Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method

R. W. M. Wedderburn

Biometrika, Vol. 61, No. 3. (Dec., 1974), pp. 439-447.

Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28197412%2961%3A3%3C439%3AQFGLMA%3E2.0.CO%3B2-F

Biometrika is currently published by Biometrika Trust.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/bio.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact support@jstor.org.

http://www.jstor.org
Thu Jul 5 06:30:46 2007

Biometrika (1974), **61**, 3, 439
Printed in Great Britain

Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method

By R. W. M. WEDDERBURN
Rothamsted Experimental Station, Harpenden, Herts.

Summary

To define a likelihood we have to specify the form of distribution of the observations, but to define a quasi-likelihood function we need only specify a relation between the mean and variance of the observations and the quasi-likelihood can then be used for estimation. For a one-parameter exponential family the log likelihood is the same as the quasi-likelihood and it follows that assuming a one-parameter exponential family is the weakest sort of distributional assumption that can be made. The Gauss–Newton method for calculating nonlinear least squares estimates generalizes easily to deal with maximum quasi-likelihood estimates, and a rearrangement of this produces a generalization of the method described by Nelder & Wedderburn (1972).

Some key words: Estimation; Exponential families; Gauss–Newton method; Generalized linear models; Maximum likelihood; Quasi-likelihood.

1. Introduction

This paper is mainly concerned with fitting regression models, linear or nonlinear, in which the variance of each observation is specified to be either equal to, or proportional to, some function of its expectation. If the form of distribution of the observations were specified, the method of maximum likelihood would give estimates of the parameters in the model. For instance, if it is specified that the observations have normally distributed errors with constant variance, then the method of least squares provides expressions for the variances and covariances of the estimates, exact for linear models and approximate for nonlinear ones, and these estimates and the expressions for their errors remain valid even if the observations are not normally distributed but merely have a fixed variance; thus, with linear models and a given error variance, the variance of least squares estimates is not affected by the distribution of the errors, and the same holds approximately for nonlinear ones.

A more general situation is considered in this paper, namely the situation when there is a given relation between the variance and mean of the observations, possibly with an unknown constant of proportionality. A similar problem was considered from a Bayesian viewpoint by Hartigan (1969). We define a quasi-likelihood function, which can be used for estimation in the same way as a likelihood function. With constant variance this again leads to least squares estimation. When other mean-variance relationships are specified, the quasi-likelihood sometimes turns out to be a recognizable likelihood function; for instance, for a constant coefficient of variation the quasi-likelihood function is the same as the likelihood obtained by treating the observations as if they had a gamma distribution.

Then computational methods are discussed. The well-known Gauss–Newton method for calculation of nonlinear least squares estimates is generalized to provide a method of calculating maximum quasi-likelihood estimates.

When there exists a function of the means that is linear in the parameters, a rearrangement of the calculations in the generalized Gauss–Newton method gives a procedure identical to that described by Nelder & Wedderburn (1972). This method produces maximum likelihood estimates by iterative weighted least squares when the distribution of observations has a certain form and there is a transformation of the mean which makes it linear in the parameters. The distributions that can be treated in this way are those for which the likelihoods are identical to the quasi-likelihoods; thus the present result generalizes that of Nelder & Wedderburn.

The approach described in this paper sheds new light on some existing data-analytic techniques, and also suggests new ones. An example is given to illustrate the method.

2. Definition of the quasi-likelihood function

Suppose we have independent observations z_i $(i = 1, \ldots, n)$ with expectations μ_i and variances $V(\mu_i)$, where V is some known function. Later we shall relax this specification and say $\text{var}(z_i) \propto V(\mu_i)$. We suppose that for each observation μ_i is some known function of a set of parameters $\beta_1, \ldots, \beta_r$. Then for each observation we define the quasi-likelihood function $K(z_i, \mu_i)$ by the relation

$$\frac{\partial K(z_i, \mu_i)}{\partial \mu_i} = \frac{z_i - \mu_i}{V(\mu_i)}, \tag{1}$$

or equivalently

$$K(z_i, \mu_i) = \int^{\mu_i} \frac{z_i - \mu_i'}{V(\mu_i')} d\mu_i' + \text{function of } z_i.$$

From now on, when convenient, the subscript i will be dropped so that z and μ will refer to an observation and its expectation, respectively. Also, the notation $S(.)$ will be used to denote summation over the observations, so that $S(z)$ is the sum of the observations. We shall find that K has many properties in common with a log likelihood function. In fact, we find that K is the log likelihood function of the distribution if z comes from a one-parameter exponential family, as will be shown in § 4.

