,y} + \beta\sigma^3_{0,y}\,\gamma_{0(0,y)} + \ldots, \tag{31}$$

etc., the general result being that

$$\kappa_{\beta,y}^{(r)} = \sum_{t=0}^{\infty} \kappa_{0,y}^{(t+r)} \frac{\beta^t}{t!}. \tag{32}$$

Bennett (1956) has obtained a special case of (29) and the resulting expansions. His choice of criterion was intuitive.

Another approximation to (29), appropriate when $y/n << 1$, is to regard $X_{0,y}$ as the sum of a random sample drawn with replacement (Haldane and Smith, 1948). This can be useful when the cumulant generating function of the population can be expressed simply.

It follows from the asymptotic normality of $X_{0,y}$ that a sufficient condition that $X_{\beta,y}$ is nearly normal when y and $n-y$ are both large is that $\beta\sqrt{(m_2 y)} = o(1)$. In fact the simplest approximation based on (30) and (31) is to take $X_{\beta,y}$ as normally distributed with mean and variance

$$m_{0,y} + \beta\,\sigma^2_{0,y} \text{ and } \sigma^2_{0,y}, \tag{33}$$

respectively. This can also be derived directly from (27), by using the normal approximation to

$$N_{x,y}\bigg/\binom{n}{y}$$

and at the same time approximating to $M_{0,y}(\beta)$ by the moment generating function of a normal random variable. Further terms of the expansions (30), etc., can be recovered by an Edgeworth expansion of (27). A normal approximation involved in calculating the power function of the exact test in a 2×2 contingency table was obtained by Patnaik (1948) using a similar argument; Stevens (1951) pointed out that if the parameter corresponding to β is not small the approximation can be badly wrong.

4.2. *The* 2×2 *contingency table.*—If the x_i take on two values, say 0 and 1, the statistic X is the entry in the lower right hand corner of the 2×2 contingency table :

		x_i	
		0	1
Y_i	0		
	1		

and the conditional distribution (27) reduces to the distribution given by Fisher (1935) as appropriate for inference about the parameter $e^{\beta} = \psi$, where

$$\psi = \frac{\text{pr}(Y_i = 1 \mid x_i = 1)}{\text{pr}(Y_i = 0 \mid x_i = 1)} \times \frac{\text{pr}(Y_i = 0 \mid x_i = 0)}{\text{pr}(Y_i = 1 \mid x_i = 0)} . \tag{34}$$

The significance test of the null hypothesis $\beta = 0$ based on the hypergeometric distribution (18) of X is, of course, the familiar exact test of association in the 2×2 table.

As noted in §4.1, the normal approximation to the non-null distribution of X based on the first terms of (30) and (31) has been given by Patnaik. An alternative normal approximation, based on an expansion around the mode of the distribution of X, has been obtained by Cornfield (1956), and this should be much the more satisfactory method when $|\beta|$ is large.

To compare the different approximations, consider the example analysed by Cornfield.

Example 1.—The data in Table 1 were obtained in a survey of physicians.

TABLE 1

Distribution of Physicians

	Controls	*Lung Cancer Patients*
Smokers	32	60
Non-smokers	11	3

Cornfield (1956) has obtained as 95 per cent. confidence limits for ψ the values $0 \cdot 030$ and $0 \cdot 623$; he shows that the corresponding exact tail area probabilities are $3 \cdot 8$ per cent. and $2 \cdot 4$ per cent. respectively, instead of the required $2 \cdot 5$ per cent. His procedure involves the iterative solution of a quartic equation.

The simplest approximate solution by the methods of §4.1 uses the first terms of (30) and (31), that is regards X as normally distributed with mean $m_{0,y} + \beta\sigma^2_{0,y}$ and variance $\sigma^2_{0,y}$, where X is observed to be 3, $y = 14$ and the finite population sampled consists of 43 values with $x_i = 0$ and 63 with $x_i = 1$, so that $n = 106$. Hence by (10) and (11)

$$m_1 = 63/106 = 0 \cdot 5943, \; m_2 = m_1(1 - m_1) = 0 \cdot 2411,$$

so that by (4) and (5)

$$m_{0,y} = 8 \cdot 321, \; \sigma^2_{0,y} = 2 \cdot 957, \; \sigma_{0,y} = 1 \cdot 720.$$

With an observed value of 3, 95 per cent. confidence limits for the true mean based on a normal approximation with a continuity correction are $(-0 \cdot 871, 6 \cdot 871)$ and since the true mean is, by (33), approximately $m_{0,y} + \beta\sigma^2_{0,y}$, it follows that the limits for β are $(-3 \cdot 109, -0 \cdot 490)$. Thus the limits for $\psi = e^{\beta}$ are $(0 \cdot 045, 0 \cdot 613)$, compared with Cornfield's values of $(0 \cdot 030, 0 \cdot 623)$.

Although the limits calculated by this simple approximation appear to agree remarkably well with Cornfield's values, there is a serious difficulty connected with the lower value. For the random variable X is non-negative, so that the approximation which gives $-0 \cdot 871$ as a possible value for its expectation cannot be sensible (see Stevens, 1951). This particular difficulty is unlikely to arise if none of the entries in the contingency table is small.

