Further Results on Tests of Separate Families of Hypotheses

By D. R. Cox
Birkbeck College, University of London

[Received May 1962]

Summary

It is required to test a composite null hypothesis. High power is desired against a composite alternative hypothesis that is not in the same parametric family as the null hypothesis. In an earlier paper a modification of the Neyman–Pearson maximum-likelihood ratio test was suggested for this problem. The present paper gives some general comments on the formulation of the problem, a general large-sample form for the test, and, finally, a number of examples.

1. Introduction

Suppose that the observed value of a random vector $\mathbf{Y} = (Y_1, \ldots, Y_n)$ is to be used to test the null hypothesis, H_f, that the probability density function (p.d.f.) is $f(\mathbf{y}, \boldsymbol{\alpha})$, where $\boldsymbol{\alpha}$ is an unknown vector parameter. Let it be required to have high power for the alternative hypothesis, H_g, that the p.d.f. is $g(\mathbf{y}, \boldsymbol{\beta})$, where $\boldsymbol{\beta}$ is an unknown vector parameter, and where $f(\mathbf{y}, \boldsymbol{\alpha})$ and $g(\mathbf{y}, \boldsymbol{\beta})$ are separate families. That is, for an arbitrary parameter value $\boldsymbol{\alpha}_0$, the p.d.f. $f(\mathbf{y}, \boldsymbol{\alpha}_0)$ cannot be approximated arbitrarily closely by $g(\mathbf{y}, \boldsymbol{\beta})$.

Cox (1961) gave examples of such problems and developed a general large-sample procedure based on a modification of the Neyman–Pearson maximum-likelihood ratio. This general procedure was shown to reduce in a rather special class of cases to a fairly simple form involving only maximum-likelihood estimators. One such example is when $Y_1, \ldots, Y_n$ are independently and identically distributed and H_f is the hypothesis that the distribution is log-normal and H_g is the hypothesis that the distribution is exponential.

The objects of the present paper are threefold:

(i) to amplify some general comments on the problem made in the previous paper;

(ii) to give a general form for the test requiring calculations rather similar to those in ordinary maximum-likelihood estimation;

(iii) to work out some examples.

2. Some General Remarks

The notation and results of the previous paper are used. Thus $L_f(\hat{\boldsymbol{\alpha}})$ is the maximized log likelihood under H_f, $L_g(\hat{\boldsymbol{\beta}})$ that under H_g, $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ denoting maximum-likelihood estimators. The test is based on

$$T_f = \{L_f(\hat{\boldsymbol{\alpha}}) - L_g(\hat{\boldsymbol{\beta}})\} - E_{\hat{\boldsymbol{\alpha}}}\{L_f(\hat{\boldsymbol{\alpha}}) - L_g(\hat{\boldsymbol{\beta}})\}, \tag{1}$$

comparing the observed difference of log likelihoods with an estimate of that to be expected under H_f. When H_f is true, $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ is asymptotically normally distributed with mean $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and a covariance matrix that can be calculated from expected derivatives of $\log f(\mathbf{y}, \boldsymbol{\alpha})$ and $\log g(\mathbf{y}, \boldsymbol{\beta}_\alpha)$. The quantity $\boldsymbol{\beta}_\alpha$, which is important in what follows, is the limit to which $\hat{\boldsymbol{\beta}}$ converges when H_f holds with parameter value $\boldsymbol{\alpha}$. Under H_f, T_f should be approximately zero, whereas under H_g, T_f should be negative.

The following are some general comments on the formulation and solution.

(i) The problem is considered here as one of significance testing rather than of discrimination. That is, the two hypotheses H_f and H_g are considered unsymmetrically, the hypothesis H_g serving only to indicate the type of alternative for which high power is required. It is not assumed that H_f and H_g are the only possible hypotheses. In a pure discrimination problem the quantity $L_f(\hat{\boldsymbol{\alpha}})-L_g(\hat{\boldsymbol{\beta}})$ would often be of direct interest.

(ii) An "exact" test can sometimes be constructed in principle as follows. The conditional distribution of the observations is taken, given the observed value of the minimal sufficient statistic for $\boldsymbol{\alpha}$. In this conditional distribution, the Neyman–Pearson lemma is applied to find, if available, a uniformly most powerful test. A large-sample test based on (1) is an approximate version of the exact argument in that we take the deviation of the log maximum-likelihood ratio $L_f(\hat{\boldsymbol{\alpha}})-L_g(\hat{\boldsymbol{\beta}})$ from its regression on the asymptotically sufficient statistic $\hat{\boldsymbol{\alpha}}$. This remark helps to explain the distributional results of sections 3 and 4.

(iii) Care is needed in thinking about large-sample approximations to the distribution of T_f. We are dealing with situations in which both probabilities of error can be made arbitrarily small by making the sample size large enough. We assume that in applications the sample size is large enough to ensure that the usual expansions of maximum-likelihood theory are good approximations, but not so large that H_f and H_g can be distinguished with negligible probability of error. (If the probabilities of error were negligible, there would hardly be need for a formal statistical test!) If a rigorous statement of limiting theorems were to be attempted, a passage to the limit like that considered by Chernoff (1956, section 6) would be necessary.

(iv) An alternative important method of tackling these problems is to introduce a comprehensive model including H_f and H_g as special cases. For example, the p.d.f. can be taken proportional to

$$\{f(\mathbf{y}, \boldsymbol{\alpha})\}^{\lambda}\{g(\mathbf{y}, \boldsymbol{\beta})\}^{1-\lambda}$$

and inferences about λ can be made in principle in the usual way. This procedure has the major advantage of leading to an estimation procedure as well as to a significance test. Usually, however, the calculations will be very complicated.

(v) If the roles of H_f and H_g as null and alternative hypotheses are interchanged, a test statistic T_g is obtained, where

$$T_g = \{L_g(\hat{\boldsymbol{\beta}})-L_f(\hat{\boldsymbol{\alpha}})\}-E_{\hat{\boldsymbol{\beta}}}\{L_g(\hat{\boldsymbol{\beta}})-L_f(\hat{\boldsymbol{\alpha}})\}. \tag{2}$$

In general, T_f and T_g will be different functions of the observations. For each test there are qualitatively three answers; thus, from T_f, we may find consistency with H_f, evidence of departure in the direction of H_g, and evidence of departure from H_f away from H_g. If both tests are applied, as will often be desirable, there are nine qualitatively different conclusions. The interpretation of all these is straightforward, except for the one in which T_f indicates a departure from H_f away from H_g, and T_g indicates a departure from H_g away from H_f. Presumably this can only happen when there is a substantial departure from both H_f and H_g. In addition to the answers to the tests T_f and T_g, it will often be useful to look at the numerical value of log likelihood ratio, $L_g(\hat{\boldsymbol{\beta}})-L_f(\hat{\boldsymbol{\alpha}})$.

3. Preliminaries on the Distribution of T_f

We now consider the distribution in large samples of the test statistic T_f, defined by (1), when H_f is true. Discussion of regularity conditions is not attempted.

For ease of exposition, suppose first that the components $Y_1, \ldots, Y_n$ are independently identically distributed with p.d.f.'s $f(y, \alpha)$ and $g(y, \beta)$, where α and β are scalar parameters. Also, consider first the distribution of part of T_f, namely $L_f(\hat{\alpha}) - E_{\hat{\alpha}}\{L_f(\hat{\alpha})\}$. All distributional calculations are made under H_f.