3. Properties of quasi-likelihoods

It is now shown that the function K has properties similar to those of log likelihoods.

Theorem 1. *Let z and K be defined as in § 2, and suppose that μ is expressed as a function of parameters $\beta_1, \ldots, \beta_m$. Then K has the following properties:*

(i) $E\left(\dfrac{\partial K}{\partial \mu}\right) = 0,$

(ii) $E\left(\dfrac{\partial K}{\partial \beta_i}\right) = 0,$

(iii) $E\left(\dfrac{\partial K}{\partial \mu}\right)^2 = -E\left(\dfrac{\partial^2 K}{\partial \mu^2}\right) = \dfrac{1}{V(\mu)},$

(iv) $E\left(\dfrac{\partial K}{\partial \beta_i}\dfrac{\partial K}{\partial \beta_j}\right) = -E\left(\dfrac{\partial^2 K}{\partial \beta_i \partial \beta_j}\right) = \dfrac{1}{V(\mu)}\dfrac{\partial \mu}{\partial \beta_i}\dfrac{\partial \mu}{\partial \beta_j}.$

Proof. First, (i) follows immediately from the definition of K. Then (ii) follows on noting that $\partial K/\partial\beta_i = (\partial K/\partial\mu)(\partial\mu/\partial\beta_i)$ and (iii) is a special case of (iv).

To prove (iv), we note that

$$E\left(\frac{\partial K}{\partial\beta_i}\frac{\partial K}{\partial\beta_j}\right) = E\left(\frac{\partial K}{\partial\mu}\right)^2 \frac{\partial\mu}{\partial\beta_i}\frac{\partial\mu}{\partial\beta_j}$$

$$= E\left[\frac{(z-\mu)^2}{\{V(\mu)\}^2}\right]\frac{\partial\mu}{\partial\beta_i}\frac{\partial\mu}{\partial\beta_j}$$

$$= \frac{1}{V(\mu)}\frac{\partial\mu}{\partial\beta_i}\frac{\partial\mu}{\partial\beta_j},$$

since $V(\mu) = \text{var}(z)$. Also we have

$$-E\left(\frac{\partial^2 K}{\partial\beta_i\partial\beta_j}\right) = -E\left[\frac{\partial}{\partial\beta_j}\left\{\frac{z-\mu}{V(\mu)}\frac{\partial\mu}{\partial\beta_i}\right\}\right]$$

$$= -E\left[(z-\mu)\frac{\partial}{\partial\beta_j}\left\{\frac{1}{V(\mu)}\frac{\partial\mu}{\partial\beta_i}\right\} - \frac{1}{V(\mu)}\frac{\partial\mu}{\partial\beta_i}\frac{\partial\mu}{\partial\beta_j}\right]$$

$$= \frac{1}{V(\mu)}\frac{\partial\mu}{\partial\beta_i}\frac{\partial\mu}{\partial\beta_j},$$

which completes the proof.

A result which will be discussed further in § 4 is as follows.

COROLLARY. *If the distribution of z is specified in terms of μ, so that the log likelihood L can be defined, then*

$$-E\left(\frac{\partial^2 K}{\partial\mu^2}\right) \leqslant -E\left(\frac{\partial^2 L}{\partial\mu^2}\right). \tag{2}$$

Proof. From the theorem just proved, the above statement is equivalent to

$$\text{var}(z) \geqslant -1/E\left(\frac{\partial^2 L}{\partial\mu^2}\right),$$

a result which follows immediately from the Cramér–Rao inequality (Kendall & Stuart, 1973, § 7.14).

4. Likelihoods of exponential families

It is possible to define a log likelihood if a one-parameter family of distributions with μ as parameter is specified for z. The following theorem shows that the log likelihood function is identical to the quasi-likelihood if and only if this family is an exponential family.

THEOREM 2. *For one observation of z, the log likelihood function L has the property*

$$\frac{\partial L}{\partial\mu} = \frac{z-\mu}{V(\mu)}, \tag{3}$$

where $\mu = E(z)$ and $V(\mu) = \text{var}(z)$, if and only if the density of z with respect to some measure can be written in the form $\exp\{z\theta - g(\theta)\}$, where θ is some function of μ.

