On the Elimination of Nuisance Parameters in the Proportional Odds Model

Peter McCullagh

Journal of the Royal Statistical Society. Series B (Methodological), Volume 46, Issue 2 (1984), 250-256.

Stable URL:
http://links.jstor.org/sici?sici=0035-9246%281984%2946%3A2%3C250%3AOTEONP%3E2.0.CO%3B2-S

Journal of the Royal Statistical Society. Series B (Methodological)

J. R. Statist. Soc. B (1984),
46, *No.* 2, *pp.* 250–256

On the Elimination of Nuisance Parameters in the Proportional Odds Model

By PETER McCULLAGH

Imperial College, London and University of British Columbia

[Received March 1982. Revised February 1983]

SUMMARY

The problem of eliminating nuisance parameters in the context of the proportional odds model is considered. The technique used involves a form of sequential conditioning but it is not equivalent to a conditional or partial likelihood. Applications to ordered responses in the context of a matched pairs design are considered but the method can also be used where the responses are continuous.

Keywords: CONDITIONAL INFERENCE; LINEAR LOGISTIC MODEL; MATCHED DATA; NUISANCE PARAMETER; NUISANCE FUNCTION; PROPORTIONAL ODDS; ORDINAL DATA

1. THE PROPORTIONAL ODDS MODEL

In those cases where the response variable can take one of a limited number of ordered values the proportional odds model is often plausible. Social science applications, where the k response categories constitute an ordinal measurement scale, are numerous and medical, biological, ecological and other applications are fairly common. In its simplest form the model involves $k-1$ nuisance parameters, here denoted by λ_j or $\lambda_j(x_0)$ (equivalent to $\exp(-\theta_j)$ of McCullagh, 1980), but when blocking or stratification is involved the number of nuisance parameters will generally be some large multiple of $k-1$. We define $\gamma_j(\mathbf{x})$ to be the probability that an individual with covariate $\mathbf{x}$ responds in category j or below. Mostly we work with the odds of "survival" beyond category j, $\lambda_j(\mathbf{x}) = \{1-\gamma_j(\mathbf{x})\}/\gamma_j(\mathbf{x})$ although we could equally well work with the reciprocal of λ.

The proportional odds model implies that

$$\lambda_j(\mathbf{x}) = \lambda_j(\mathbf{x}_0)\exp\{\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x}-\mathbf{x}_0)\}, \quad j = 1, \ldots, k-1, \tag{1}$$

where $\mathbf{x}_0$ is some arbitrary known base-line value and $\{\lambda_j(\mathbf{x}_0)\}$ is a set of $k-1$ arbitrary nuisance parameters. McCullagh (1980), Bennett (1983) and most other authors who have worked with the proportional odds model have used unconditional maximum likelihood to estimate simultaneously the $k-1$ nuisance parameters $\theta_j = -\log\lambda_j(\mathbf{x}_0)$ and the p-vector of regression parameters $\boldsymbol{\beta}$. These estimates are computationally straightforward and, in most applications, the usual asymptotic approximations are satisfactory. Note in particular that according to the formulation just given it is not necessary that the observations be grouped. For the asymptotic approximations to be satisfactory we require the total number of observations in the sample to be large.

Suppose now that observations all bearing on the same issue have been obtained from several blocks or strata indexed by $i = 1, \ldots, N$. Often it is reasonable to assume that a model of the form (1) applies to each stratum where the stratum parameters $\lambda_j^{(i)}(\mathbf{x}_0)$ vary in an arbitrary way but where $\boldsymbol{\beta}$ is constant across strata. Here the "usual" asymptotic methods are quite unsatisfactory particularly when the number of individuals per stratum is small and N is large. Even in the simple case (1) where there is a single stratum of moderate size the unconditional method may be suspect if k is a substantial fraction of the stratum total.

Present address: Maths Dept, Huxley Bldg, Imperial College, 180 Queen's Gate, London SW7 2BZ, UK.

