Likelihood Ratio Power and Sample Size Calculations

BARRY W. BROWN, JAMES LOVATO, and KATHY RUSSELL*

This paper describes sample size and power calculations for experiments in which hypotheses are tested using likelihood ratio methods. The underlying mathematical results are briefly reviewed and examples of their use are presented. The methods are extended to cases in which the null hypothesis specifies equality of parameter values. Several examples demonstrate the handling of covariates and illustrate the use of constraints to reduce the dimension of the parameter space. Finally, computer implementation and available software are discussed.

KEY WORDS: Sample size; Power; Likelihood ratio; Asymptotic methods

1 Introduction

Likelihood ratio (LR) methods for calculating power and requisite sample sizes were published in Wilks (1938), but the required computations precluded their use in all but the simplest cases. Despite advances in computation, LR methods are still under-utilized. Reasons for the lack of use include the difficulty of formulating reasonable specific alternative hypotheses and an insufficient appreciation of the straightforward nature of the calculations. Here, we examine algorithms used to

*B. Brown is Professor of Biomathematics in the Department of Biomathematics, Box 237, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030. J. Lovato was and K. Russell is a member of the applications development staff in the same Department. This work was supported by Cancer Center Core grant CA16672 from the National Cancer Institute, a cooperative study agreement with IBM, and the personal generosity of Pat and Larry McNeil. The authors thank Vickie William and Betty Schwarz for expert editorial assistance.

compute sample size and power using LR methods, and we provide examples of the framing of alternative hypotheses.

Related work includes Muller and Peterson (1984) and O'Brien (1986a) who discuss exact power calculations for tests involving normally distributed error. Software for these calculations is available in Heitjan (1993). Since exact calculations are available for the normal distribution, we do not discuss those here. O'Brien (1986b) provides an expanded discussion of the log-linear model. Other asymptotic LR methods have been derived in Self and Mauritsen (1988) and Self, Mauritsen, and Ohara (1992).

This paper is organized as follows: a) Section 2 states the mathematical results for calculations and presents a straightforward application of them; b) Section 3 contains methods for testing the equality of model parameters and includes an example; c) Section 4 provides non-trivial examples with covariates and high-dimensional parameter spaces and d) Section 5 discusses computer implementation and available software.

2 Preliminaries

2.1 Example

The following is a prototypical trial. Some calculations relating to this trial are considered in Section 4.1. The purpose of the trial is to describe the probability of response versus dose; specifically, whether a logistic model with a quadratic term improves the fit over a linear logistic model. Doses of the agent, one dose per subject, are administered and responses are recorded. Linear and quadratic logistic models are fit to the data and the difference in log likelihoods of the two models are calculated. If twice this difference exceeds 3.84 (the upper 95th percentile of the χ^2_1 distribution), there is evidence at the 0.05 significance level that the quadratic term improves the fit.

The planning of such studies, particularly determining the sample size required to provide a specified power, requires a specification of the presumed true state of nature at which the calculation will be made. This state includes the specification of values for the parameters of the model, $\tilde{\theta}$, which in this case consists of the intercept, linear, and quadratic coefficients for the logistic model. The state also includes the design of the experiment, i.e., the dosages to be employed and the proportion of subjects to receive each dosage. The design can greatly influence the sample size required; however, the subject of obtaining good designs is outside the scope of the current work.

2.2 Mathematical Background

This presentation follows Cox and Hinkley (1974), Section 9.3.

Let $\mathbf{y} = \{y_1, \ldots, y_n\}$ be a realization of independent, identically distributed random variables,$\{Y_1, \ldots, Y_n\}$, having a density $f(y;\theta)$, where θ is a vector of unknown parameters. The log likelihood of the observations, $\mathbf{y}$, at the parameter value θ is

$$l(\mathbf{y};\theta) = \sum_{s=1}^{n} \log(f(y_s;\theta)). \tag{1}$$

Let $\tilde{\theta}$ denote the alternative hypothesis value of θ. The expected information matrix $\mathbf{i}$ is defined by

$$\mathbf{i}_{j,k} = -\sum_{l=1}^{n} \mathcal{E}_{\tilde{\theta}} \left(\frac{\partial^2 l(Y_l;\theta)}{\partial\theta_j \partial\theta_k} \right). \tag{2}$$

This expectation can be written in computable form as

$$\mathcal{E}_{\tilde{\theta}} \left(\frac{\partial^2 l(Y_l;\theta)}{\partial\theta_j \partial\theta_k} \right) = \int \frac{\partial^2 l(\mathbf{y};\theta)}{\partial\theta_j \partial\theta_k} f(y;\tilde{\theta}) dy, \tag{3}$$

where the integration is over the set of possible values of Y. Note that the expectation is calculated at the assumed true value of θ. If the Y_i assume discrete values, $f(y|\theta)$ is the probability that $Y = y$, and the integral in (3) is replaced by a sum over all possible outcomes, y.