Since $L_f'(\hat{\alpha}) = 0$, we have the usual expansions that approximately

$$L_f(\alpha) = L_f(\hat{\alpha}) + \tfrac{1}{2}(\hat{\alpha} - \alpha)^2 L_f''(\hat{\alpha}) \tag{3}$$

and

$$\hat{\alpha} - \alpha = \frac{L_f'(\alpha)}{L_f''(\hat{\alpha})}. \tag{4}$$

We replace $L_f''(\hat{\alpha})$ by $-nV\{\partial \log f(Y, \alpha)/\partial\alpha\}$. Thus

$$L_f(\hat{\alpha}) = L_f(\alpha) + \tfrac{1}{2}n(\hat{\alpha} - \alpha)^2 V\{\partial \log f(Y, \alpha)/\partial\alpha\} \tag{5}$$

and

$$\begin{aligned} E_\alpha\{L_f(\hat{\alpha})\} &= E_\alpha\{L_f(\alpha)\} + \tfrac{1}{2} \\ &= n\int f(y, \alpha) \log f(y, \alpha)\, dy + \tfrac{1}{2}. \end{aligned} \tag{6}$$

Hence

$$\begin{aligned} E_{\hat{\alpha}}\{L_f(\hat{\alpha})\} = n\int f(y, \alpha) \log f(y, \alpha)\, dy \\ + n(\hat{\alpha} - \alpha)\int \frac{\partial}{\partial\alpha}\{f(y, \alpha) \log f(y, \alpha)\}\, dy \\ + \tfrac{1}{2}n(\hat{\alpha} - \alpha)^2 \int \frac{\partial^2}{\partial\alpha^2}\{f(y, \alpha) \log f(y, \alpha)\}\, dy + \tfrac{1}{2}. \end{aligned} \tag{7}$$

It is useful to write

$$F_i = \log f(Y_i, \alpha), \quad F_{i,\alpha} = \frac{\partial \log f(Y_i, \alpha)}{\partial\alpha}, \quad F_{i,\alpha\alpha} = \frac{\partial^2 \log f(Y_i, \alpha)}{\partial\alpha^2}, \tag{8}$$

dropping the suffix i whenever Y_i is replaced by Y.

We then have results such as that

$$\int \frac{\partial}{\partial\alpha}\{f(y, \alpha) \log f(y, \alpha)\}\, dy = E_\alpha(FF_\alpha) = C_\alpha(F, F_\alpha),$$

$$\int \frac{\partial^2}{\partial\alpha^2}\{f(y, \alpha) \log f(y, \alpha)\}\, dy = E_\alpha\{F_\alpha^2 + F(F_\alpha^2 + F_{\alpha\alpha})\}.$$

It follows from (5) and (7) that

$$\begin{aligned} L_f(\hat{\alpha}) - E_{\hat{\alpha}}\{L_f(\hat{\alpha})\} = \Sigma\left\{F_i - E_\alpha(F) - \frac{C_\alpha(F, F_\alpha)}{V_\alpha(F_\alpha)} F_{i,\alpha}\right\} \\ - \tfrac{1}{2}n(\hat{\alpha} - \alpha)^2 E_\alpha\{F(F_\alpha^2 + F_{\alpha\alpha})\} - \tfrac{1}{2}. \end{aligned} \tag{9}$$

The first line of (9) is in general of order $\sqrt{n}$ in probability, whereas the second line is of order 1. For the moment ignore the second line. Now $C_\alpha(F, F_\alpha)/V_\alpha(F_\alpha)$ is the regression coefficient of F on F_α, under the hypothesis H_f. Hence we have the sum of deviations of the F_i from their linear regression on $F_{i,\alpha}$. Therefore the asymptotic distribution is normal with mean zero and variance

$$n\left\{V_\alpha(F)-\frac{C_\alpha^2(F,F_\alpha)}{V_\alpha(F_\alpha)}\right\}, \tag{10}$$

provided that (10) is not zero, in which case (9) is of order 1 in probability.

We shall in the remainder of the paper deal only with the $O(\sqrt{n})$ term. It would, however, certainly be possible to obtain improved approximations by incorporating the $O(1)$ term. Alternatively, a correction for bias could be incorporated by taking $L_f(\hat{\alpha})-E_{\hat{\alpha}}\{L_f(\hat{\alpha})\}$ to have mean not zero but

$$-\frac{E_\alpha\{F(F_\alpha^2+F_{\alpha\alpha})\}}{2E_\alpha(F_\alpha^2)}-\tfrac{1}{2}, \tag{11}$$

the expectation of the right-hand side of (9).

An example may help the understanding of these formulae. Let $f(y,\alpha)$ be the p.d.f. of a normal distribution of mean α and unit variance. Then

$$L_f(\hat{\alpha}) = -\tfrac{1}{2}n\log(2\pi)-\tfrac{1}{2}\Sigma(Y_i-\hat{\alpha})^2,$$

where $\hat{\alpha}=\Sigma Y_i/n$. Thus

$$E_\alpha\{L_f(\hat{\alpha})\} = -\tfrac{1}{2}n\log(2\pi)-\tfrac{1}{2}(n-1),$$

so that

$$L_f(\hat{\alpha})-E_{\hat{\alpha}}\{L_f(\hat{\alpha})\} = -\tfrac{1}{2}\{\Sigma(Y_i-\hat{\alpha})^2-(n-1)\}.$$

This has mean zero, variance $\tfrac{1}{2}(n-1)$ and has a translated chi-squared distribution. To apply the asymptotic theory, we calculate

$$F=-\tfrac{1}{2}\log(2\pi)-\tfrac{1}{2}(Y-\alpha)^2, \quad F_\alpha = Y-\alpha.$$

Thus $V_\alpha(F)=\tfrac{1}{2}$, $V_\alpha(F_\alpha)=1$, $C_\alpha(F,F_\alpha)=0$, so that the asymptotic distribution is, from (10), normal with mean zero and variance $\tfrac{1}{2}n$.

4. The Distribution of T_f

We now consider the distribution under H_f of T_f itself, considering only the terms of order $\sqrt{n}$ in probability. The extension of (3) is

$$L_f(\hat{\alpha})-L_g(\hat{\beta}) = L_f(\alpha)-L_g(\beta_\alpha). \tag{12}$$

Equations (6) and (7) are replaced by

$$E_\alpha\{L_f(\hat{\alpha})-L_g(\hat{\beta})\} = n\int f(y,\alpha)\log\left\{\frac{f(y,\alpha)}{g(y,\beta_\alpha)}\right\}dy$$

and

$$E_{\hat{\alpha}}\{L_f(\hat{\alpha})-L_g(\hat{\beta})\} = n\int f(y,\alpha)\log\left\{\frac{f(y,\alpha)}{g(y,\beta_\alpha)}\right\}dy$$
$$+n(\hat{\alpha}-\alpha)\int\frac{\partial}{\partial\alpha}\left[f(y,\alpha)\log\left\{\frac{f(y,\alpha)}{g(y,\beta_\alpha)}\right\}\right]dy. \tag{13}$$

Let
$$G_i = \log g(Y_i, \beta_\alpha), \quad G_{i,\beta} = \partial \log g(Y_i, \beta_\alpha)/\partial\beta, \tag{14}$$
the suffix i being omitted when Y_i is replaced by Y. Now
$$\frac{\partial}{\partial\alpha}\left[f(y,\alpha)\log\left\{\frac{f(y,\alpha)}{g(y,\beta_\alpha)}\right\}\right] = \frac{\partial f(y,\alpha)}{\partial\alpha}\log\left\{\frac{f(y,\alpha)}{g(y,\beta_\alpha)}\right\} + \frac{\partial f(y,\alpha)}{\partial\alpha} - \frac{f(y,\alpha)}{g(y,\beta_\alpha)}\frac{\partial g(y,\beta_\alpha)}{\partial\alpha}\frac{d\beta_\alpha}{d\alpha}.$$