 0035–9246–84–46250 $2.00

We now investigate a method of estimation that enables us to eliminate the nuisance parameters. The method involves conditioning but is not equivalent to the construction of any conditional or partial likelihood function in the sense of Cox (1975). It does however produce, in a fairly automatic way, consistent parameter estimates and consistent estimates of their precision. Here consistency refers to the limit $N \to \infty$ or, where N is fixed, the limit as the strata totals becomes large.

2. THE TWO-SAMPLE PROBLEM

We consider here the two-sample problem, which involves the elimination of $k-1$ nuisance parameters. From an applied viewpoint this is perhaps the least interesting case because the difference between the unconditional estimates and the "conditional" estimates used here is almost negligible even when k is large and many of the cells are empty. Similarly, the unconditional and "conditional" estimates of precision are almost identical. The two-sample problem is chosen partly because the computations involved are simple and partly because it is easy to see precisely what is being done.

Let x be the indicator variable for group membership, 1 for sample 1, 2 for sample 2 and set $x_0 = 1$ so that $\lambda_j(1)$ is the odds for the first group of falling in category $j+1$ or above. The proportional odds model (1) can be written

$$\lambda_j(2)/\lambda_j(1) = \exp(\beta), \quad j = 1, \ldots, k-1, \tag{2}$$

where we denote the common value of the ratio by $\psi = \exp(\beta)$. We suppose that the data y_{ij} can be considered as two multinomial samples with totals n_1, n_2 such that the parameter vectors satisfy (2). Here y_{ij} is the number of observations in the ith sample falling in category j. It is convenient to form the cumulative totals

$$s_j = y_{2.j+1} + \ldots + y_{2.k} \quad \text{and} \quad m_j = y_{.j+1} + \ldots + y_{.k}, \tag{3}$$

where the dot subscript indicates summation. Let $n = n_1 + n_2$ be the total sample size.

If inference for β were to be based on the random variable S_j alone for some fixed, specified j, this would usually be done via the reference distribution conditional on m_j which is

$$\Pr\{S_j = s_j \mid m_j\} = \binom{n_2}{s_j}\binom{n_1}{m_j - s_j} \exp(\beta s_j) \Big/ \sum_r \binom{n_2}{r}\binom{n_1}{m_j - r} \exp(\beta r), \tag{4}$$

where the sum in the denominator extends from $\max(0, m_j - n_1)$ to $\min(n_2, m_j)$. The advantage of the reference distribution (4) is that it is independent of λ and thus generates similar regions for β. There are $k-1$ such conditional distributions each one generating a different set of similar regions for β. We now show how this information may be combined.

The notation here deliberately suggests that we think of the $(k-1)$-vector $\mathbf{S}$ as if it were a sufficient statistic but in doing so it is important to remember that the reference set in (4) varies from one component of $\mathbf{S}$ to another. Let μ_j and v_j, both functions of β, be the mean and variance of S_j computed from (4). Suppose now that it makes sense to refer to the covariance matrix $\mathbf{V}$ of $\mathbf{S}$ and that it has the form $v_{ij} = c_{ij}(v_i v_j)^{1/2}$ for some known constants c_{ij}. This apparently empty assumption is in fact false because of the way the conditioning event changes from one component of $\mathbf{S}$ to another. A similar conceptual difficulty arises in the construction of partial likelihoods (Cox, 1975). Nevertheless we proceed to use $\mathbf{V}$ where necessary. The justification for doing so is that the precise form of $\mathbf{V}$ does not affect the consistency of our estimates, or, in some cases, even their numerical value, but only their efficiency and even here the effect is negligible unless the c's are grossly inappropriate.

