REFERENCES

ERDELYI, A. *et al.* (1953). *Higher Transcendental Functions*, **2**. New York: McGraw-Hill Book Co. Inc.
KENDALL, M. G. (1957). The moments of the Leipnik distribution. *Biometrika*, **44**, 270.
LEIPNIK, R. B. (1947). Distribution of the serial correlation coefficient in a circularly correlated universe. *Ann. Math. Statist.* **18**, 86.
QUENOUILLE, M. H. (1948). Some results in the testing of serial correlation coefficients. *Biometrika*, **35**, 261.
WATSON, G. N. (1944). *Theory of Bessel Functions*. Cambridge University Press.
WHITE, J. H. (1957). Approximate moments for the serial correlation coefficient. *Ann. Math. Statist.* **28**, 798.

Two further applications of a model for binary regression

BY D. R. COX
Birkbeck College, University of London

1. *Introduction.* In a recent paper (Cox, 1958), I have discussed some aspects of a logistic model for analysing regression when the dependent variable can take only two values, say 0 and 1. In the present note two further applications are presented of what is essentially the same model. The first is to the analysis of 2×2 contingency tables based on matched pairs, and the second is to the testing of the agreement between an observed binary sequence and a corresponding sequence of probabilities.

2. *The 2×2 contingency table with matched pairs.* Consider the form taken by a simple comparison of matched pairs when the observations are (0, 1) variables. Let there be n pairs of individuals, the pairing usually being such that the two individuals in any one pair tend to be alike. Let one member of each pair belong to group A, the other to group B, the assignment being randomized if a comparative experiment is involved. An observation, taking one of two values 0 and 1, is made on each individual. For the ith pair, let these be represented by random variables Y_{ia}, Y_{ib}. The possible observations on a pair, writing that on A first, are (0, 0), (0, 1), (1, 0) and (1, 1).

It is possible to form a 2×2 contingency table from the data

	Group A	Group B
0		
1		
	n	n

McNemar (1947) seems to have been the first to point out that the usual χ^2 significance test for such a table is invalid, because it ignores the correlation induced by pairing. He recommended that the significance of the difference between A and B should be tested by rejecting the pairs (0, 0) and (1, 1), and by examining whether the proportion of (1, 0)'s among the remaining 'mixed' observations (0, 1) and (1, 0) is consistent with binomial variation with chance $\frac{1}{2}$. Mosteller (1952) and Cochran (1950) have given further accounts of this test and Cochran has discussed extensions to the comparison of more than two groups. Stuart (1957) has recently obtained a test equivalent to McNemar's by arguments based on the theory of stratified sampling.

This work raises two problems. Are there circumstances under which the test is optimum, and is there a corresponding estimation procedure? To deal with these questions we must set up a parametric model covering the non-null case. The simplest such model seems to be the following. Let all random variables be mutually independent and let there be a parameter λ_i characteristic of the ith pair and a parameter ψ describing the true difference between A and B, such that

$$\Pr(Y_{ia} = 1)/\Pr(Y_{ia} = 0) = \lambda_i, \tag{1}$$

$$\Pr(Y_{ib} = 1)/\Pr(Y_{ib} = 0) = \psi\lambda_i. \tag{2}$$

If we write $\lambda_i = e^{\alpha_i}$, $\psi = e^{\beta}$, we have the logistic model of the earlier paper.

It follows by the arguments of that paper, in particular of § 4·5, that the jointly sufficient set of statistics consists of (i) $\sum_{i=1}^{n} Y_{ib}$, (ii) the pair totals $(Y_{ia} + Y_{ib})$, $i = 1, \ldots, n$. Further, optimum inference about ψ, regarding $\lambda_1, \ldots, \lambda_n$ as unknown nuisance parameters, is based on the distribution of (i) condi-

tionally on the set (ii). Now whenever $Y_{ia}+Y_{ib}\neq 1$, the contribution of the ith pair to (i) is fixed. Hence, the conditional distribution just mentioned is equivalent to that of R = number of pairs (0, 1) conditionally on the observed value of M = number of pairs (0, 1) or (1, 0).