Therefore
$$E_{\hat\alpha}\{L_f(\hat\alpha) - L_g(\hat\beta)\} = nE_\alpha(F-G) + n(\hat\alpha-\alpha)E_\alpha\{F_\alpha(F-G)\} - n(\hat\alpha-\alpha)\frac{d\beta_\alpha}{d\alpha}E_\alpha(G_\beta). \tag{15}$$

But $E_\alpha(G_\beta) = 0$ (Cox, 1961, equation 25). Therefore from (4), (12) and (15), we have that asymptotically
$$T_f = \Sigma\left\{F_i - G_i - E_\alpha(F-G) - \frac{C_\alpha(F-G, F_\alpha)}{V_\alpha(F_\alpha)}F_{i,\alpha}\right\}. \tag{16}$$

This is the sum of deviations of the $F_i - G_i$ from their regression on $F_{i,\alpha}$. Hence, under H_f, the statistic T_f is asymptotically normally distributed with zero mean and variance
$$n\left\{V_\alpha(F-G) - \frac{C_\alpha^2(F-G, F_\alpha)}{V_\alpha(F_\alpha)}\right\}. \tag{17}$$

It will usually be necessary in applications to estimate (17) consistently from the data.

This result can be generalized in two main ways. First, if $Y_1, \ldots, Y_n$ are independently but not identically distributed, (17) generalizes immediately to
$$\Sigma V_\alpha(F_i - G_i) - \frac{\{\Sigma C_\alpha(F_i - G_i, F_{i\alpha})\}^2}{\Sigma V_\alpha(F_{i.\alpha})}. \tag{18}$$

A second generalization applies if $\boldsymbol{\alpha}$ is a vector parameter; the previous argument applies without change if $\boldsymbol{\beta}$ is a vector. Let
$$F_{i,\alpha_j} = \frac{\partial \log f(Y_i, \boldsymbol{\alpha})}{\partial\alpha_j}. \tag{19}$$

Suppose, for simplicity, that $Y_1, \ldots, Y_n$ are identically distributed. Then the maximum-likelihood estimators, $\hat\alpha_j$, satisfy approximately equations analogous to (4), namely
$$n\sum_k E_\alpha(F_{\alpha_j}F_{\alpha_k})(\hat\alpha_k - \alpha_k) = -\sum_i F_{i,\alpha_j}$$
from which
$$\hat\alpha_k - \alpha_k = -\frac{1}{n}\sum_l E_\alpha(F.F.)^{kl}\sum_i F_{i,\alpha_l},$$
where $E_\alpha(F.F.)^{kl}$ is an element of the matrix inverse to $E_\alpha(F_{\alpha_j}F_{\alpha_k})$.

The extension of (16) is
$$T_f = \sum_i\{F_i - G_i - E_\alpha(F-G) - \sum_{k,l}E_\alpha(F.F.)^{kl}F_{i,\alpha_l}\}, \tag{20}$$

and hence, under H_f, the statistic T_f is asymptotically normally distributed with mean zero and with variance n times the variance of $F-G$ adjusted for multiple linear regression on the F_{α_j}.

We shall not consider here the extensions to dependent random variables that would be necessary if the present methods were to be applied to problems about time series.

5. The Calculation of the Test Statistic

The application of the above procedure involves essentially four steps, namely the calculation of

(i) the maximum-likelihood estimators $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$;
(ii) the log likelihood ratio $L_f(\hat{\boldsymbol{\alpha}})-L_g(\hat{\boldsymbol{\beta}})$;
(iii) the expected value $E_{\hat{\boldsymbol{\alpha}}}\{L_f(\hat{\boldsymbol{\alpha}})-L_g(\hat{\boldsymbol{\beta}})\}$;
(iv) the asymptotic variance of T_f.

The asymptotic distribution theory will still hold if instead of maximum-likelihood estimators any asymptotically equivalent method of estimation is used. We shall not consider here simplifications to the whole procedure that might be possible in particular cases by the use of asymptotically equivalent statistics.

Point (ii) is straightforward. Point (iii) is often the one to cause difficulty. The expectation can be obtained theoretically, or by numerical calculation of the appropriate sums or integrals, or by approximate evaluation using a series expansion. It may sometimes be convenient to find the expectation of a function asymptotically equivalent to the maximized log likelihood ratio. Two general methods are available for (iv). One is to find (17), or its equivalent, by calculating the appropriate variances and covariances mathematically. The other method is to evaluate the log likelihood ratio F_i-G_i for each observation, replacing $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\alpha}}$ and $\boldsymbol{\beta}_{\hat{\boldsymbol{\alpha}}}$, and also to find F_{i,α_j}. The variance of F_i-G_i adjusting for regression on F_{i,α_j} is then found by standard linear regression methods. If the observations are divided into groups with a different distribution in each group, a pooled regression within groups is taken.

In the remaining sections of the paper some examples are given. These are chosen primarily to illustrate the tests in fairly simple terms. The most interesting applications are likely to be to situations where two alternative rather specific theories are available.

6. The Log-Normal Distribution *versus* The Exponential Distribution

Suppose that $Y_1, \ldots, Y_n$ are independently and identically distributed. Let the null hypothesis H_f be that the p.d.f. is log-normal, namely

$$\frac{1}{y\sqrt{(2\pi\alpha_2)}}\exp\left\{-\frac{(\log y-\alpha_1)^2}{2\alpha_2}\right\}$$

and let H_g be the hypothesis that the p.d.f. is exponential, namely

$$\frac{1}{\beta}e^{-y/\beta}.$$

Cox (1961, section 9) developed the test for this problem by an argument slightly different from that used here. It is easily verified that the same results follow from the formulae of section 4.

Briefly, $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are the sample mean and variance of the log Y_i and $\hat{\beta}$ is the sample mean of the Y_i. Under H_f, $\hat{\beta}$ converges in probability to $\beta_\alpha = e^{\alpha_1 + \frac{1}{2}\alpha_2}$. It is easily shown that

$$T_f = n \log(\hat{\beta}/\beta_{\hat{\alpha}}). \tag{21}$$

To calculate the asymptotic variance of (21), we have that

$$F - G = -\tfrac{1}{2}\log(2\pi\alpha_2) - \log Y - \frac{(\log Y - \alpha_1)^2}{2\alpha_2} + \log\beta - \frac{Y}{\beta}, \tag{22}$$

$$F_{\alpha_1} = \frac{\log Y - \alpha_1}{\alpha_2}, \tag{23}$$

$$F_{\alpha_2} = -\frac{1}{2\alpha_2} + \frac{(\log Y - \alpha_1)^2}{2\alpha_2^2}. \tag{24}$$

In (22) β is to be replaced by $\beta_\alpha = e^{\alpha_1 + \frac{1}{2}\alpha_2}$.

The procedure is now to find the multiple regression equation of (22) on (23) and (24) by evaluating the appropriate covariance matrix. The asymptotic variance of T_f is, from the generalization of (17),

$$n(e^{\alpha_2} - 1 - \alpha_2 - \tfrac{1}{2}\alpha_2^2).$$

Hence under H_f

$$\frac{\log(\hat{\beta}/\beta_{\hat{\alpha}})}{\{(e^{\hat{\alpha}_2} - 1 - \hat{\alpha}_2 - \frac{1}{2}\hat{\alpha}_2^2)/n\}^{\frac{1}{2}}} \tag{25}$$

is asymptotically normally distributed with mean zero and unit variance, whereas under H_g the expectation of (25) is negative.