If this were a linear exponential family problem where the components of $\mathbf{S}$ were independent, the maximum likelihood estimate $\tilde{\beta}_c$ would be given by the equation

$$\sum_j \{s_j - \mu_j(\tilde{\beta}_c)\} = 0. \tag{5}$$

In other words, the sufficient statistic is reduced to $S_.$ and the estimate $\tilde{\beta}_c$ is found by equating the observed $s_.$ to its conditional expectation, $\Sigma\mu_j(\tilde{\beta}_c)$. Even in the present context the estimator (5), which is numerically unaffected by the choice of $\mathbf{V}$, is consistent and remarkably efficient. Its assymptotic variance, assuming $\mathbf{V}$ to be a satisfactory measure of the dispersion of $\mathbf{S}$, is

$$\text{var}\,(\tilde{\beta}_c) = \mathbf{1}^{\mathrm{T}}\mathbf{V}\mathbf{1}/(\Sigma v_j)^2, \tag{6}$$

where $\mathbf{1}$ is the unit vector of length $k-1$. Expression (6) is valid as $n_1, n_2 \to \infty$, and k need not be fixed.

Numerical values for $\mathbf{C} = \{c_{lj}\}$ are required in (6) though not in (5). In the special case $\beta = 0$ the unconditional correlation between S_j and S_l is consistently estimated by

$$c_{lj} = [m_j(n-m_l)/\{m_l(n-m_j\}]^{\frac{1}{2}}, \quad l \leqslant j \tag{7}$$

with a similar expression for $l > j$. Furthermore, for $\beta = 0$, (7) is the exact conditional correlation of S_j and S_l given $(m_1, \ldots, m_{k-1})$. It seems plausible that (7) should hold in some approximate sense even for $\beta \neq 0$ and the calculations that follow are based on this assumption.

In the above discussion we have indicated that the mean vector and covariance matrix in the appropriate reference distribution for $\mathbf{S}$ are $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_{k-1})$ and $\mathbf{V}$, both involving known functions of β only. Using this information alone we can compute the weighted least squares estimate $\hat{\beta}_c$, re-computing $\mathbf{V}$ as a function of $\hat{\beta}_c$ at each iteration. It is readily shown (McCullagh, 1983) that $\hat{\beta}_c$ has minimum asymptotic variance among all linear influence estimators and is given by

$$\sum_j w_j\{s_j - \mu_j(\hat{\beta}_c)\} = 0 \tag{8}$$

where $\mathbf{w} = \mathbf{V}^{-1}\mathbf{v}$ and $\mathbf{v}$ is the $(k-1)$-vector of conditional variances. Here we have made use of the fact that $v_j = d\mu_j/d\beta$. In the two-sample problem studied here, $\mathbf{V}$ is a Green's matrix and $\mathbf{V}^{-1}$ is tri-diagonal, a property that greatly simplifies the computations in (8).

In fact, the expression on the left of (8) can be thought of as the derivative of a log likelihood function in the sense that the asymptotic variance of $\hat{\beta}_c$ can be obtained by further differentiation. We find that

$$\text{var}\,(\hat{\beta}_c) = (\Sigma w_j v_j)^{-1} \tag{9}$$

which is numerically slightly smaller than (6). Expression (9) is valid as $n_1, n_2 \to \infty$ and, again, k need not be fixed.

Comparing (8) with (5) we see that the statistic $S_.$ has been replaced by the weighted statistic $\Sigma w_j S_j$ with w depending on β. This makes sense in a general way because there is no reason outside of the linear exponential family to expect that a single statistic should have optimum properties uniformly for all β. Note that, at $\beta = 0$, $\Sigma w_j s_j$ is just the Wilcoxon statistic indicating that, locally near $\beta = 0$, $\hat{\beta}_c$ is fully efficient. This result can be verified directly by showing that, for β near zero, the variance of the unconditional maximum likelihood estimator is given by (9).

3. AN EXAMPLE

The following numerical example is used to indicate the magnitude of the likely differences between the conditional and unconditional estimates. The data given in Table 1a were obtained as part of an investigation into the effect of cheese additives on consumer preferences. Here $k = 9$, which is large, and the response categories range from 1 = strong disliking to 9 = strong liking. In fact the same 52 panelists tasted both cheeses but the correlations so induced are ignored in the present analysis. It is known that to ignore such positive correlations leads to conservative inferences.