Proof. If $\partial L/\partial\mu$ has the form (3), then integrating with respect to μ and defining

$$\theta = \int\frac{d\mu}{V(\mu)},$$

we have the required result. Conversely, suppose for some measure m on the real line, the distribution of z is given by $\exp\{z\theta - g(\theta)\}\,dm(z)$. Then $\int e^{z\theta}\,dm(z) = e^{g(\theta)}$, and so the moment generating function $M(t)$ of z is

$$\int e^{z(\theta+t)}\,e^{-g(\theta)}\,dm(z) = e^{g(\theta+t)-g(\theta)}.$$

It follows that $g(\theta+t)-g(\theta)$, regarded as a function of t, is the cumulant generating function of z. Hence $g'(\theta) = \mu$ and $g''(\theta) = V(\mu)$; also $d\mu/d\theta = g''(\theta) = V(\mu)$. Then we have

$$\frac{\partial L}{\partial \mu} = \{z - g'(\theta)\}\frac{d\theta}{d\mu} = \frac{z-\mu}{V(\mu)}.$$

This completes the proof.

If K really is a log likelihood, then the theorem shows that, given $V(\mu)$, we can construct θ and $g(\theta)$ by integration. Theorem 9 of Lehmann (1959) shows that V and g must be analytic functions and that the characteristic function $\phi(t)$ of z is analytic on the whole real line and given by $\phi(t) = \exp\{g(\theta+it)-g(\theta)\}$. Thus, in principle, the problem of determining whether or not K is a log likelihood is reduced to the problem of determining whether a given complex function $\phi(t)$, analytic over the whole real line, is a characteristic function; this point will not be pursued further here.

In the corollary to Theorem 1 in § 3 it was shown that

$$-E\left(\frac{\partial^2 K}{\partial \mu^2}\right) \leqslant -E\left(\frac{\partial^2 L}{\partial \mu^2}\right).$$

Then Theorem 2 shows that this inequality becomes an equality for a one-parameter exponential family. Thus for a given mean-variance relationship, a one-parameter exponential family minimizes the information $-E(\partial^2 L/\partial\mu^2)$, provided that an exponential family exists for that relationship.

It seems reasonable to regard $-E(\partial^2 K/\partial\mu^2)$, which is equal to $1/\mathrm{var}\,(z)$, as a measure of the information z gives concerning μ when only the mean-variance relationship is known, and to regard $E\{\partial^2(K-L)/\partial\mu^2\}$, which is always nonnegative, as the additional information provided by knowing the distribution of z. From this point of view, assuming a one-parameter exponential family for z is equivalent to making no assumption other than the mean-variance relationship.

5. Estimation using quasi-likelihoods

This section discusses maximum quasi-likelihood estimates and shows that their precision may be estimated from the expected second derivatives of K in the same way as the precision of maximum likelihood estimates may be estimated from the expected second derivatives of the log likelihood.

For each observation, let u be the vector whose components are $\partial K/\partial\beta_i$. Then, from Theorem 1, u has mean 0 and dispersion matrix with elements

$$-E\left(\frac{\partial^2 K}{\partial\beta_i\,\partial\beta_j}\right).$$

Let $H = \partial^2 S(K)/\partial\beta_i\,\partial\beta_j$; then, if the observations are independent, $S(u)$ has mean 0 and dispersion $D = -E(H)$. Now let $\hat\beta$ be the maximum quasi-likelihood estimate of β, obtained by setting $S(u)$ equal to its expectation, 0. To first order in $\beta-\hat\beta$ we have $S(u) \simeq H(\beta-\hat\beta)$, whence $\beta-\hat\beta \simeq H^{-1}S(u)$.

Approximating to H by its expectation, $-D$, we have,

$$\hat{\beta} \simeq \beta + D^{-1}S(u).$$

Now $D^{-1} S(\mu)$ has dispersion D^{-1}; hence we have deduced, rather informally, the following result.

THEOREM 3. *Maximum quasi-likelihood estimates have approximate dispersion matrix* $D^{-1} = \{E(H)\}^{-1}$, *where H is the matrix of second derivatives of* $S(K)$.