Now a simple calculation from (1) and (2) shows that

$$\Pr(Y_{ia}=0, Y_{ib}=1 \mid Y_{ia}+Y_{ib}=1)=\psi/(1+\psi)=\theta, \text{ say.} \tag{3}$$

Therefore R, conditionally on the observed value of M, has a binomial distribution

$$\Pr(R=r \mid M=m)=\binom{m}{r}\theta^r(1-\theta)^{m-r}. \tag{4}$$

In particular the optimum test of the null hypothesis $\psi=1$, $\theta=\frac{1}{2}$ is McNemar's test, and confidence intervals for θ and hence for ψ are obtained in the usual way for a binomial parameter. The significance test can be looked on as the very special case of Haldane & Smith's (1948) test for a serial order effect obtained when each series contains just two items.

Example. Mosteller (1952) illustrated the test on an experiment in which each of 100 subjects used both of two drugs A and B, the response being a dichotomy 'not-nausea', 'nausea' (0 and 1, say). 81 subjects never had nausea, i.e. gave the observation (0, 0), 9 subjects gave (1, 0), i.e. had nausea with A but not with B, 1 subject gave (0, 1) and 9 gave (1, 1). The significance test of the null hypothesis that the drugs are equally liable to induce nausea amounts to testing whether a division of 10 trials into (9, 1) is significantly extreme in a binomial distribution with chance $\frac{1}{2}$. The exact significance level in a two-sided test is $11/512\simeq 0{\cdot}021$; as an approximation to this, we get from a χ^2 test, corrected for continuity, that significance is attained at very nearly the 0·025 level. A table of 95 % confidence limits for the binomial probability (Hald, 1952) gives (0·003, 0·445) as the limits for θ and hence the odds factor ψ is between 1/300 and 4/5.

Tests and interval estimates comparing the values of ψ in different experiments can be done by familiar techniques for binomial variates.

3. *Test of agreement between a sequence and a set of probabilities.* Let $Y_1,\ldots, Y_n$ be mutually independent random variables each taking the values (0, 1) and let $p_1,\ldots, p_n$ be a given set of numbers, $0\leqslant p_i\leqslant 1$. Suppose that it is required to use observations on $Y_1,\ldots, Y_n$ to test the hypothesis that

$$\Pr(Y_i=1)=p_i, \quad (i=1,\ldots,n). \tag{5}$$

For example, a weather forecaster might put forward each day a number purporting to be the probability that it will rain the following day. It might then be required to test whether the observed occurrences of rain are consistent with these probabilities.

If n is large, we may group the trials into sets each with nearly constant p_i; then the observed proportion of 1's in each set can be compared with the corresponding p_i. Let n be too small for this test to be used.

One method of deriving a small sample test, when special alternatives to (5) are not available, is to consider a family of probabilities derived from (5). This family is characterized by a continuous parameter β and

$$\log\{\Pr_\beta(Y_i=1)/\Pr_\beta(Y_i=0)\}=\beta\log\{p_i/(1-p_i)\}. \tag{6}$$

The null hypothesis (5) corresponds to $\beta=1$. If $\beta>1$, the suggested probabilities p_i show the right general pattern of variation, but do not vary enough. If $0<\beta<1$, the suggested probabilities vary too much. If $\beta<0$, the p_i vary in the wrong direction and if $\beta=-1$, the p_i are the complements of the true probabilities.

The log likelihood under (6) of an observed series $y_1,\ldots, y_n$ is

$$\beta\Sigma y_i\log p_i+\beta\Sigma(1-y_i)\log(1-p_i)-\Sigma\log\{p_i^\beta+(1-p_i)^\beta\}. \tag{7}$$

Hence, the sufficient statistic is obtained by scoring

$$X_i=\begin{cases}\log(2p_i) & \text{when } Y_i=1;\\ \log[2(1-p_i)] & \text{when } Y_i=0,\end{cases} \tag{8}$$

and by considering a total score $X=\Sigma X_i$. The factor 2 is included to make the expected score positive and to arrange that an event of probability $\frac{1}{2}$ scores 0.

Under the null hypothesis $\beta=1$,

$$E_1(X)=n\log 2+\Sigma p_i\log p_i+\Sigma(1-p_i)\log(1-p_i), \tag{9}$$

$$V_1(X)=\Sigma p_i(1-p_i)\{\log[p_i/(1-p_i)]\}^2. \tag{10}$$

Provided that n is not very small and that none of the p_i is near 0 or 1, the distribution of X is nearly normal.