If the exponential distribution is taken as the null hypothesis instead of the log-normal, the test statistic has an entirely different form.

7. The Poisson Distribution *versus* the Geometric Distribution

The next example is, so far as I know, unlikely to be of practical importance, but may be of interest in that it shows the relation between "exact" and asymptotic theory.

Let $\mathbf{Y} = (Y_1, \ldots, Y_n)$ be independent and identically distributed. Let H_f be the hypothesis that the distribution is of the Poisson form with probability function

$$\frac{e^{-\alpha}\alpha^y}{y!} \quad (y = 0, 1, 2, \ldots), \tag{26}$$

and H_g the hypothesis of a geometric distribution with probability function

$$\frac{\beta^y}{(1+\beta)^{y+1}} \quad (y = 0, 1, 2, \ldots). \tag{27}$$

The mean of (27) is equal to β; note the range of the distribution. The distributions (26) and (27) are of roughly similar shape if the mean is less than one.

The maximum-likelihood estimators $\hat{\alpha}$ and $\hat{\beta}$ are both equal to the sample mean $\bar{Y}$. The log likelihood-ratio is

$$L_f(\hat{\alpha}) - L_g(\hat{\beta}) = n(1 + \bar{Y})\log(1 + \bar{Y}) - n\bar{Y} - \Sigma \log Y_i! \tag{28}$$

The expectation of (28) under the Poisson distribution (26) has now to be calculated. Series expansion shows that

$$E_\alpha\{n(1+\bar{Y})\log(1+\bar{Y})\} = n(1+\alpha)\log(1+\alpha)+O(1).$$

Let
$$l_f(\alpha) = E_\alpha(\log Y!) = \sum_{y=0}^{\infty} \log y!\,\frac{e^{-\alpha}\,\alpha^y}{y!}. \tag{29}$$

Then the test statistic is
$$T_f = -\Sigma\log Y_i!+nl_f(\bar{Y}). \tag{30}$$

To find the asymptotic variance of T_f, we put $\beta = \beta_\alpha = \alpha$ and consider the random variables

$$F-G = (Y+1)\log(1+\alpha)-\log Y!-\alpha, \tag{31}$$

$$F_\alpha = -1+Y/\alpha. \tag{32}$$

The covariance matrix of $F-G$ and F_α is

$$\begin{bmatrix} \alpha\{\log(1+\alpha)\}^2+V_\alpha(\log Y!)-2\log(1+\alpha)\,C_\alpha(Y,\log Y!) & \log(1+\alpha)-C_\alpha(Y,\log Y!)/\alpha \\ \log(1+\alpha)-C_\alpha(Y,\log Y!)/\alpha & 1/\alpha \end{bmatrix}$$

and hence the variance of $F-G$ adjusted for regression on F_α is

$$V_\alpha(\log Y!)-\frac{C_\alpha^2(Y,\log Y!)}{\alpha} = v_f(\alpha), \tag{33}$$

say. The test is thus to treat

$$\frac{-\Sigma\log Y_i!+nl_f(\bar{Y})}{\{nv_f(\bar{Y})\}^{\frac{1}{2}}} \tag{34}$$

as having a standard normal distribution under H_f, negative values being expected under H_g.

Suppose now that the roles of H_f and H_g are interchanged, the geometric distribution being taken as the null hypothesis. Parallel calculations lead to a test based on the asymptotic standard normal distribution of

$$\frac{\Sigma\log Y_i!-nl_g(\bar{Y})}{\{nv_g(\bar{Y})\}^{\frac{1}{2}}} \tag{35}$$

where
$$l_g(\beta) = E_\beta(\log Y!) = \sum_{y=0}^{\infty}\log y!\,\frac{\beta^y}{(1+\beta)^{y+1}}, \tag{36}$$

and
$$v_g(\beta) = V_\beta(\log Y!)-\frac{C_\beta^2(Y,\log Y!)}{\beta(\beta+1)}. \tag{37}$$

An alternative method of obtaining the standard errors in the denominators of (34) and (35) is by estimating by standard regression methods the variance of $\log Y_i!$ adjusting for linear regression on Y_i. The general procedure is to find the variance of F_i-G_i adjusting for linear regression on $F_{i,\alpha}$; for calculations under H_f, α and β are replaced by $\hat{\alpha}$ and $\beta_{\hat{\alpha}}$. In the more complex situations, this approach is likely to be the best.

Table 1 is a short table of the functions needed to compute (34) and (35); the extension of the table would be straightforward.

TABLE 1

Functions required for testing Poisson and geometric distributions

y	$\log y!$	y	$l_f(y)$	$v_f(y)$	$l_g(y)$	$v_g(y)$
0,1	0	0·2	·0134	·0082	·0256	·0152
2	0·69	0·4	·0524	·0284	·0957	·0697
3	1·79	0·6	·1169	·0554	·203	·164
4	3·18	0·8	·199	·0859	·341	·270
5	4·79	1·0	·304	·117	·508	·398
6	6·58	1·2	·428	·149	·698	·560

In order to illustrate how the results of applying the tests of this paper may be interpreted it is best to consider a numerical example.

Example. The second column of Table 2 specifies a sample of 30 observations, in fact drawn from a Poisson distribution of mean 0·8. Fitted frequencies from Poisson and geometric distributions are also given.

We have that $\bar{Y} = 0{\cdot}867$, $\Sigma \log Y_i! = 5{\cdot}93$. The value of the maximized log likelihood-ratio is, by (28), 4·05. Since $e^{4\cdot05} \simeq 57$, this corresponds to a fairly substantial odds in favour of the Poisson distribution. (The precise interpretation of the likelihood ratio is, however, not completely clear because of the two maximizations. It is assumed that no conditional prior distributions of the parameters α and β are available, so that a Bayesian approach is inapplicable.)

TABLE 2

Example illustrating test for Poisson and geometric distributions

Variate value	*Observed frequency*	*Estimated Poisson frequency*	*Estimated geometric frequency*
0	12	12·61	16·07
1	11	10·93	7·46
2	6	4·74	3·47
3	1	1·37	1·61
⩾4	0	0·35	1·40

Graphical interpolation in Table 1 gives $l_f(\bar{Y}) \simeq 0{\cdot}226$, $v_f(\bar{Y}) \simeq 0{\cdot}95$, so that from (30) $T_f = 0{\cdot}85$ and T_f divided by its standard error is $0{\cdot}85/\sqrt{(30 \times 0{\cdot}95)} = 0{\cdot}16$, indicating excellent agreement with the Poisson distribution in the respect tested. Such departure as there is leads away from the geometric distribution, which would be expected to give negative values of T_f. A similar test based on T_g, taking the geometric distribution as the null hypothesis, gives a critical ratio $-2{\cdot}00$, indicating a significant departure at the 5 per cent. level from the null hypothesis, in the direction of the Poisson distribution Note that a two-tailed test is used because, if a large positive T_g had occurred, it would have been interpreted as evidence of a departure from the geometric distribution away from the Poisson distribution, and not as indicating consistency with the null hypothesis.

Thus the final conclusion from the tests is that there is a ratio of roughly 60 to 1 in favour of the Poisson distribution as opposed to the geometric distribution, that the data are completely consistent with the Poisson distribution in the respect tested, and that there is a significant departure from the geometric distribution at about the 5 per cent. level.