The effect of taking additional terms in the expansion for the cumulants of $X_{\beta,y}$ is discouraging. We find directly from (12) and (16) that $\gamma_{1(0,y)} = -0\cdot08112$ and is so small that the inclusion of terms of an Edgeworth expansion to replace the normal multiplier $1\cdot96$ is hardly worth considering. (The true skewness of $X_{\beta,y}$ for the smaller values of β in the confidence range is large and positive.) If we assume $X_{\beta,y}$ to be normal with mean

$$m_{0,y} + \beta\sigma^2{}_{0,y} + \frac{1}{2}\beta^2\sigma^3{}_{0,y}\,\gamma_{1(0,y)}$$

and variance

$$\sigma^2{}_{0,y} + \beta\gamma_{1(0,y)}\,\sigma^3{}_{0,y},$$

we obtain as new confidence limits (0·0624, 0·643), and the left-hand limit is worse than that given by the first approximation. The reason is that, as already noted, the first inproximation to the value of γ_1 has the wrong sign. I have not investigated the effect of including further terms in the series expansions.

Example 1 has been dealt with in some detail, partly because the problem of getting a simple method of calculating limits for the odds-ratio in a 2×2 table is of intrinsic interest, and partly in order to illustrate the type of difficulty to be expected in general in applying the formulae of §4.1. The general conclusion is that the first approximation has worked well up to a point and that the refinement of the approximation by the inclusion of further terms is likely to be unsuccessful when one of the observed frequencies in the table is small. It seems reasonable to expect that the inclusion of further terms will be useful if none of the frequencies is small and β is not large, but the extensive numerical investigation that is desirable to specify at all precisely just when the method works has not been undertaken.

4.3. *The $2 \times k$ contingency table.*—A direct generalization of the situation just considered, where the independent variable x took on two values, arises when the independent variable takes on a comparatively small number, k say, of values, there being several observations of Y_i at each value of x. That is, we have a $2 \times k$ contingency table, in which the k-fold classification carries a numerical score.

It is well known that in such cases a test of significance of trend can be obtained by extracting from χ^2_{k-1} for the whole table, a single degree of freedom for linear regression on x (Yates, 1948; Armitage, 1955; Cochran, 1955). This procedure is almost equivalent to the test based on (33); see especially Armitage's paper. In fact, Armitage obtains his test by a formal linear regression analysis of the (0, 1) variables Y_i on x_i, using the usual formulae of least-squares theory. It is clear, however, that with the x's and Y, the total number of 1's, regarded as fixed, X is equivalent to the linear regression coefficient.

Now as the number of observations Y_i recorded at each x value increases, the total at each x value becomes normally distributed so that the usual results about least-squares theory for normal random variables imply the asymptotic efficiency of the χ_1 test against linear alternatives. A new result from §4.1 is that the test statistic is uniquely appropriate for testing against all logistic alternatives, even in small samples, in that it uses the sufficient statistic X. A further point is that by regarding X in the way we have as the sum of a random sample obtained without replacement, a routine procedure is available for refining the normal approximation, or, if special precision is desired in measuring significance, the exact probability may be obtained by enumeration; see also Haldane (1940).

As noted in §2, the procedure is non-parametric in that the logistic assumption enters only into the derivation of the test statistic and not into the specification of the null hypothesis.

Example 2.—Suppose that there are seven equally spaced values of x, $-3, \ldots, 3$, and that eight trials are made at each value. The finite population of x values is thus discrete rectangular and formulae (10)–(13) give $m_1 = m_3 = 0$, $m_2 = 4$, $m_4 = 28$. Suppose further that of the 56 observed Y_i, there are 23 observations equal to 1 and 33 equal to 0. Then $n = 56$, $y = 23$ in the formulae (4)–(7) [or (14)–(17)] and we find that $m_{0,y} = 0$, $\sigma^2{}_{0,y} = 55 \cdot 2$ and that the fourth cumulant of $X_{0,y}$ is $-200 \cdot 4$, giving $\gamma_{2(0,y)} = -0 \cdot 066$. Of course $\gamma_{1(0,y)} = 0$.

Therefore significance is tested by seeing whether X, equal to $\sum_{-3}^{3} iy_i$, where y_i is the number of 1's when $x = i$, is significantly extreme in a distribution of mean 0, variance $55 \cdot 2$, $\gamma_1 = 0$ and $\gamma_2 = -0 \cdot 066$. The tables of Pearson and Merrington (1951) can be used to assess the change, due to non-normality, in the usual significance limits, or alternatively the Fisher-Cornish inversion of the Edgeworth series can be employed.

If we are to use the present method to estimate β, the expansions

$$E(X_{\beta,y}) = 55.2\,\beta - 33.4\,\beta^3 + \ldots,$$
$$V(X_{\beta,y}) = 55.2 - 100.2\,\beta^2 + \ldots$$

must be applied. The second terms are small relative to the first if $|\beta| \leqslant \frac{1}{4}$ so that we can reasonably expect good results from the terms just displayed only in this range. Otherwise other methods must be employed.