The estimates together with their estimated standard errors are given in Table 2. The numerical differences between the estimates are almost negligible indicating that the eight nuisance

TABLE 1a
Response category in a cheese tasting experiment

Cheese	*Dislike*	2	3	4	5	6	7	8	*Like*	*Total*
B	6	9	12	11	7	6	1	0	0	52
C	1	1	6	8	23	7	5	1	0	52
Total	7	10	18	19	30	13	6	1	0	104

TABLE 1b
Fitted values for cheese tasting data

Cheese	*Dislike*	2	3	4	5	6	7	8	*Like*	*Total*
B	5.80	7.92	12.80	10.76	10.71	2.80	1.04	0.16	–	52
C	1.20	2.08	5.20	8.24	19.29	10.20	4.96	0.84	–	52

TABLE 2
Log odds ratio estimates for cheese tasting data

Parameter	*Estimate*	*S.E.*
Unconditional $\hat{\beta}_u$	1.637	0.379
Variable weights $\hat{\beta}_c$	1.624	0.391
Fixed weights $\tilde{\beta}_c$	1.660	0.402

parameters have little effect on the bias of $\hat{\beta}_u$. It would appear therefore that, in problems such as this one, unconditional estimation is entirely satisfactory. It would make sense therefore to choose between the estimates on the basis of computational convenience. The conditional estimators would appear to be simpler particularly if the computations are done by hand. One interesting property of both $\hat{\beta}_u$ and $\hat{\beta}_c$ is that unused categories may be deleted without affecting the estimate. This property is not shared by $\tilde{\beta}_c$ unless the unused categories occur at the beginning or at the end.

The analysis just given strongly suggests that, where model checking is to be based on residuals, these should be defined by $(s_j - \hat{\mu}_j)\,\hat{v}_j^{-\frac{1}{2}}$, $j = 1, \ldots, k-1$. We do not explore the implications of such a definition or compare it with the more conventional definition based on cell residuals. The two definitions yield quite different patterns and consequently give quite different impressions of the quality of the fit. In the present example the seven residuals based on s are (0.24, 0.22, −1.75, 0.35, 0.27, 0.86, 0.20), a striking pattern though apparently of no statistical significance. Fitted values based on $\hat{\beta}_c$ are given in Table 1b.

4. COMBINATION OF ODDS RATIOS FROM SEVERAL STRATA

We now consider problems where the number of nuisance parameters increases without limit and where unconditional maximum likelihood estimates are unsatisfactory. We suppose that the experiment comprises N strata where N is large. In the ith stratum there is a control group for which the odds of a response in category $j+1$ or above is $\lambda_j^{(i)}$. The corresponding odds for the treatment group are $\lambda_j^{(i)} \exp(\beta)$ so that the treatment effect, β, is constant across strata. Within each stratum, the problem is equivalent to that discussed in Section 2: the notation of Section 2 is generalized by adding the superscript i. Let $\mu_j^{(i)}$ and $v_j^{(i)}$ be the conditional mean and variance of $S_j^{(i)}$ given $m_j^{(i)}$. The corresponding mean vector is $\boldsymbol{\mu}^{(i)}$ and the covariance matrix, $\mathbf{V}^{(i)}$ is obtained using (7).

Now choose any fixed j, $1 \leqslant j < k$, and consider how inference for β would be made based on the dichotomy at j. This reduced problem has the linear exponential family form. From (4), each stratum contributes the factor

$$\exp\{l_j^{(i)}(\beta)\} = \exp(\beta s_j^{(i)}) \Big/ \sum_r \binom{n_2^{(i)}}{r} \binom{n_1^{(i)}}{m_j^{(i)} - r} \exp(\beta r) \tag{10}$$

to the likelihood function so that β would be estimated by equating the observed value of the sufficient statistic $s_j^{(\cdot)}$ to its conditional mean, $\mu_j^{(\cdot)}(\beta)$. Thus we restrict attention to the vector $\mathbf{S}^{(\cdot)} = (S_1^{(\cdot)}, \ldots, S_{k-1}^{(\cdot)})$ of sufficient statistics. Prior to the combination of information from the components of $\mathbf{S}^{(\cdot)}$ it is convenient to think of the log-likelihood function as a $(k-1)$-vector