Essentially the same conclusion can be reached in several other ways. Thus the fitted frequencies for the Poisson distribution agree closely with the observed frequencies, whereas there is a fairly substantial departure from the geometric distribution (goodness-of-fit χ^2 nearly significant at the 5 per cent. level). Again, an obvious difference between the Poisson and geometric distributions of the same mean is the greater dispersion of the geometric distribution. This suggests the use of dispersion tests. For the Poisson distribution this takes the form

$$\frac{\Sigma(Y_i - \bar{Y})^2}{\bar{Y}} = 24{\cdot}8,$$

the distribution under the null hypothesis being closely χ^2 with 29 degrees of freedom. The corresponding expression for testing the geometric distribution is

$$\frac{\Sigma(Y_i - \bar{Y})^2}{\bar{Y}(1+\bar{Y})} = 13{\cdot}3.$$

The expectation of this is about 29, but the distribution under the null hypothesis does not seem to have been examined. It is likely to be more dispersed than the χ^2 distribution with 29 degrees of freedom. Of course, given $\bar{Y}$, the statistics $\Sigma(Y_i - \bar{Y})^2$ and $\Sigma \log Y_i!$ are quite highly correlated, so that it is unlikely that appreciably different conclusions would result from the dispersion tests and from the T_f and T_g tests.

We now compare briefly the above asymptotic test with the uniformly most powerful similar test for the problem. Let H_f be the null hypothesis. Then $\bar{Y}$ is a complete sufficient statistic for the parameter α and we therefore consider the conditional distribution of the observations, given $\bar{Y}$; the appropriateness of this distribution for testing conformity to the Poisson distribution was pointed out by Fisher (1950). The special feature of the present problem is that the conditional distributions under both null and alternative hypotheses do not involve unknown parameters. A direct application of the Neyman–Pearson lemma then shows that the uniformly most powerful similar test against H_g has a critical region formed from large values of $\Sigma \log Y_i!$ using the conditional distribution

$$\text{prob}\{Y_i = y_i (i = 1, \ldots, n) \,|\, \bar{Y} = \bar{y}, H_f\} = \frac{(n\bar{y})!}{\Pi y_i!} \cdot \frac{1}{n^{n\bar{y}}}. \tag{38}$$

The "exact" test thus has exactly the form of (34) except that an exact distribution is used rather than an asymptotic normal approximation.

We shall not examine the adequacy of the normal approximation or the efficiency of the above procedure relative to other possible tests, for example those based on the sample variance or on the sample proportion of zeros. Asymptotic calculations of power can easily be made from the information in Table 1.

8. QUANTAL RESPONSES

We consider now the simplifications brought about when the random variables $Y_1, \ldots, Y_n$ take values 0 and 1 only. Suppose further that we can group the Y's into k sets such that the Y's in a set are identically distributed under both H_f and H_g. Thus we can express the problem in terms of random variables $Z_1, \ldots, Z_k$ which are independently binomially distributed with indices $n_1, \ldots, n_k$ and with parameters $f_1(\alpha), \ldots, f_k(\alpha)$ under H_f and $g_1(\beta), \ldots, g_k(\beta)$ under H_g. For simplicity we treat here only the situation where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are scalar parameters, α and β.

An interesting special case arises when H_f and H_g correspond to one-hit and two-hit curves. That is, we have observations on groups of individuals at doses $x_1, \ldots, x_k$, the number of individuals giving positive response being recorded in each group. Under H_f, the probability of positive response at dose x_j is

$$f_j(\alpha) = 1 - e^{-\alpha x_j}, \tag{39}$$

whereas under H_g the corresponding probability is

$$g_j(\beta) = 1 - e^{-\beta x_j} - \beta x_j\, e^{-\beta x_j}. \tag{40}$$

The log likelihoods are, except for constants, in general

$$L_f(\alpha) = \Sigma Z_j \log f_j(\alpha) + \Sigma(n_j - Z_j) \log\{1 - f_j(\alpha)\}, \tag{41}$$

$$L_g(\beta) = \Sigma Z_j \log g_j(\beta) + \Sigma(n_j - Z_j) \log\{1 - g_j(\beta)\}. \tag{42}$$

Maximum-likelihood estimators $\hat{\alpha}$ and $\hat{\beta}$ can be obtained by standard methods. In principle asymptotically equivalent estimates can always be obtained non-iteratively by applying the method of least squares with estimated weights to suitably transformed observations. We do not do that here.

Explicit expressions for $\hat{\alpha}$ and $\hat{\beta}$ are not usually available so that direct evaluation of the expectations of $L_f(\hat{\alpha})$ and $L_g(\hat{\beta})$ is not possible. However, the general argument of section 3 shows that $L_f(\hat{\alpha}) - L_f(\alpha)$ is of order one in probability. Hence it is enough to evaluate $E_\alpha\{L_f(\alpha)\}$ and to replace α by $\hat{\alpha}$ in the answer. This is easy from (41). In fact,

$$E_\alpha\{L_f(\alpha)\} = \Sigma n_j f_j(\alpha) \log f_j(\alpha) + \Sigma n_j\{1 - f_j(\alpha)\} \log\{1 - f_j(\alpha)\}. \tag{43}$$

Similarly, $L_g(\hat{\beta}) - L_g(\beta_\alpha)$ is of order one in probability and we can replace $E_\alpha\{L_g(\hat{\beta})\}$ by

$$E_\alpha\{L_g(\beta_\alpha)\} = \Sigma n_j f_j(\alpha) \log g_j(\beta_\alpha) + \Sigma n_j\{1 - f_j(\alpha)\} \log\{1 - g_j(\beta_\alpha)\}. \tag{44}$$

For a given α, the value of β_α is found from the maximum-likelihood equation for $\hat{\beta}$, replacing the observations by their expectations under H_f.

The statistic T_f can be calculated from (41)–(44), replacing α and β_α by $\hat{\alpha}$ and $\beta_{\hat{\alpha}}$. An analogous calculation gives T_g.

To find the asymptotic variance of T_f under H_f, we use the results of section 4. In fact, let $Y_{j1}, \ldots, Y_{jn_j}$ give the individual responses forming Z_j, a positive response corresponding to $Y = 1$, a negative response to $Y = 0$. Then in the jth group

$$F - G = \begin{matrix} \log\{f_j(\alpha)/g_j(\beta)\} & (Y = 1), \\ \log[\{1 - f_j(\alpha)\}/\{1 - g_j(\beta)\}] & (Y = 0), \end{matrix} \tag{45}$$

and
$$F_\alpha = \begin{array}{ll} f_j'(\alpha)/f_j(\alpha) & (Y=1), \\ -f_j'(\alpha)/\{1-f_j(\alpha)\} & (Y=0). \end{array} \tag{46}$$

Define
$$h_j(\alpha,\beta) = \log\left[\frac{f_j(\alpha)\{1-g_j(\beta)\}}{\{1-f_j(\alpha)\}\,g_j(\beta)}\right]. \tag{47}$$

Then, for an observation in the jth group

$$V_\alpha(F-G) = f_j(\alpha)\{1-f_j(\alpha)\}\,h_j^2(\alpha,\beta), \tag{48}$$

$$V_\alpha(F_\alpha) = \{f_j'(\alpha)\}^2/[f_j(\alpha)\{1-f_j(\alpha)\}], \tag{49}$$

$$C_\alpha(F-G,F_\alpha) = f_j'(\alpha)\,h_j(\alpha,\beta). \tag{50}$$

Therefore, summing over all observations and groups and applying (18), we have that the asymptotic variance of T_f is

$$\Sigma n_j f_j(\alpha)\{1-f_j(\alpha)\}\,h_j^2(\alpha,\beta) - \frac{\{\Sigma n_j f_j'(\alpha)\,h_j(\alpha,\beta)\}^2}{\Sigma n_j\{f_j'(\alpha)\}^2/[f_j(\alpha)\{1-f_j(\alpha)\}]}. \tag{51}$$

In this α and β are to be replaced by $\hat{\alpha}$ and $\beta_{\hat{\alpha}}$.