In general, at least four methods are available for interval estimation of the parameter β in this and similar situations. First it would be possible to work from the distribution (26) which is appropriate for inference about β and which is free of nuisance parameters. Unless the value of y is such that the possible values of X more extreme than the observed value are fairly small in number and are such that the $N_{x,y}$ can be found reasonably simply, this method is impracticable.

The second method is to use formulae based on (29), i.e. the expansions (30), etc. As has already been remarked, this approach gives linear regression formulae as a first approximation and these may be expected to work reasonably well when the slope is small and particularly when the theoretical probabilities lie between $\frac{1}{5}$ and $\frac{4}{5}$, over which range the logistic curve is substantially linear. In the examples discussed above the inclusion of a small number of further terms is not successful. As noted in §4.2, further investigation is desirable to find just when additional terms in these expansions can be put in with advantage.

The third and fourth methods do not use the conditional distribution (26). Maximum likelihood may be applied to (23) and tables for this, and a simple iterative calculating scheme, are given by Berkson (1957). The asymptotic standard error can be found in the usual way. The fourth method, which is not iterative, is that of minimum logit χ^2 (Berkson, 1955), i.e. the method of least squares with empirical weights applied to suitably transformed observations; again an asymptotic standard error can be found by the usual regression formulae. The last two methods are asymptotically equivalent; Berkson (1955) has shown for a range of special cases that from the point of view of minimizing the mean square error of a point estimate of β, the latter method is to be preferred. The relevance or otherwise of this for general practical use raises very interesting issues: see Berkson's papers and Silverstone (1957).

It is recommended that where the significance test is of primary importance the method of the present paper should be used and that if estimation is of primary importance and the sample size is not too small, one of the last two methods should be applied.

4.4. *The serial order test.*—If $x_i = i$ and if x is interpreted to correspond to serial order, the test statistic X is the sum of the rank numbers of those observations giving the outcome 1. The use of this as a test of randomness has been proposed by Haldane and Smith (1948), who gave tables for testing the statistic in small samples, as well as the large sample approximation to its distribution. They dealt also with the case when observations corresponding to some serial numbers are missing.

It follows from §4.1 that this test is optimum against logistic alternatives. It is also of interest to remark that the test amounts to Wilcoxon's two-sample test (see, for instance, Siegel, 1957, p. 72 and Birnbaum, 1956) applied to the ranks corresponding to (*a*) the 0's and (*b*) the 1's. This then is a situation where Wilcoxon's test is optimum, although the problem is different from that for which Wilcoxon's test is most commonly applied, namely the comparison of the locations of two samples. It would be interesting to know if there are any location problems for which Wilcoxon's test is optimum.

The approximate procedures analogous to those described in §4.3 lead to confidence intervals for β. Another approach, from which I have not succeeded in getting a useful working method, is to use (19) and (29) to obtain the cumulant generating function of $X_{\beta,y}$, to approximate to the resulting sums by integrals and hence to obtain an approximation to the distribution when Min $(y, n - y)$ is large and β is not necessarily small.

If the probability that any particular trial has outcome 1 is always small, so that $e^{\alpha+\beta i}/(1 + e^{\alpha+\beta i}) \sim e^{\alpha+\beta i} << 1$ and if x is thought of as time, we have very nearly a point process in continuous time. The test statistic is the sum of the times at which a 1 is observed and the procedure and set-up become those considered by Cox (1955).

4.5. *The combination of data from several series.*—In some applications, including that of the previous subsection, there may be several series for analysis, all assumed to have the same value of β. Thus in the illustrative example discussed by Haldane and Smith, the children of 51 families are classified as normal or phenylketonuric. The independent variable is the serial number of the birth, so that there are 51 series for combination.

If it is assumed that the probability that a child is normal depends only on the serial number of the birth, and is the same for all the families involved, we consider the total rank sum and test it conditionally on the grand total number of normals; this is in effect the procedure of §4.3 using a $2 \times k$ contingency table with the column classification determined by birth order.

The procedure used by Haldane and Smith depends on the total rank sum considered conditionally on the separate total numbers of normals, family by family. This is easily seen to be the theoretically optimum procedure when the nuisance parameters α are different for each family and the logistic regression coefficient β is constant. In fact if the affix j

refers to the j^{th} series (family), the likelihood is from (23)

$$\exp \{\Sigma\, \alpha^{(j)}\, Y^{(j)} + \beta\, \Sigma\, X^{(j)}\} / \Pi_i\, (1 + e^{\alpha^{(j)} + \beta x_{ij}^{(j)}}), \tag{35}$$

so that $Y^{(1)}, \ldots, Y^{(k)}, X = \Sigma\, X^{(j)}$ are the jointly sufficient statistics, and the nuisance parameters $\alpha^{(1)}, \ldots, \alpha^{(k)}$ are eliminated by considering the distribution of X conditionally on $Y^{(1)}, \ldots, Y^{(k)}$.

4.6. *Test with α known.*—So far we have supposed that α is an unknown nuisance parameter and this has led us to consider the distribution of X conditionally on the observed value of Y, the total number of 1's. Suppose now the null hypothesis is that 1's occur independently at constant known probability $\theta_0 = e^{\alpha_0}/(1 + e^{\alpha_0})$ and it is required for the test to be sensitive against alternatives in which the probability at the i^{th} trial depends on x_i. In particular if $x_i = i$, we require a test for trend from the value θ_0.