$$\mathbf{l}^{(\cdot)}(\beta) = \{l_1^{(\cdot)}(\beta), \ldots, l_{k-1}^{(\cdot)}(\beta)\}, \tag{11}$$

where the jth component satisfies the equation

$$\partial l_j^{(\cdot)}(\beta)/\partial\beta = s_j^{(\cdot)} - \mu_j^{(\cdot)}. \tag{12}$$

Thus each element of (12) is $O_p(N^{\frac{1}{2}})$ with zero mean and covariance matrix $\mathbf{V}^{(\cdot)}$ which is $O(N)$ and this fact guarantees the consistency of the estimators described below.

Estimates of β may be found using either of the two methods described in Section 2. The fixed weight estimator $\tilde{\beta}_c$ is given by

$$\sum_j \{s_j^{(\cdot)} - \mu_j^{(\cdot)}(\tilde{\beta}_c)\} = 0 \tag{13}$$

with asymptotic variance

$$\operatorname{var}(\tilde{\beta}_c) = \mathbf{1}^{\mathrm{T}}\mathbf{V}^{(\cdot)}\mathbf{1}/(\Sigma v_j^{(\cdot)})^2. \tag{14}$$

The more efficient weighted least squares estimator $\hat{\beta}_c$ is given by

$$\sum_j w_j\{s_j^{(\cdot)} - \mu_j^{(\cdot)}(\beta_c)\} = 0, \tag{15}$$

where the vector $\mathbf{w}$ is given by $\mathbf{w} = \mathbf{V}^{(\cdot)-1}\mathbf{v}^{(\cdot)}$ with $\mathbf{v}^{(\cdot)} = (v_1^{(\cdot)}, \ldots, v_{k-1}^{(\cdot)})^{\mathrm{T}}$. The asymptotic variance of $\hat{\beta}_c$ is

$$\operatorname{var}(\hat{\beta}_c) = \{\Sigma w_j v_j^{(\cdot)}\}^{-1} \tag{16}$$

which is slightly less than (14).

Equation (15) but not (13) can be interpreted as the derivative of an artificially constructed quasi-likelihood function (Wedderburn, 1974), ensuring that the asymptotic variance (16) can be obtained by further differentiation.

We have referred to $\hat{\beta}_c$ as the weighted least squares estimator of β based on the vector $\mathbf{S}^{(\cdot)}$. The computations require iteration and the weights must be re-computed at each cycle. It is important to distinguish clearly between $\hat{\beta}_c$ and the weighted least squares estimator β^* based on $\mathbf{S}^{(1)}, \ldots, \mathbf{S}^{(N)}$ and its block-diagonal covariance matrix, $\operatorname{diag}\{\mathbf{V}^{(1)}, \ldots, \mathbf{V}^{(N)}\}$. The equation for β^* has the form

$$\sum_i \sum_j w_j^{(i)}\{s_j^{(i)} - \mu_j^{(i)}(\beta^*)\} = 0. \tag{17}$$

The example in the following section demonstrates that β^* is not consistent as $N \to \infty$ except when $k = 2$. This result seems paradoxical because weighted least squares estimates are consistent regardless of how the weights are chosen so that the dependence on i of the weights in (17) would seem to be irrelevant with regard to consistency. Inconsistency here is a result of selection induced by the weights $w_j^{(i)}$ which are random variables.

5. MATCHED PAIRS

Here we consider a design where each stratum consists of exactly one control and one treated individual (or case in the context of retrospective studies). It is readily demonstrated that the unconditional estimate of β is inconsistent and in fact that its probability limit is 2β. For a proof of this result in the case $k = 2$ see Andersen (1973).