The statistic T_g and its asymptotic variance under H_g are obtained by interchanging the roles of $f_j(\alpha)$ and $g_j(\beta)$.

Example. Pereira and Kelly (1957) have given dilution series of adenovirus in tissue culture (HeLa cell culture). Table 3 gives one set of results for a 9-day inoculation period. The two hypotheses H_f and H_g of interest are the one-hit and two-hit curves of (39) and (40). The log likelihoods can be written down directly from (41) and (42) and the following maximum-likelihood estimates obtained: $\hat{\alpha} = 0{\cdot}413$, $\hat{\beta} = 0{\cdot}904$. From these, the expected values in Table 3 can be calculated, and also the values of $L_f(\hat{\alpha})$ and $L_g(\hat{\beta})$ obtained.

TABLE 3

Dilution series of adenovirus in tissue culture (Pereira and Kelly's data)

Concn. x_j	*No. of indiv.* n_j	*Obs. pos.* Z_j	*Obs. neg.*	*One-hit: expected pos.*	*One-hit: expected neg.*	*Two-hit: expected pos.*	*Two-hit: expected neg.*
0·5	32	3	29	6·0	26·0	2·4	29·6
1	32	10	22	10·8	21·2	7·3	24·7
2	30	19	11	16·9	13·1	16·2	13·8
4	31	27	4	25·1	5·9	27·2	3·8
8	32	30	2	30·8	1·2	31·8	0·19

First consider the application of standard χ^2 goodness-of-fit tests. For comparison with the one-hit curve, we have $\chi_3^2 = 3{\cdot}8$. The construction of a test of H_f against more specific alternatives is likely to be roughly equivalent to extracting a single degree of freedom from χ_3^2. Now, not only is $\chi_3^2 = 3{\cdot}8$ not significant, but also a single degree of freedom χ^2 of this size is short of the 5 per cent. point. Therefore it is unlikely that decisive evidence against H_f will be obtained by any test.

The value of χ^2_3 for testing H_g is 19·8, of which 17·2 is contributed by one cell with expected frequency 0·2. Such an expected frequency is too small for the applicability of the standard formulae. Yet if the results for this dose-level are pooled with those at the next level, the resulting value of χ^2 is only 2·5 and it is clear that most of the information bearing on the choice between the two curves has been lost. The general conclusion thus seems to be that the observation at dose-level 8 is crucial and is probably evidence against the two-hit curve. Apart from this, both curves seem to fit well. It is unlikely that much can be added for this individual example, but, as a general procedure, this is unsatisfactory. In particular, the χ^2 goodness-of-fit tests will be insensitive, especially when the degrees of freedom are larger. A more incisive method of analysis is desirable. While it is unlikely that the asymptotic distribution theory of sections 3 and 4 is adequate here, it is worth while seeing the answers obtained from the test statistics T_f and T_g.

The following additional calculations are necessary:

(i) The calculation of $\beta_{\hat\alpha}$ and $\alpha_{\hat\beta}$. To find, for example, $\beta_{\hat\alpha}$ a maximum-likelihood estimation of β is made using as "observations" the expected frequencies under H_f. Thus in all four maximum-likelihood estimations are required. The method of calculation was to find quick preliminary estimates from the responses at the middle doses. Then the derivative of log likelihood was computed at three values of α (or β) and the value giving zero derivative found by graphical interpolation and confirmed by direct calculation. The further work in finding $\beta_{\hat\alpha}$ and $\alpha_{\hat\beta}$, once $\hat\alpha$ and $\hat\beta$ have been found, is not great.

(ii) The values of $E_{\hat\alpha}\{L_f(\hat\alpha)\}$ and $E_{\hat\alpha}\{L_g(\hat\beta)\}$ are now obtained from (43) and (44), replacing α in these equations by $\hat\alpha$. The values of $E_{\hat\beta}\{L_g(\hat\beta)\}$ and $E_{\hat\beta}\{L_f(\hat\alpha)\}$ are obtained from corresponding equations. Note that all the log likelihoods are functions of the information type based on some combination of the observed frequencies and the expected frequencies under $\hat\alpha$, $\hat\beta$, $\alpha_{\hat\beta}$ and $\beta_{\hat\alpha}$.

(iii) The standard error of T_f is obtained directly from (51), with a corresponding equation for the standard error of T_g.

The numerical results are summarized in Table 4.

TABLE 4

Likelihood calculations on the data of Table 3

Estimates		*Log likelihoods*		*Test statistics*	
$\hat\alpha$	0·413	$L_f(\hat\alpha)$	−71·07	T_f	−2·79
$\alpha_{\hat\beta}$	0·403	$L_g(\hat\beta)$	−73·08	st. error	3·51
$\hat\beta$	0·904	$E_{\hat\alpha}(L_f(\hat\alpha)-L_g(\hat\beta))$	4·80	T_g	−5·71
$\beta_{\hat\alpha}$	0·915	$E_{\hat\beta}(L_g(\hat\beta)-L_f(\hat\alpha))$	3·70	st. error	2·32

First, the observed log likelihood ratio $L_f(\hat\alpha)-L_g(\hat\beta)$ is 2·01. If this could be interpreted approximately as if it referred to two simple hypotheses, it would represent a likelihood ratio of about 7 to 1, modestly in favour of the one-hit curve, and would be the relevant quantity in a pure discrimination problem. When multiplied by the prior odds, it would give the posterior odds in favour of the one-hit curve. However, such an interpretation, even asymptotically, is debatable when there are nuisance parameters (Cox, 1961, p. 109) and is defensible only under a special assumption about the conditional prior density functions of α and β. It does seem reasonable, however, to regard the ratio of maximized likelihoods as a measure of relative goodness of fit, the numbers of nuisance parameters being the same in H_f and H_g.

The negative value of T_f indicates that the log likelihood ratio is rather less in favour of H_f; the discrepancy is well within permissible limits of sampling error. Under H_g there is a discrepancy of log likelihood of $-5{\cdot}71$, about $2\frac{1}{2}$ times the standard error, and if the asymptotic distribution theory could be trusted, this would indicate quite strong evidence of inconsistency with H_g. The values of T_f and T_g seem to me of descriptive interest, even apart from their use in a formal significance test.

Examination of the calculations suggests that T_f and T_g are less sensitive to the single extreme observation than the χ^2 test of H_g, but this would need confirmation. Of course, in practice conclusions about H_f and H_g would hardly be reached from a single set of data. Also account would need to be taken of the possibility of a component of variance between trials in α and β.

9. A MULTIPLICATIVE *versus* AN ADDITIVE MODEL

We now consider a very simple problem involving a comparison of a multiplicative and an additive model. Let $Y_{11}, \ldots, Y_{1m}$; $Y_{21}, \ldots, Y_{2m}$ be two independent random samples. Let H_f be the usual normal theory model for the two-sample problem applying to the log Y's and let H_g be the corresponding model in terms of the Y's. That is, under H_f the log Y's are independently normally distributed with means α_1 and α_2 and with common variance α_3, whereas under H_g the Y's are independently normally distributed with means β_1 and β_2 and with common variance β_3.