It follows from (13) that the simple sufficient statistic for β is again $X = \Sigma\, x_i Y_i$, but this time we work with its unconditional distribution, so that X is now the sum of n independent random variables. Under the null hypothesis

$$E(X) = \theta_0\, \Sigma\, x_i, \tag{36}$$

$$V(X) = \theta_0\, (1 - \theta_0)\, \Sigma\, x_i^2, \tag{37}$$

and the distribution is asymptotically normal as $n \to \infty$ provided, for example, that for some $r > 2$

$$\lim_{n \to \infty} \sum_{i=1}^{n} |\, x_i \,|^{r} / \{ \sum_{i=1}^{n} x_i^2 \}^{r/2} = 0. \tag{38}$$

This condition, which follows immediately from Lyapunov's form of the central limit theorem, covers in particular the case $x_j = j$. For small n, an exact test can be made quite easily.

We can examine the non-null distribution for small enough β either by replacing the logistic law by the asymptotically equivalent linear regression or by a direct argument based on approximating by the central limit theorem to the terms in the exact distribution of X. The result is that if as $n \to \infty$ the X's and β are such that $\beta^2 \sum_1^n x_i^2 = o(1)$, X is normal with mean $\theta_0\, \Sigma\, x_i + \beta\theta_0(1 - \theta_0)\, \Sigma\, x_i^2$ and variance $\theta_0(1 - \theta_0)\, \Sigma\, x_i^2$. Thus, as would be expected on general grounds, we get equivalent results in testing β in large samples by (a) regarding α as unknown and using §4.2 or by (b) estimating θ_0, i.e. α_0, from the average proportion of 1's and then treating this as a known true value, using the present method.

Example 3.—Suppose that an experimental task is attempted 9 times and that the correct response at each trial, which may be L or R, is determined by randomization. Success or failure is recorded at each trial and it is required to test the null hypothesis that successes occur randomly with probability $\frac{1}{2}$ against the alternative that there is a trend in the success rate. Suppose that the results of one run are $F\,S\,F\,S\,S\,F\,S\,S\,S$.

If success S is denoted by 1 and if $x_i = i$, i.e. we examine for regression on serial order, the test statistic is the sum of the ranks of the S's, i.e. 35. From (36) and (37) the theoretical mean is

$$\frac{\theta_0\, n(n+1)}{2} = 22\tfrac{1}{2}$$

and the theoretical variance

$$\theta_0(1-\theta_0)\,\frac{n(n+1)(2n+1)}{6} = \frac{285}{4}$$

giving a standard error of 8·44. Hence, after correcting for continuity, the deviation from expectation is 1·422 times the standard error, corresponding to significance at the $15\tfrac{1}{2}$ per cent. level in a two-sided test, using the normal approximation. A simple enumeration shows the exact level to be $41/256 = 0 \cdot 160$.

5. Two Predetermined Independent Variables

Suppose now that the Y_i are mutually statistically independent and that there are two independent regression variables for each trial, x_i and u_i say. We generalize the logistic law (2) into

$$\operatorname{pr}(Y_i = 1) = e^{\alpha+\beta x_i+\gamma u_i}/(1 + e^{\alpha+\beta x_i+\gamma u_i}), \tag{39}$$

where β and γ are partial regression coefficients. A direct generalization of (23) leads to the consideration of the joint distribution of $X = \Sigma\, x_i Y_i$ and $U = \Sigma\, u_i Y_i$, conditionally on the observed value of $Y = \Sigma\, Y_i$. Further the relation for the cumulant generating function corresponding to (29) is that

$$K_{\beta,\gamma,y}(s_1, s_2) = K_{0,y}(\beta + s_1, \gamma + s_2) - K_{0,y}(\beta, \gamma). \tag{40}$$

If we simplify the notation by denoting the cumulants of (X, U) when $\beta = \gamma = 0$ by κ_{ij}, instead of by $\kappa_{ij(0,y)}$ (section 3) we have

$$E(X - \kappa_{10}) = \kappa_{20}\beta + \kappa_{11}\gamma + \tfrac{1}{2}\kappa_{30}\beta^2 + \kappa_{21}\beta\gamma + \tfrac{1}{2}\kappa_{12}\gamma^2 + \ \ldots\ , \tag{41}$$

$$E(U - \kappa_{01}) = \kappa_{11}\beta + \kappa_{02}\gamma + \tfrac{1}{2}\kappa_{21}\beta^2 + \kappa_{12}\beta\gamma + \tfrac{1}{2}\kappa_{03}\gamma^2 + \ \ldots\ , \tag{42}$$

etc. If the second degree terms are dropped, we can construct unbiased estimates of β and γ and these are the estimates given by a formal linear regression analysis of the (0, 1) observations on the x's and the u's. Improved estimates may for small β and γ be found by equating the right hand sides of (41) and (42) to $X - \kappa_{10}$ and $U - \kappa_{01}$ respectively.