The computations of Section 4 reduce to the following. All pairs for which the responses are identical contribute a fixed quantity (10) to the likelihood function and can therefore be ignored. Apart from an additive constant associated with the identical pairs, $s_j^{(\cdot)}$ is the number of pairs for which the control responded in category j or below and for which the treated individual responded in category $j+1$ or above. The corresponding number with treatments and controls interchanged is $m_j^{(\cdot)} - s_j^{(\cdot)}$. The jth component of the likelihood vector (11) treats $S_j^{(\cdot)}$ as binomial with index $m_j^{(\cdot)}$ and parameter $\pi = \exp(\beta)/\{1 + \exp(\beta)\}$. The covariance matrix $\mathbf{V}^{(\cdot)}$ has the form $\pi(1-\pi)\{\text{a matrix of constants}\}$ where the (j, l) element in the matrix of constants is the number of pairs for which the ordered response in an obvious notation is $(\leqslant j, > l)$ or $(> l, \leqslant j)$ for $j \leqslant l$. Thus to compute $\hat{\beta}_c$ no iteration is required because the model is linear ($\mu_j^{(\cdot)} = m_j^{(\cdot)}\pi$) and the weights do not depend on β. Effectively therefore, we treat $\Sigma w_j S_j^{(\cdot)}$ as the uniformly most powerful statistic for tests concerning β. The estimator $\hat{\beta}_c = \log\{\Sigma w_j s_j^{(\cdot)}/\Sigma w_j(m_j^{(\cdot)} - s_j^{(\cdot)})\}$ was given by McCullagh (1977): its asymptotic variance is $\{\Sigma w_j v_j^{(\cdot)}\}^{-1}$.

6. AN EXAMPLE INVOLVING MATCHED PAIRS

The distance vision data of Stuart (1953) provides a suitable example where the response categories are ordered and where the design involves matched pairs. These data have been thoroughly examined elsewhere (see, for example, Bishop *et al.*, 1975, p. 284), although not by the methods suggested here and generally not explicitly from the viewpoint of matched pairs or ordered categories. We find $\mathbf{s}^{(\cdot)} = (456, 700, 349)$, $\mathbf{m}^{(\cdot)} = (843, 1297, 646)$ giving three non-independent estimates 0.164, 0.159 and 0.161 which are remarkably close, perhaps too close. Apart from the factor $\pi(1-\pi)$, the diagonal elements of $\mathbf{V}^{(\cdot)}$ are equal to $\mathbf{m}^{(\cdot)}$: the three off-diagonal elements are 343, 102, 262. Thus we find $\mathbf{w} = (0.6380, 0.7076, 0.6123)$ giving $\hat{\beta}_c = 0.16105$ with standard error 0.0466. The fixed weight estimator is $\tilde{\beta}_c = \log(1505/1281) = 0.16115$ with standard error $0.0467 = \{\mathbf{1}^{\mathrm{T}}\mathbf{V}^{(\cdot)}\mathbf{1}/(\Sigma v_j^{(\cdot)})^2\}^{\frac{1}{2}}$.

A total of 1171 women had better vision in the right eye than in the left whereas 1010 women had better vision in the left eye. The inconsistent least squares estimator β^* is given by $\beta^* = \log(1171/1010) = 0.1479$. It is fairly easy to show that the probability limit of β^*/β lies between about $\frac{2}{3}$ and 1.0 depending on k and on the nuisance parameters.

For $k > 2$ the goodness of fit of at least one aspect of the proportional odds model may be tested by comparing the three separate estimates of β with the combined value $\hat{\beta}_c$ or $\tilde{\beta}_c$. This may be done by inspection or by constructing the appropriate quadratic form. In the present example such a test merely confirms what is already obvious.

7. REGRESSION MODELS

The principal idea in Section 4 is that the full log likelihood function or its derivative is formed by vector addition of the contributions from each stratum. The covariance matrix $\mathbf{V}^{(\cdot)}$ is formed in a similar way. The methods of Section 2 are then used to combine information from the components of the likelihood vector.