There are two general points to make about this formulation. First, the ranges of the Y's are different under the two hypotheses. We assume, however, that under H_g the probability of a negative observation is negligible; an alternative formulation would replace the normal distributions in H_g by truncated normal distributions. Secondly, note that if it were required to examine solely distributional shape we would allow the two populations to have different variances. But suppose that two treatments are under comparison in a randomized experiment and that we are concerned with whether the treatment effect is additive or multiplicative. Then, on the multiplicative hypothesis, the standard deviation of Y should be proportional to the mean; hence relevant information is contained in the sample standard deviations and the present formulation is sensible. Similar problems arise in connection with all the standard analysis of variance situations.

Now, except for a constant,

$$L_f(\hat{\boldsymbol{\alpha}}) = -m(\hat{\alpha}_1 + \hat{\alpha}_2) - m \log \hat{\alpha}_3, \tag{52}$$

$$L_g(\hat{\boldsymbol{\beta}}) = -m \log \hat{\beta}_3, \tag{53}$$

where $\hat{\alpha}_3$ and $\hat{\beta}_3$ are the pooled variances within samples for $\log Y$ and Y. The statistics $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are the sample means for $\log Y$ and enter from the term $1/y$ in the log–normal density function.

To evaluate the expectation of (52) under H_f, we have first that $E_{\boldsymbol{\alpha}}(\hat{\alpha}_j) = \alpha_j$ $(j = 1, 2)$, $E_{\boldsymbol{\alpha}}(\log \hat{\alpha}_3) \sim \log \alpha_3$. Further, $E_{\boldsymbol{\alpha}}(\hat{\beta}_3) = \frac{1}{2} e^{\alpha_3}(e^{\alpha_3} - 1)(e^{2\alpha_1} + e^{2\alpha_2})$, since the variance of Y

in the jth population is $e^{2\alpha_j+\alpha_3}(e^{\alpha_3}-1)$. Thus

$$T_f = m \log(\hat{\beta}_3/\beta_{\hat{\alpha},3}), \tag{54}$$

where

$$\log \beta_{\hat{\alpha},3} = \hat{\alpha}_3 + \log(e^{\hat{\alpha}_3}-1) + \log \frac{e^{2\hat{\alpha}_1}+e^{2\hat{\alpha}_2}}{2}. \tag{55}$$

This statistic is of the type considered in the previous paper. The asymptotic variance of (54) can be obtained by the method of the previous paper, evaluating first the covariance matrix of $(\hat{\boldsymbol{\alpha}}, \hat{\beta}_3)$, or by the method of section 4 of the present paper. It is found after some elementary but tedious calculation that asymptotically

$$\begin{aligned}\frac{V_\alpha(T_f)}{m} &= \frac{(e^{4\alpha_1}+e^{4\alpha_2})}{(e^{2\alpha_1}+e^{2\alpha_2})}(e^{4\alpha_3}+2e^{3\alpha_3}+3e^{2\alpha_3}-4-4\alpha_3) - \frac{(2e^{\alpha_3}-1)^2}{(e^{\alpha_3}-1)^2}\alpha_3^2 \\ &= v_f^{(1)}(\alpha_1-\alpha_2)\, v_f^{(2)}(\alpha_3) - v_f^{(3)}(\alpha_3),\end{aligned} \tag{56}$$

say, where the functions occurring in (56) are tabulated in Table 5.

TABLE 5

Functions required in test of multiplicative null hypothesis

$\lvert\alpha_1-\alpha_2\rvert$	$v_f^{(1)}(\alpha_1-\alpha_2)$	α_3	$v_f^{(2)}(\alpha_3)$	$v_f^{(3)}(\alpha_3)$
0	0·500	0·00	2·00	1·00
0·2	0·520	0·05	2·66	1·16
0·4	0·572	0·10	3·46	1·32
0·6	0·644	0·15	4·41	1·51
0·8	0·720	0·20	5·55	1·70
1·0	0·790	0·25	6·90	1·90
2·0	0·965	0·30	8·51	2·12
∞	1·000	0·35	10·61	2·36

The variance of T_f is then estimated from (42) by inserting maximum-likelihood estimators and the test is made as in the other cases.

Consider now the associated problem in which H_g is the null hypothesis. Equations (52) and (53) are unchanged. A new point arises in evaluating the expectation of (52). For if X is a random variable normally distributed with mean μ and variance σ^2, properties of $\log X$ can be considered only conditionally on $X>0$. However, we shall only be interested in cases where prob $(X<0)$ is negligible, for otherwise choice between H_f and H_g will be easy. Hence it is in order to write

$$\begin{aligned}\log X &= \log\mu + \log\left(1+\frac{X-\mu}{\mu}\right) \\ &= \log\mu + \frac{X-\mu}{\mu} - \frac{(X-\mu)^2}{2\mu^2} + \dots,\end{aligned}$$

and to derive expansions such as that

$$E(\log X) = \log \mu - \frac{\sigma^2}{2\mu^2} - \frac{3\sigma^4}{2\mu^4}, \tag{57}$$

$$V(\log X) = \frac{\sigma^2}{\mu^2} + \frac{\sigma^4}{2\mu^4}. \tag{58}$$

We neglect powers of σ/μ higher than the fourth.

Thus we have the following approximate expressions for the limits under H_g of the estimates $\hat{\boldsymbol{\alpha}}$:

$$\alpha_{\boldsymbol{\beta},1} = \log \beta_1 - \frac{\beta_3}{2\beta_1^2} - \frac{3\beta_3^2}{\beta_1^4}, \tag{59}$$

$$\alpha_{\boldsymbol{\beta},2} = \log \beta_2 - \frac{\beta_3}{2\beta_2^2} - \frac{3\beta_3^2}{\beta_2^4}, \tag{60}$$

$$\alpha_{\boldsymbol{\beta},3} = \frac{\beta_3}{2}\left(\frac{1}{\beta_1^2} + \frac{1}{\beta_2^2}\right) + \frac{\beta_3^2}{4}\left(\frac{1}{\beta_1^4} + \frac{1}{\beta_2^4}\right). \tag{61}$$

The test statistic T_g is given by

$$T_g = m(\hat{\alpha}_1 + \hat{\alpha}_2 - \alpha_{\hat{\boldsymbol{\beta}},1} - \alpha_{\hat{\boldsymbol{\beta}},2}) + m \log(\hat{\alpha}_3/\alpha_{\hat{\boldsymbol{\beta}},3}). \tag{62}$$

The calculation of the asymptotic standard error of T_g under H_g by the formulae of section 4, or by the method of the previous paper, is tedious. The corresponding calculation would almost certainly be impracticable if we were dealing with a situation more complex than the simple two-sample problem under consideration here. In such cases, the following method is preferable.

Write down for each observation, Y_i, the value of $G_i - F_i$, replacing $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\alpha}$ by $\boldsymbol{\alpha}_{\hat{\boldsymbol{\beta}}}$. Write down similarly the values of the derivatives G_{i,β_l}. Construct the multiple analysis of covariance table of $G_i - F_i$ on the G_{i,β_l} and estimate the variance of T_g as the sum of squares within samples, adjusting for regression.