Next, we consider the case where the mean-variance relation is not known completely, but the variance is known to be proportional to a given function of the mean, i.e. var $(z) = \gamma V(u)$, where V is a known function but γ is unknown. Clearly the maximum quasi-likelihood estimate of β is not affected by the value of γ, so that we can calculate $\hat{\beta}$ as if γ was known to be 1. To obtain error estimates we need some estimate of γ. Assuming that μ is approximately linear in β and that $V(\hat{u})$ differs negligibly from $V(\mu)$, we have the approximation

$$E\left[S\left\{\frac{(z-\mu)^2}{V(\mu)}\right\}\right] \simeq n-m,$$

which leads to an estimate of γ given by

$$\hat{\gamma} = \frac{1}{n-m} S\left\{\frac{(z-\mu)^2}{V(\mu)}\right\}.$$

For normal linear models, this gives the usual estimate of variance.

6. A GENERALIZATION OF THE GAUSS–NEWTON METHOD

When $V(\mu) = 1$, maximum quasi-likelihood estimation reduces to least squares. One method of calculating the estimates is then the Gauss–Newton method. This is an iterative process in which one calculates a regression of the residuals on the quantities $\partial\mu/\partial\beta_i$ by linear least squares, the residuals and $\partial\mu/\partial\beta_i$ being calculated from the current estimate of $\hat{\beta}$. The resulting regression coefficients are then used as corrections to $\hat{\beta}$. It will now be shown that to calculate maximum quasi-likelihood estimates with a general V, the Gauss–Newton method can be modified simply by using the current estimate of $1/V(\mu)$ as a weight variate in the least squares calculation.

Writing v_i for $\partial\mu/\partial\beta_i$, and r for $z-\mu$, we have

$$\frac{\partial S(K)}{\partial \beta_i} = S\left\{\frac{rv_i}{V(\mu)}\right\},$$

and using Theorem 2

$$-E\left\{\frac{\partial^2 S(K)}{\partial\beta_i \partial\beta_j}\right\} = S\left\{\frac{v_i v_j}{V(\mu)}\right\}.$$

Then if we obtain successive approximations to $\hat{\beta}$ using the Newton–Raphson method with the second derivatives of K replaced by their expectations we obtain corrections $\delta\beta$ to the estimates given, for $i = 1, \ldots, n$, by

$$\sum_j S\left\{\frac{v_i v_j}{V(\mu)}\right\} \delta\beta_j = S\left\{\frac{rv_i}{V(\mu)}\right\}. \tag{4}$$

Hence we have proved the following:

THEOREM 4. *Using the Newton–Raphson method with expected second derivatives of K to calculate $\hat{\beta}$ is equivalent to calculating iteratively a weighted linear regression of the residuals, r, on the derivatives of μ with respect to the β's with weight $1/V(\mu)$, and using the regression coefficients as corrections to $\hat{\beta}$.*

Here $V(\mu)$ and the derivatives of μ are calculated at the current estimate of $\hat{\beta}$.

7. GENERALIZED LINEAR MODELS

We now derive a result which includes the result of Nelder & Wedderburn (1972) as a special case. Suppose that some function of the mean $f(\mu)$ can be expressed in the form

$$f(\mu) = \Sigma\beta_i x_i = Y,$$

say, where the x's are known variables. Then in the notation of the previous section $v_i = x_i d\mu/dY$. Hence (4) may be rewritten

$$\Sigma S\left\{\frac{1}{V(\mu)}\left(\frac{d\mu}{dY}\right)^2 x_i x_j\right\}\delta\beta_j = S\left\{\frac{r}{V(\mu)}\frac{d\mu}{dY}x_i\right\}.$$

Then if β denotes the current estimates and $\beta^* = \beta + \delta\beta$, the corrected ones, and if $Y = \Sigma\beta_i x_i$ we have

$$\sum_j S\left\{\frac{1}{V(\mu)}\left(\frac{d\mu}{dY}\right)^2 x_i x_j\right\}\beta_j^* = S\left\{\frac{1}{V(\mu)}\left(\frac{d\mu}{dY}\right)^2\left(Y + r\frac{dY}{d\mu}\right)\right\},$$

which proves the next theorem.