We shall consider the special problem of testing the null hypothesis $\beta = 0$ with γ an arbitrary nuisance parameter.

The method is to construct a statistic having expectation β, as far as the terms retained in (41) and (42). Let $X' = X - \kappa_{10}$, $U' = U - \kappa_{01}$ and write

$$\begin{pmatrix} \kappa_{20} & \kappa_{11} \\ \kappa_{11} & \kappa_{02} \end{pmatrix}^{-1} = \begin{pmatrix} \kappa^{20} & \kappa^{11} \\ \kappa^{11} & \kappa^{02} \end{pmatrix},$$

denoting by Δ the determinant of the maxtrix on the right. As noted above, the first approximations to unbiased estimates of β, γ are respectively

$$B_1 = \kappa^{20} X' + \kappa^{11} U', \; C_1 = \kappa^{11} X' + \kappa^{02} U'. \tag{43}$$

If we evaluate the expectation of B_1 using (41) and (42) and replace the second order terms by their sample estimates with sign reversed, we get for the required estimate of β

$$B_1 - \tfrac{1}{2}\Delta[(\kappa_{30}\kappa_{02} - \kappa_{21}\kappa_{11})(B_1^2 - \kappa^{20}) + 2(\kappa_{21}\kappa_{02} - \kappa_{12}\kappa_{11})(B_1C_1 - \kappa^{11}) + (\kappa_{12}\kappa_{02} - \kappa_{03}\kappa_{11})(C_1^2 - \kappa^{02})]. \tag{44}$$

The first approximation to the variance of (43) is κ^{20}; the second approximation is very complicated.

Alternative procedures analogous to those of §4.3 are to apply maximum likelihood or minimum logit χ^2; these result in standard multiple regression calculations, iterative in the first method, non-iterative in the second.

6. Regression on the Result of the Preceding Trial

In this section we deal briefly with the situation in which the probability of a 1 at the i^{th} trial may depend on the observation at the preceding trial. Suppose that the probability of a 1 following a 1 is

$$e^{\alpha+\beta}/(1 + e^{\alpha+\beta}) \tag{45}$$

and following a 0 is

$$e^{\alpha}/(1 + e^{\alpha}); \tag{46}$$

this is, of course, merely a special way of writing the transition probabilities in a simple Markov chain.

The likelihood of $Y_2, \ldots, Y_n$ given Y_1 is

$$\frac{e^{\alpha \sum_2^n Y_i + \beta \sum_2^n Y_i Y_{i-1}}}{(1 + e^{\alpha})^{\sum_2^n (1-Y_{i-1})}(1 + e^{\alpha+\beta})^{\sum_2^n Y_{i-1}}} \tag{47}$$

the likelihood having to be taken conditionally on Y_1, since the distribution of Y_1 is undefined. The expression (47) is complicated by end effects and the jointly sufficient set of statistics consists of Y_1 and Y_n in addition to $\Sigma\, Y_i$ and $\Sigma\, Y_i Y_{i-1}$. A formal simplification results from making the set-up circular, by defining $Y_0 = Y_n$. Then the likelihood is

$$\frac{e^{\alpha \Sigma Y_i + \beta \Sigma Y_i Y_{i-1}}}{(1 + e^{\alpha})^{n-\Sigma Y_i}(1 + e^{\alpha+\beta})^{\Sigma Y_i}}, \tag{48}$$

and $\Sigma\, Y_i = Y$ and $\Sigma\, Y_i Y_{i-1} = X$ are jointly sufficient. The distribution of X conditionally on $Y = y$ is

$$\begin{aligned} \operatorname{pr}(X = x \mid Y = y) &= \frac{M_{x,y}\, e^{\beta x}}{\Sigma\, M_{x,y}\, e^{\beta x}} \\ &= \frac{M_{x,y}\, e^{-\frac{1}{2}\beta w}}{\Sigma\, M_{x,y}\, e^{-\frac{1}{2}\beta w}}, \end{aligned} \tag{49}$$

say, where w is the total number of runs in the circular array of 0's and 1's, equal to $2(y-x)$, and where $M_{x,y}$ is the number of distinct samples with y 1's for which $X=x$. The difference from (26) is that the null distribution is that of the number of runs in random circular permutation of y 1's and $n-y$ 0's.

There is now the familiar difficulty, that arises also in the study of ordinary serial correlation coefficients, of deciding whether to deal in practice with a quantity defined artificially in order to simplify the mathematics. Unless there is some physical justification of he circular definition, it seems best to take the test statistic, W, equal to the number of runs of 0's and 1's in the observed non-circular sequence Y_i and to take its distribution conditionally on $Y=y$ as given approximately by the analogue of (49), namely

$$\text{pr}(W=w \mid Y=y) = \frac{M'_{w,y}\, e^{-\frac{1}{2}\beta w}}{\Sigma\, M'_{w,y}\, e^{-\frac{1}{2}\beta w}}, \tag{50}$$

where $M'_{w,y}$ is the number of different samples of y 1's and $n-y$ 0's having the total number of runs equal to w.