In the present application a sensible compromise is to calculate some of the terms theoretically, leaving the more complex ones to be estimated empirically. In fact, for observations in the first sample we have that

$$V_{\boldsymbol{\beta}}(G_{\beta_1}) = 1/\beta_3, \quad C_{\boldsymbol{\beta}}(G_{\beta_1}, G_{\beta_3}) = 0, \quad V_{\boldsymbol{\beta}}(G_{\beta_3}) = 1/(2\beta_3^2), \tag{63}$$

$$C_{\boldsymbol{\beta}}(G-F, G_{\beta_1}) = \frac{1}{\beta_1} + \frac{\beta_3}{\beta_1^3} + \frac{1}{\alpha_{\boldsymbol{\beta},3}}\left(\frac{\beta_3}{2\beta_1^3} + \frac{7\beta_3^2}{2\beta_1^5}\right), \tag{64}$$

with similar expressions in the second sample, replacing G_{β_1} by G_{β_2} and β_1 by β_2.

This leaves only $V_{\boldsymbol{\beta}}(F-G)$ and the average over the two samples of $C_{\boldsymbol{\beta}}(G-F, G_{\beta_3})$ to be estimated empirically. The analysis of covariance table can then be completed. Because G_{β_1}, G_{β_2} and G_{β_3} are uncorrelated, adjustments for regression can be made separately for the three independent variables and hence the variance of T_g estimated.

Example. Table 6 gives some artificial data obtained by random sampling from normal distributions of variance one and means 4 and 5. That is, H_g is true. The first part of Table 7 gives the means and variances of Y and log Y, the further statistics necessary to compute T_f and T_g, and the values of $L_f(\hat{\boldsymbol{\alpha}})-L_g(\hat{\boldsymbol{\beta}})$, T_f and T_g.

TABLE 6

Artificial data for comparison of multiplicative and additive models

Sample I:	4·1, 3·1, 5·0, 4·5, 3·0, 4·8, 4·3, 2·4, 2·6, 4·6, 4·2, 4·5, 4·0, 2·1, 6·1, 3·2, 4·3, 4·2, 2·7, 5·7
Sample II:	5·1, 2·5, 4·8, 4·7, 6·2, 3·8, 4·9, 4·8, 3·3, 5·0, 6·1, 6·0, 4·4, 4·4, 5·4, 5·4, 5·3, 4·4, 4·4, 4·8

TABLE 7

Summary of calculations on data of Table 6

Parameters of H_f: multiplicative model		*Parameters of H_g: additive model*	
$\hat{\alpha}_1$ 1·341	$\alpha_{\hat{\beta},1}$ 1·338	$\hat{\beta}_1$ 3·970	
$\hat{\alpha}_2$ 1·546	$\alpha_{\hat{\beta},2}$ 1·540	$\hat{\beta}_2$ 4·785	
$\hat{\alpha}_3$ 0·0619	$\alpha_{\hat{\beta},3}$ 0·0518	$\hat{\beta}_3$ 0·9427	$\beta_{\hat{\alpha},3}$ 1·2460
$L_f(\hat{\boldsymbol{\alpha}})$	const $-2·11$	$L_g(\hat{\boldsymbol{\beta}})$	const $+1·18$
T_f	$-5·58$	T_g	3·76
st. error 2·32		st. error	2·35 (theoretical)
			3·43 (empirical)

The only part of the calculation that calls for comment is that of the standard error of T_g. For this the values of $G-F$, G_{β_1}, G_{β_2}, G_{β_3} were obtained for each observation, $\boldsymbol{\beta}$ being replaced by $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\alpha}$ by $\boldsymbol{\alpha}_{\hat{\boldsymbol{\beta}}}$. Table 8 gives the matrix of pooled

TABLE 8

Matrix of corrected sums of squares and products within samples for $G-F$, G_{β_1}, G_{β_2}, G_{β_3}: (i) theoretical, (ii) empirical

		$G-F$	G_{β_1}	G_{β_2}	G_{β_3}
$G-F$	(i)	—	9·46	6·89	—
	(ii)	12·00	$-0·69$	1·93	$-0·60$
G_{β_1}	(i)		21·22	0	0
	(ii)		25·12	0	2·14
G_{β_2}	(i)			21·22	0
	(ii)			17·31	$-5·29$
G_{β_3}	(i)				22·51
	(ii)				19·73

sums of squares and products within samples, theoretical values from (63) and (64) being given first, and empirical values second. There is reasonably good agreement between (i) and (ii) except perhaps for the sum of products for $G-F$ with G_{β_1} and G_{β_2}. The sum of squares adjusted for regression using theoretical values whenever possible is 5·53 and that using empirical values in the first line of Table 8 is 11·78. These lead to the standard errors in Table 7.

Elementary examination of the data suggests that H_g fits better than H_f. First, the corrected sums of squares within the separate samples are 22·32 and 15·39 for Y and 1·623 and 0·854 for $\log Y$. These are more nearly constant for Y than for $\log Y$. Secondly, there is a little evidence of negative skewness for $\log Y$ and no evidence of skewness for Y. Note particularly one very low observation in the second sample. A further method often used in practice to choose between different scales of measurement is to see which scale gives the larger treatment effect relative to error. The values of the standard t statistic are 2·55 for $\log Y$ and 2·59 for Y, so that there is nothing to choose between the two hypotheses on this basis.

The value of $L_f(\hat{\boldsymbol{\alpha}})-L_g(\hat{\boldsymbol{\beta}})$ is $-3{\cdot}29$, which, subject to the qualification noted in discussing the previous example, indicates a likelihood ratio of about 27 to 1 in favour of the additive hypothesis. The ratio of T_f to its standard error is 2·4, significant at the 2 per cent. level in a two-sided test, indicating quite strong evidence against the multiplicative hypothesis; the negative sign of T_f shows that the departure is in the general direction of H_g. The positive sign of T_g indicates that the apparent departure from H_g is away from H_f; this is presumably a reflection of the fact that the sample with the higher mean value of Y has the smaller variance. The value of T_g is, however, only 1·6 or 1·1 times the estimated standard error, depending on which error is used, so that the data are reasonably consistent with H_g in the respect tested.

It is unlikely that in simple situations like this the present tests will show up anything that could not be detected by very elementary methods. Nevertheless, the present methods may be useful partly because the statistics $L_f(\hat{\boldsymbol{\alpha}})-L_g(\hat{\boldsymbol{\beta}})$, T_f and T_g are interesting ways of summarizing the data, and partly because approximate levels of significance are obtained.

10. Discussion

Possible further developments of the work in this paper include:

(i) the working out of more special cases, including tests analogous to those of section 9, but applying to more complicated analysis of variance situations;

(ii) the examination of the adequacy of the limiting normal approximation to the distribution of T_f, in particular the inclusion of the terms of order 1 in probability in the expansions of sections 3 and 4;

(iii) the calculation of power functions and the comparison with alternative tests, for which the distribution of T_f under H_g would be required;

(iv) the replacement of the test statistics by asymptotically equivalent statistics that are easier to compute;

(v) the extension to deal with problems about time series;

(vi) the investigation of possible asymptotic optimum properties.

Acknowledgement

I am very grateful to Dr J. J. Gart for suggesting the application to the comparison of one-hit and two-hit quantal dose response curves.

REFERENCES

CHERNOFF, H. (1956), "Large sample theory: parametric case", *Ann. math. Statist.*, **27**, 1–22.

COX, D. R. (1961), "Tests of separate families of hypotheses", *Proc. 4th Berkeley Symp.*, **1**, 105–123.

FISHER, R. A. (1950), "The significance of deviations from expectation in a Poisson series", *Biometrics*, **6**, 17–24.

PEREIRA, H. G. and KELLY, B. (1957), "Toxic and infective action of adenovirus", *J. gen. Microbiol.*, **17**, 517–524.