A similar method may be used where the covariates are quantitative instead of indicator variables. It is sufficient to consider a single stratum and the dichotomy at category j. For an individual with covariate $\mathbf{x}$ the odds of responding in category $j+1$ or above are given by

$$\lambda_j(\mathbf{x}) = \lambda_j(\mathbf{x}_0)\exp\{\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x} - \mathbf{x}_0)\}, \tag{18}$$

where the odds function $\lambda_j(x_0)$ varies from one stratum to another. The total number of observations above category j, m_j, is taken as fixed. Let $\mathbf{s}_j$ be the sum of the covariate vectors over these m_j observations. The contribution to the likelihood function of the dichotomy at category j is

$$\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{s}_j) \Big/ \sum_{r \in R_j} \exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{r}), \tag{19}$$

where R_j is a set containing $\binom{n}{m_j}$ elements each element being the sum of the $\mathbf{x}$'s over a subset of size m_j. Essentially (19) gives the probability that a random sample of size m_j from the finite population $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ was the one actually observed, assuming that each subset has selection probability proportional to $\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{r})$ where $\mathbf{r}$ is the sum of the $\mathbf{x}$'s in that subset.

Equation (19) gives the contribution of the ith stratum to the jth element of the likelihood vector. Moments of the distribution corresponding to (19) are required for computing parameter estimates. Approximations are given in the following section. Following the methods of Section 3, parameter estimates are found by equating either the sum of the observed $s_j^{(\cdot)}$ or a weighted sum of these to the corresponding combination of their conditional means.

8. COMPUTATIONAL ISSUES

Computational difficulties have already been mentioned in Section 7. In Sections 2 and 3 the only computational difficulty concerns the computation of the conditional mean and variance of the non-central hypergeometric distribution (4). In the numerical examples discussed here the exact calculation was carried out provided that the range, $\min(n_1, n_2, m_j, n - m_j)$, did not exceed 10. For values of 10 or more we use the approximation

$$v = \frac{n}{n-1}\left\{\frac{1}{\mu} + \frac{1}{n_1 - \mu} + \frac{1}{m_j - \mu} + \frac{1}{n_2 - m_j + \mu}\right\}^{-1} \tag{20}$$

together with the exact equation for μ in terms of v

$$(\psi - 1)\mu^2 - \mu\{n + (m_j + n_1)(\psi - 1)\} + m_j n_1 \psi + v(\psi - 1) = 0 \tag{21}$$

which can be obtained from Mantel and Hankey (1975). In (21) $\psi = \exp(\beta)$. As an initial estimate of v we use the variance at $\beta = 0$ which is $n_1 n_2 m_j (n - m_j)/\{n^2(n-1)\}$. These two equations give perfectly satisfactory approximations for maximum likelihood calculations and involve only modest computational effort.

ACKNOWLEDGEMENTS

I wish to thank Dr Graeme Newell who provided data from which Table 1a was extracted and Dr S. Bennett for helpful discussions.

REFERENCES

Andersen, E. B. (1973) *Conditional Inference and Models for Measuring*. Copenhagen: Mentalhygiejnisk Forlag.

Bennett, S. (1983) Analysis of survival data by the proportional odds model. *Statistics in Medicine*, **2**, 279–285.

Cox, D. R. (1975) Partial likelihood. *Biometrika*, **62**, 269–276.

McCullagh, P. (1977) A logistic model for paired comparisons with ordered categorical data. *Biometrika*, **64**, 449–453.

——— (1980) Regression models for ordinal data. *J. R. Statist. Soc.* B, **42**, 109–142.

——— (1983) Quasi-likelihood functions. *Ann. Statist.*, **11** 59–67.

Mantel, N. and Hankey, W. (1975) The odds ratio of a 2 × 2 contingency table. *Amer. Statistn.*, **29**, 143–145.

Stuart, A. (1953) The estimation and comparison of strengths of association in contingency tables. *Biometrika*, **40**, 105–110.

Wedderburn, R. W. M. (1974) Quasi-likelihood functions, generalized linear models and the Gauss-Newton method. *Biometrika*, **61**, 439–447.