Now W is formally identical with Wald and Wolfowitz's run statistic for testing for the difference between two populations on the basis of a random sample drawn from each. In that application (Wald and Wolfowitz, 1942), with samples of size $n-y$ and y, the observations are ranked and then those from the first sample are scored 0 and those from the second sample scored 1. The total number of runs is used as a test statistic: the authors show that this test is consistent for all possible alternatives, although the procedure is clearly very inefficient for many particular types of alternative.

The null distribution is determined by the numbers $M'_{x,y}$ and the exact distribution has been given by Wald and Wolfowitz and earlier by Stevens (1939), who considered both the circular and non-circular cases.* Wald and Wolfowitz show also that the total number of runs W is asymptotically normal as $n \to \infty$ with y/n fixed, with

$$E(W) \sim \frac{2y(n-y)}{n}, \tag{51}$$

$$V(W) \sim \frac{4y^2(n-y)^2}{n^3}. \tag{52}$$

It follows from (5), using a relation for cumulant generating functions similar to (29), that if β is small, W is asymptotically normal with mean

$$\frac{2y(n-y)}{n} - \tfrac{1}{2}\beta V(W) \tag{53}$$

and variance (52), so that approximate estimation of β is possible by the methods used before.

A small sample test can be made from the tables of Swed and Eisenhart (1943) or alternatively by reduction to a 2×2 contingency table in the way very clearly set out by Stevens. This can be done exactly in the circular case. In the non-circular case, it is usually probably best to work not with the total number of runs but with say W_1, the number of runs of 1's, which may be equal to or be one more or one less than the number

* The problem had been considered before this as a combinatorial one.

of runs of 0's. The quantity W_1 may be tested by the usual exact test applied to the 2×2 contingency table

W_1	$y - W_1$
$n - y + 1 - W_1$	$W_1 - 1$

The more general problems of inference in Markov chains with more than two states have been considered by a number of authors (Bartlett, 1951; Hoel, 1954; Good, 1955; Anderson and Goodman, 1957).

7. Cumulative Score as an Independent Varibale

As was explained in §1, it is in some applications right to assume that the probability of a 1 on the i^{th} trial is a function of the number of 1's occurring in the first $(i - 1)$ trials, i.e. of $Y_1 + \ldots + Y_{i-1}$. A significance test for such dependence could be obtained by using the serial order test of §4.5, although these would presumably be some loss of power in doing this. For estimation purposes, however, it is necessary to treat the problem afresh.

Suppose that the probability that a trial has outcome 1, given that there have been v 1's in the previous trials, is

$$e^{\alpha+\beta v}/(1 + e^{\alpha+\beta v}). \tag{54}$$

Consider a sequence $Y_1, \ldots, Y_n$ in which there are y 1's and let r_0 be the serial number of the first 1, $r_0 + r_1$ the serial number of the second 1 and so on, $r_0 + \ldots + r_{y-1}$ being the serial number of the last 1, and r_y being defined as $n - r_0 - \ldots - r_{y-1}$. Then the likelihood is

$$\frac{e^{y\alpha+\frac{1}{2}y(y-1)\beta}}{\prod_{k=0}^{y} (1 + e^{\alpha+k\beta})^{r_k}}, \tag{55}$$

so that the full sequence $r_0, \ldots, r_y$ is required for sufficiency.

A more intuitive approach is to consider the sum of products analogous to those that have occurred in the previous sections, namely $\Sigma\, Y_i(Y_1 + \ldots + Y_{i-1})$. This, however, is equal to $\frac{1}{2}y(y-1)$, so that this approach leads nowhere. Another possibility is to note that if β is small, or large, the likelihood (55) involves the r's only through $\Sigma\, kr_k$ and $\Sigma\, r_k = n$. This suggests that the use of $R = \Sigma\, kr_k$ as a test statistic will be locally optimum, large values of R corresponding to negative values of β.

Under the null hypothesis $\beta = 0$ each ordering of the r_i has equal probability, with the exception that if one of the r_i is 0 it must occur at a fixed position, namely at the end. Let $y' = y + 1$ if all r_i are non-zero, i.e. if the series ends with a 0, and let $y' = y$ if $r_y = 0$. Thus y' is the number of non-zero r's, and the $y'!$ permutations of the r's are equally likely. If y' is not too small a satisfactory test should be obtained from the distribution of R in the universe of permutations of the non-zero r's observed. In this

$$E_p(R) = \tfrac{1}{2}(y' - 1)\, n, \tag{56}$$

$$V_p(R) = \frac{y'(y' + 1)}{12} \Sigma\, (r_i - \bar{r})^2, \tag{57}$$

where E_p, V_p refer to expectations over the permutations just mentioned and the sum of squares in (57) is over the non-zero r's. The last formula is most neatly proved by noting that for arbitrary $\{r_0, \ldots, r_y\}$ we have by symmetry and invariance considerations that

$$V_p(R) \equiv E_p\{\Sigma\, k(r_k - \bar{r})\}^2 = B\,\Sigma\,(r_i - \bar{r})^2, \tag{58}$$

where B depends only on y. The expectations of both sides are easily calculated when $r_0, \ldots, r_y$ are uncorrelated random variables of mean zero and unit variance, and B is then determined. Asymptotic normality follows from Hoeffding's theorem (Hoeffding, 1951).