THEOREM 5. *When $Y = f(\mu) = \Sigma\beta_i x_i$ a method equivalent to the generalized Gauss–Newton method already described is to calculate repeatedly a weighted linear regression of*

$$y = Y + (z-\mu)\frac{dY}{d\mu}$$

on $x_1, \ldots, x_m$ using as weighting variate

$$w = \left(\frac{d\mu}{dY}\right)^2 \Big/ V(\mu). \tag{5}$$

Nelder & Wedderburn showed that this technique could be used to obtain maximum likelihood estimates when there was a linearizing transformation of the mean $f(\mu)$, and the distribution of z could be expressed in the form

$$\pi(z;\, \theta, \phi) = \alpha(\phi)\exp\{z\theta - g(\theta) + h(z)\} + \beta(z, \phi),$$

where θ is a function of μ and ϕ is a nuisance parameter. For fixed ϕ this gives a one-parameter exponential family, so that the likelihood is the same as the quasi-likelihood. Also, by a simple extension of the argument used in Theorem 1 we have $\mathrm{var}(z) = g''(\theta)/\alpha(\phi)$. Hence the mean-variance relationship is of the form given in (3), and the result of Nelder & Wedderburn is a special case of Theorem 5.

A good starting approximation in this process is usually given by setting $\mu = z$ and calculating w from (5) and y as $f(z)$, but some modification may be needed when f has singularities at the ends of the range of possible z.

8. Example

J. F. Jenkyn in an unpublished Aberystwyth Ph.D. thesis, discussed the data of Table 1 which gives estimates of the percentage leaf area of barley infected with *Rhynchosporium secalis*, or leaf blotch, for 10 different varieties grown at 9 different sites in a variety trial in 1965.

Jenkyn applied the angular transformation to the data, and then applied the method of Finlay & Wilkinson (1963), calculating, for each variety, regressions of the transformed percentages on the site means of the transformed percentages. He found a marked relationship between the variety means and the slopes of the regressions and also between the variety means and the residual variances from the regression. Thus the angular transformation failed to produce additivity or to stabilize the variance; in fact, it appeared that a transformation with a more extreme effect at the ends of the range, or at least at the lower end, was needed; Jenkyn suggested a logarithmic transformation. Two others suggest themselves: the logistic transformation, $\log(p/q)$, and the complementary log log transformation, $\log(-\log q)$.

Table 1. *Incidence of* R. secalis *on leaves of* 10 *varieties grown at nine sites; percentage leaf area affected*

	Variety										
Site	1	2	3	4	5	6	7	8	9	10	Mean
1	0·05	0·00	0·00	0·10	0·25	0·05	0·50	1·30	1·50	1·50	0·52
2	0·00	0·05	0·05	0·30	0·75	0·30	3·00	7·50	1·00	12·70	2·56
3	1·25	1·25	2·50	16·60	2·50	2·50	0·00	20·00	37·50	26·25	11·03
4	2·50	0·50	0·01	3·00	2·50	0·01	25·00	55·00	5·00	40·00	13·35
5	5·50	1·00	6·00	1·10	2·50	8·00	16·50	29·50	20·00	43·50	13·36
6	1·00	5·00	5·00	5·00	5·00	5·00	10·00	5·00	50·00	75·00	16·60
7	5·00	0·10	5·00	5·00	50·00	10·00	50·00	25·00	50·00	75·00	27·51
8	5·00	10·00	5·00	5·00	25·00	75·00	50·00	75·00	75·00	75·00	40·00
9	17·50	25·00	42·50	50·00	37·50	95·00	62·50	95·00	95·00	95·00	61·50
Mean	4·20	4·77	7·34	9·57	14·00	21·76	24·17	34·81	37·22	49·33	

An attempt was made to analyze the data using a logistic transformation. To do this, zero had to be replaced by some suitably small value; since the value 0·01 % occurs in the data we could hardly replace zero by something greater than this, unless 0·01 % were to be increased too. It was found that some of these small values in the data gave large negative residuals which had a serious distorting effect on the analysis, and only when these values were ignored was it possible to obtain a satisfactory analysis.

The logistic transformation appeared to be about right for stabilizing variance and producing additivity except for the undesirable effect of the small observations. This led to a different formulation of the model which was the same to a first order of approximation, but which avoided the problems caused by applying a logistic transformation to small observations.

Let p_{ij} denote the proportion of leaf area infected in the ith variety at the jth site. Let $P_{ij} = E(p_{ij})$ and $Q_{ij} = 1-P_{ij}$. The model is stated as $\text{logit}\, P_{ij} = Y_{ij} = m+a_i+b_j$ and $\text{var}(p_{ij}) = P_{ij}^2 Q_{ij}^2$.