In principle it would be possible to calculate confidence intervals for β from an observed value $X = x$. If x significantly exceeds (9), this is evidence that $\beta > 1$.

Example. Suppose that there are 16 trials, 8 of which have outcome 1 and 8 have outcome 0. Let the p_i corresponding to the zero observations be 0·1, 0·1, 0·2, 0·2, 0·4, 0·5, 0·6, 0·7, and corresponding to the unit observations 0·3, 0·3, 0·5, 0·6, 0·6, 0·8, 0·9, 0·9.

Thus the score for the first observation recorded as 0 is $\log[2(1-p_i)] = \log 1{\cdot}8 = 0{\cdot}255$, and the score for the first observation recorded as 1 is $\log(2p_i) = \log 0{\cdot}6 = -0{\cdot}222$. We find that the total observed score $x = 1{\cdot}106$ and that under the hypothesis $\beta = 1$, equations (9) and (10) give

$$E_1(X) = 1{\cdot}030, \quad V_1(X) = 0{\cdot}785,$$

so that there is excellent agreement with expectation. Under the hypothesis $\beta = 0$, i.e. that 1's occur randomly with constant chance $\frac{1}{2}$, we find

$$E_0(X) = n\log 2 + \tfrac{1}{2}\Sigma\log[p_i(1-p_i)] = -1{\cdot}329,$$
$$V_0(X) = \tfrac{1}{4}\Sigma\{\log[p_i/(1-p_i)]\}^2 = 1{\cdot}314.$$

The observed value differs significantly from $E_0(X)$ at the 5 % level. Thus the data support the idea that 1's do not occur with constant chance $\frac{1}{2}$ and are in excellent agreement with the suggested probabilities.

The family (6), on which the test just described is based, is especially appropriate when the sequence $\{p_i\}$ is known to be correct at and near $p = \frac{1}{2}$ but possibly incorrectly spread around $p = \frac{1}{2}$. Thus we may call the test based on (9) and (10) a test for spread. A natural generalization is to replace (6) by

$$\log\{\Pr_{\beta,\alpha}(Y_i = 1)/\Pr_{\beta,\alpha}(Y_i = 0)\} = \beta\log\{p_i/(1-p_i)\} + \alpha, \tag{11}$$

the null hypothesis being that $\beta = 1$, $\alpha = 0$. The pair of sufficient statistics are X, as defined previously, and $Y = \Sigma Y_i$. Under the null hypothesis, X, Y are nearly jointly normally distributed with the mean and variance of X given by (9) and (10) and with

$$E_1(Y) = \Sigma p_i, \quad V_1(Y) = \Sigma p_i(1-p_i), \tag{12}$$

$$C_1(X, Y) = \Sigma p_i(1-p_i)\log[p_i/(1-p_i)]. \tag{13}$$

Note that if the p_i are symmetrically arranged about $\frac{1}{2}$, X and Y are uncorrelated.

A test for bias ignoring spread will be based on Y alone, i.e. solely on the observed total number of 1's. If both bias and spread are of interest, it is necessary to specify the relative importance to be attached to each, if an optimum small-sample procedure is to be found. Since it is rarely possible to do this, a sensible practical approach is to find the observed values x and y and to see whether

$$(x-E_1(X), y-E_1(Y))\begin{pmatrix} V_1(X) & C_1(X, Y) \\ C_1(X, Y) & V_1(Y) \end{pmatrix}^{-1}\begin{pmatrix} x-E_1(X) \\ y-E_1(Y) \end{pmatrix} \tag{14}$$

is significantly large in the χ^2 distribution with 2 degrees of freedom. The expression (14) is, except for a factor $\frac{1}{2}$, the exponent in the bivariate normal distribution of X and Y; it is the likelihood ratio statistic for testing the hypothesis that X, Y have the bivariate normal distribution (9), (10), (12) and (13), against the hypothesis that X, Y have arbitrary means, but the same covariance matrix as under the null hypothesis. This, of course, does not allow for the fact that the covariance matrix varies in a determined way with the parameters α and β. However, the determination of the correct likelihood ratio criterion requires the maximum likelihood estimation of α and β, which is tedious.