Example 4.—In a learning experiment the following sequence is obtained in 20 trials, a correct response being denoted by 1 and an incorrect one by 0:

$$0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0, 1, 0, 1.$$

Here $r_0 = r_1 = 4$, $r_2 = r_5 = r_8 = r_9 = 2$, $r_3 = r_4 = r_6 = r_7 = 1$, $r_{10} = 0$, with $n = 20$, $y' = y = 10$, and $R = 72$. From (56) and (57), the theoretical mean on the null hypothesis is 90 and the variance

$$\frac{10 \times 11}{12}\left\{52 - \frac{20^2}{10}\right\} = 110.$$

Hence the standardized deviation from expectation, after correction for continuity, is $17{\cdot}5/10{\cdot}49 = 1{\cdot}67$, significant at about the $9\frac{1}{2}$ per cent. level in the normal tables.

If information from several series is to be combined, the values of R are added.

The statistic R does not seem to lend itself to a simple approximate estimation procedure. If an estimate of β is required, as would usually be the case, the simplest procedure is to consider the formal maximum likelihood or minimum χ^2 procedures from (55).

One industrial application in which this sort of situation might arise is in the study of failure rates for mechanisms under modification. At each trial the mechanism may not fail (0) or may fail (1). After a failure, but not after a non-failure, the mechanism is modified. The parameter $-\beta$ measures the rate of decrease, due to modification, in the probability of failure.

A more frequently occurring application, however, is probably the one suggested above to the analysis of learning experiments. There has been much interesting work recently on stochastic models for the representation of learning in simple situations (Bush and Mosteller, 1955; Cane, 1956; Audley, 1956, 1957; Audley and Jonckheere, 1956) and some of these studies have led to formidable statistical problems of fitting and testing. When these studies aim at linking the observations to a neuro-physiological mechanism, it is reasonable to take the best model practicable and to wrestle as vigorously as possible with the resulting statistical complications. If, however, the object is primarily the reduction of data to a manageable and revealing form, it seems fair to take for the probability of a success, 1, as simple an expression as possible that seems to the right general shape and which is flexible enough to represent the various possible dependencies that one wants to examine. For this the logistic seems a good thing to consider.

The simplest case is the one considered above where the probability of a correct response is assumed to depend only on the number of previous correct responses, or in a slightly more complicated case, only on the number of previous rewarded correct responses. The equation (54) may be compared with that of Audley (1956), which in the present notation is that the probability that a trial has outcome 1 is

$$(\alpha + \beta v)/(1 + \beta v). \tag{59}$$

A difficulty that might arise with some applications of (59), although not with the present one, is that β must be non-negative. The maximum likelihood solution for (59) has been used by Audley.

A fairly general model that should cover a variety of cases is to set up a logistic dependence between the probability of a 1 and such quantities as (i) the number of preceding correct responses (or rewarded responses), (ii) the number of (penalized) incorrect responses, and (iii) the result of the preceding trial. Much of the interest of such an analysis would presumably lie in finding whether, under a range of circumstances, the dependencies can be represented with a small number of such independent variables. Maximum likelihood fitting, using the tables of Berkson (1957), results in a straightforward iterative multiple regression calculation. The number of linear equations to be solved at each step is one more than the number of independent variables.

An alternative, very simple non-iterative, method of analysis is to use the method of minimum logit χ^2; grouping of the independent variables would be required to ensure that there are few combinations for which the responses are all 0 or all 1. This method is recommended for finding a first approximation for use in the iterative solution of the maximum likelihood equations.

Some of this work was done at the I.M.S. Summer Institute, 1957, and support from the National Science Foundation is gratefully acknowledged. I wish also to thank Dr. A. Birnbaum for some helpful discussions.

REFERENCES

ANDERSON, T. W. & GOODMAN, L. A. (1957), "Statistical inference about Markov chains", *Ann. Math. Statist.*, **28**, 89–110.

ARMITAGE, P. (1955), "Tests for linear trends in proportions and frequencies", *Biometrics*, **11**, 375–386.

AUDLEY, R. J. (1956), *Stochastic Processes and the Description of Learning Behaviour in Choice Situations*. Ph.D. thesis, Univ. of London.

—— (1957), "A stochastic description of the learning behaivour of an individual subject", *Q. J. Exptl. Psychol.*, **9**, 12–20.

—— & JONCKHEERE, A. R. (1956), "Stochastic processes and learning behaviour", *Brit. J. Statist. Psychol.*, **9**, 87–94.

BARTLETT, M. S. (1951), "The frequency goodness of fit for probability chains", *Proc. Camb. Phil. Soc.*, **47**, 86–95.

BENNETT, B. M. (1956), "On a rank order test for the equality of probability of an event", *Skand. Akt.* **20**, 11–18.

BERKSON, J. (1955), "Maximum likelihood and minimum χ^2 estimates of the logistic function", *J. Amer. Statist. Assoc.*, **50**, 130–162.

—— (1957), "Tables for the maximum likelihood estimate of the logistic function", *Biometrics*, **13**, 28–34.

BIRNBAUM, Z. W. (1956), "On the use of the Mann-Whitney statistic", *Proc. 3rd. Berkeley Symp. on Math. Statist. and Prob.*., **1**, 13–17.