When the method of § 6 is applied the weighting variate is equal to 1. This is useful because it means that the iterative analysis remains orthogonal. The modified y variate, $f(\mu)$, takes the form of a 'working logit', namely

$$y_{ij} = Y_{ij} + \frac{p_{ij} - P_{ij}}{P_{ij} Q_{ij}}.$$

The quantities $r_{ij} = (p_{ij} - P_{ij})/(P_{ij} Q_{ij})$ can be regarded as weighted residuals. They are proportional to residuals divided by their estimated standard error. Evidently the calculations are like the usual ones for fitting a logistic model to quantal data, but simpler, because the weights are equal to 1.

When the model was fitted by the method described, there were, of course, clear differences between sites, and between varieties. The estimate of γ obtained from the residual mean square came to 0·995, a value which indicates high variability in the data. It implies, for instance, that if P_{ij} is 0·20, the standard deviation of p_{ij} is about 0·16. Clearly, for this to happen, p_{ij} would have to have a rather skew distribution for P_{ij} not near 0·5; examination of the weighted residuals showed this skewness.

Table 2. *Means over sites of fitted values of logit* $\hat{P}_{ij}$

	Variety									
	1	2	3	4	5	6	7	8	9	10
Mean of logit $\hat{P}_{ij}$	−4·05	−4·51	−3·96	−3·09	−2·69	−2·71	−1·71	−0·78	−0·91	−0·16

(Standard error ± 0·331.)

The mean values of Y_{ij} for each variety are shown, with their standard errors, in Table 2. Clearly there are differences between varieties; there seem to be 3 highly resistant varieties and 3 less so, while the remaining 4 are much more susceptible.

Starting with the final set of working logits, the technique of Finlay & Wilkinson (1963) was applied noniteratively, but there was no sign of any interaction; nor was there any when a single degree of freedom for nonadditivity (Tukey, 1949) was isolated.

It seems, then, that a simple summary of the data has been achieved which makes it easy to see what conclusions can be drawn. The simpler method of working with logit P_{ij} might have worked better if the variance had not been so large; part of the trouble is that with such a large variance the approximations

$$E(\text{logit}\, p_{ij}) \simeq \log P_{ij}, \quad \text{var}\,(\text{logit}\, p_{ij}) \simeq \text{var}\,(p_{ij})/(P_{ij}^2 Q_{ij}^2)$$

break down.

9. Conclusions

It may be difficult to decide what distribution one's observations follow, but the form of the mean-variance relationship is often much easier to postulate; this is what makes quasi-likelihoods useful. It has been seen how maximum quasi-likelihood estimation produced a satisfactory analysis of rather difficult data, and how these estimates can be computed.

Some procedures used in the past are best understood in terms of quasi-likelihoods. For instance, in probit analysis, when the variance of the observations is found to be greater than that predicted by the binomial distribution, it is common to accept the maximum likelihood estimates regardless, while estimating the degree of heterogeneity as in Chapter 4 of Finney (1971). If the variance is still proportional to binomial variance then this procedure can be justified in terms of quasi-likelihoods. Also Fisher (1949), finding that in some data the variance was proportional to the mean, treated them effectively as if they had a Poisson distribution, even though the measurement involved was a continuous one. Thus quasi-likelihoods improve understanding of some past procedures, as well as providing new ones.

The author wishes to thank the director of the National Institute of Agricultural Botany and Dr J. F. Jenkyn for their permission to use the data, Mr M. J. R. Healy whose comments on an earlier version of the paper improved the presentation and Mr R. W. Payne for running the calculations on the GENSTAT statistical program developed at Rothamsted.

References

FINLAY, K. W. & WILKINSON, G. N. (1963). The analysis of adaptation in a plant breeding programme. *Aust. J. Agric. Res.* **14**, 742–54.

FINNEY, D. J. (1971). *Probit Analysis*, 3rd edition. Cambridge University Press.

FISHER, R. A. (1949). A biological assay of tuberculins. *Biometrics* **5**, 300–16.

HARTIGAN, J. A. (1969). Linear Bayesian models. *J.R. Statist. Soc.* B **31**, 446–54.

KENDALL, M. G. & STUART, A. L. (1973). *The Advanced Theory of Statistics*, Vol. II, 3rd edition. London: Griffin.

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. New York: Wiley.

NELDER, J. A. & WEDDERBURN, R. W. M. (1972). Generalized linear models. *J.R. Statist. Soc.* A **135**, 370–84.

TUKEY, J. W. (1949). One degree of freedom for non-additivity. *Biometrics* **5**, 232–42.

[*Received November* 1973. *Revised June* 1974]