Example. Consider the data that were analysed previously. We have that the observed value of Y is $y = 8$ and that $E_1(Y) = 7{\cdot}7$, $V_1(Y) = 2{\cdot}930$, $C_1(X, Y) = -0{\cdot}090$. Therefore, the observed value of Y, as well as that of X, agrees well with its expectation under the suggested scheme of probabilities and the need for a combined test hardly arises. The formal details of such a test are that

$$(1{\cdot}106-1{\cdot}030,\ 8-7{\cdot}7)\begin{pmatrix} 0{\cdot}785 & -0{\cdot}090 \\ -0{\cdot}090 & 2{\cdot}930 \end{pmatrix}^{-1}\begin{pmatrix} 1{\cdot}106-1{\cdot}030 \\ 8-7{\cdot}7 \end{pmatrix} \tag{15}$$

is to be tested as χ^2 with 2 degrees of freedom. The value of expression (15) is 0·01: a value smaller than this would arise by chance only about 1 in 100 times.

There are further problems connected with the general situation discussed here. First, the same set of observations can be consistent with several alternative sequences of probabilities and it may be

required to consider which sequence is preferable. It seems reasonable to prefer that sequence of probabilities for which the information in Shannon's sense is a minimum, for this implies minimum uncertainty concerning the outcome of the realized sequence. According to (9), this amounts to preferring the probabilities for which $E_1(X)$ is a minimum. Secondly, it happens in some applications that the probabilities p_i are not given, but have to be estimated from data by fitting a particular type of model, often to the same data with which goodness of fit is to be tested. In such cases, the most satisfactory test of goodness of fit is likely to be obtained by fitting a model containing additional parameters and testing estimates of the additional parameters for significance from zero. The approach of the present section is relevant only when there are available no special forms of alternative specific to the problem.

REFERENCES

COCHRAN, W. G. (1950). The comparison of percentages in matched samples. *Biometrika*, **37**, 256–66.

COX, D. R. (1958). The regression analysis of binary sequences. *J.R. Statist. Soc.* B, **20**, to appear.

HALD, A. (1952). *Statistical Tables and Formulas*. New York: Wiley and Sons.

HALDANE, J. B. S. & SMITH, C. A. B. (1948). A simple exact test for birth-order effect. *Ann. Eugen., Lond.*, **14**, 117–24.

MCNEMAR, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, **12**, 153–7.

MOSTELLER, F. (1952). Some statistical problems in measuring the subjective response to drugs. *Biometrics*, **8**, 220–6.

STUART, A. (1957). Comparison of frequencies in matched samples. *Brit. J. Statist. Psychol.* **10**, 29–32.

A note on a series solution of a problem in estimation*

BY IRWIN GUTTMAN

University of Alberta and Princeton University

1. INTRODUCTION AND SUMMARY

If $t(x)$ is a sufficient statistic for the family of probability functions $\{\mathscr{P}_\theta^x \mid \theta \epsilon \Omega\}$ defined over the real line, and if $f(x)$ is an unbiased estimator of a real valued function of the parameter, say $g(\theta)$, then it is well known that the function

$$h(t) = E\{f(X) \mid t\}$$

is an unbiased estimator of $g(\theta)$, and that it has smaller variance and risk (for strictly convex loss functions) than $f(x)$, unless of course $f(x) = h(t(x))$ almost everywhere $\{\mathscr{P}_\theta^x\}$. Further, if t is also a complete statistic, then $h(t)$ is the unique Uniformly Minimum Variance (UMV) unbiased estimate of $g(\theta)$.

The above holds for continuous and discrete probability functions $\mathscr{P}_\theta^x$. We discuss here the case where $\mathscr{P}_\theta^x$ are discrete probability distribution functions defined on the real line, with probability densities $p(x)$, where $x = 0, 1, 2, \ldots$.

Under certain regularity conditions given in § 2, a method of determining $h(t)$, without considering unbiased estimators $f(x)$ of $g(\theta)$ at all, is given. This has the feature, then, of avoiding the evaluation of conditional expectations. The method also allows for a solution of a problem raised by Girshick, Mosteller & Savage (1946). This is discussed in § 3, where some examples are given to illustrate the theorem of § 2. It is interesting to note that a special case of the method has been used by Lehmann & Scheffé (1950) to prove completeness of some statistics.

* Prepared in connexion with research sponsored by the Office of Naval Research.