BUSH, R. R. & MOSTELLER, F. (1955), *Stochastic Models for Learning*. New York: J. Wiley.

CANE, V. R. (1956), "Some statistical problems in experimental psychology", *J. R. Statist. Soc.* **B**, **18**, 177–201.

COCHRAN, W. G. (1955), "Some methods for strengthening the common χ^2 tests", *Biometrics*, **10**, 417–451.
CORNFIELD, J. (1956), "Statistical problem arising from retrospective studies", *Proc. 3rd. Berkeley Symp. on Math. Statist. and Prob.*, **3**, 135–148.
COX, D. R. (1955), "Some statistical methods connected with series of events", *J. R. Statist. Soc.*, B, **17**, 129–164.
FISHER, R. A. (1935), "The logic of inductive inference", *J. R. Statist. Soc.*, **98**, 39–54.
GARWOOD, F. (1941), "The application of maximum likelihood to dosage-mortality curves", *Biometrika*, **32**, 46–58.
GOOD, I. J. (1955), "The likelihood ratio test for Markoff chains", *Biometrika*, **42**, 531–533.
HALDANE, J. B. S. (1940), "The mean and variance of χ^2 when used as a test of homogeneity when the expectations are small", *Biometrika*, **31**, 346–355.
—— & SMITH, C. A. B. (1948), "A simple exact test for birth-order effect", *Ann. Eugenics*, **14**, 117–124.
HOEFFDING, W. (1951), "A combinatorial central limit theorem", *Ann. Math. Statist.*, **22**, 558–566.
HOEL, P. G. (1954), "A test for Markoff chains", *Biometrika*, **41**, 430–433.
IRWIN, J. O. & KENDALL, M. G. (1944), "Sampling moments of moments for a finite population", *Ann. Eugenics*, **12**, 138–142.
MADOW, W. G. (1948), "On the limiting distributions of estimates based on samples from finite universes", *Ann. Math. Statist.*, **19**, 535–545.
MOTOO, M. (1957), "On Hoeffding's combinatorial central limit theorem", *Ann. Inst. Statist. Math.*, **8**, 145–154.
PATNAIK, P. B. (1948), "The power function of the test for the difference between two proportions in a 2×2 table", *Biometrika*, **35**, 157–173.
PEARSON, E. S. & MERRINGTON, M. (1951), "Tables of the 5% and 0·5% points of Pearson curves expressed in standard measure", *Biometrika*, **38**, 4–10.
SIEGEL, S. (1957), *Non-parametric Methods of Statistics*. New York: McGraw-Hill.
SILVERSTONE, H. (1957), "Estimating the logistic curve", *J. Amer. Statist. Assoc.*, **52**, 567–577.
STEVENS, W. L. (1939), "Distribution of groups in a sequence of alternatives", *Ann. Eugenics*, **9**, 10–17.
—— (1951), "Mean and variance of an entry in a contingency table", *Biometrika*, **38**, 468–470.
SWED, F. S. & EISENHART, C. (1943), "Tables for testing randomness of grouping in a sequence of alternatives", *Ann. Math. Statist.*, **14**, 66–87.
WALD, A. & WOLFOWITZ, J. (1940), "On a test of whether two samples are from the same population", *Ann. Math. Statist.*, **11**, 147–162. Reprinted in Selected Papers in Statistics and Probability, A. Wald, McGraw Hill.
YATES, F. (1948), "The analysis of contingency tables with groupings based on quantitative characters", *Biometrika*, **35**, 176–181.

DISCUSSION ON DR. COX'S PAPER

Dr. C. A. B. SMITH: It is a great pleasure to propose the vote of thanks to a paper so well set out. Dr. Cox points out clearly the value of the logistic function in simplifying the mathematical theory, especially in such places as equations (29) and (40) which express the general cumulant generating function in terms of that for the null hypothesis. I am also very much interested to see that certain non-parametric significance tests (including the Haldane-Smith rediscovery of Wilcoxon's test) are optimal against alternatives involving a linear regression of the logit. Opinions about the merits of non-parametric tests may vary, but they give me a feeling of honest confidence which is the proper conclusion of a well-conducted scientific investigation—provided, that is, that they are reasonably near optimal.

At the moment, however, I am concerned with some problems similar to those discussed by Cox but involving estimation rather than significance: and an important question there is how far they can be reduced to a linear regression of the logit. Of course, in an assay problem a small dose of a given drug will in most cases give no response, a large enough dose will give 100 per cent. response, and the curve relating probability of response to dose will have an "ogive" shape. It is then hardly surprising that by a suitable transformation of the dosage it can be coaxed into a good approximation to the logistic or cumulative Gaussian form. However, in other cases this may not be so. I have been looking at some results of Dr. D. B. Fry, of the University College Phonetics Department. (He has kindly given me permission to refer to these as yet unpublished results.) He was investigating what constitutes "stress" in phonetics, i.e., what distinguishes, say, the CONtent of a book from conTENT of mind. In ordinary conversation the words may be different in force, sound, tone and timing. Dr. Fry has tried to examine the effect of