The information matrix is the inverse of the asymptotic covariance matrix of the maximum likelihood estimates of θ. This will be used in Section 3.

The presence of covariates in the model requires another level of computation to obtain the information matrix. When the experimenter chooses covariate values, the information matrix is computed separately for each value, and the results are summed. In observational studies in which covariates are random, the information matrix is integrated, component by component, over the population covariate distribution.

The null hypothesis determines the values of specific components of θ, which is partitioned into (ψ, λ), where the first set of parameter values, ψ, have been set by the null hypothesis to the value ψ_0. The remaining parameters, λ, are not constrained by the null hypothesis. In the example, the null hypothesis constrains the coefficient of the quadratic term in the logistic model to be zero; the values of the intercept and linear coefficient are not constrained.

The expected information matrix, $\mathbf{i}$, is partitioned in correspondence to θ into

$$\begin{pmatrix} \mathbf{i}_{\psi\psi} & \mathbf{i}_{\psi\lambda} \\ \mathbf{i}_{\lambda\psi} & \mathbf{i}_{\lambda\lambda} \end{pmatrix}.$$

Let $\hat{\theta}$ be the maximum likelihood estimate of θ, $\hat{\theta}_0$ be the value that maximizes the likelihood subject to $\psi = \psi_0$, and $\tilde{\psi}$ be the presumed true value of ψ. The results central to LR power and sample-size calculations are as follows:

Result 1. The asymptotic distribution of $2[l(\hat{\theta}) - l(\hat{\theta}_0)]$ is noncentral χ^2 with noncentrality parameter

$$\left[\tilde{\psi} - \psi_0\right]^T \left[\mathbf{i}_{\psi\psi} - \mathbf{i}_{\psi\lambda}\mathbf{i}_{\lambda\lambda}^{-1}\mathbf{i}_{\lambda\psi}\right] \left[\tilde{\psi} - \psi_0\right] \tag{4}$$

and degrees of freedom equal to the number of elements in ψ_0.

Result 2 provides the distribution of a linearly transformed multivariate normal distribution; it is used in Section 3 for null hypotheses that impose equality constraints on parameters. For proof of this result, see Tong (1990), Section 3.1.1.

Result 2. If X is p-variate normal with mean vector μ and covariance matrix $\mathbf{\Sigma}$ and if A is an $(n \times p)$ matrix, then $A\ X$ is multivariate normal with mean $A\mu$ and covariance $A\ \mathbf{\Sigma}\ A^T$.

2.3 Example: Single-Group Binomial

What sample size is required to reject with power 0.8 the null hypothesis that a binomial probability is $p = 0.2$ when actually $p = 0.4$ using a 0.05 level, two-sided test?

The only parameter of the model is the probability of an event, p, so the information matrix is a scalar. For a single observation, the log likelihood is $\log(p)$ if the event occurs and $\log(1-p)$ if the event does not occur.

$$\begin{aligned} \mathbf{i} &= -\left[p \times \frac{\partial^2 \log(p)}{\partial p^2} + (1-p) \times \frac{\partial^2 \log(1-p)}{\partial p^2}\right] \\ &= \frac{1}{p(1-p)}. \end{aligned} \tag{5}$$

The expected information for a single observation is $1/(0.4 \times 0.6) = 4.17$, so for n observations, the expected information will be $4.17n$. Using (4), the noncentrality parameter for a null hypothesis value of p_0 is $(p_0 - 0.4)^2 \times 4.17n$, which at $p_0 = 0.2$ is $0.167n$.

The noncentrality parameter, ν, must be chosen to provide the desired power. We will reject the null hypothesis if twice the LR is greater than the 0.95 quantile of the (central) χ^2_1 distribution, i.e., 3.84. The probability of exceeding 3.84 is

$$1 - \chi^2_1(3.84|\nu),$$

where $\chi^2_d(x|\nu)$ is the cumulative distribution function of the noncentral chi-square distribution with d degrees of freedom and noncentrality parameter ν. The requisite value of ν was found using DSTATTAB, which is described in Section 5. This value is 7.85; hence n is 47.

The LR test is multi-sided as is the analysis of variance; the test detects any difference from the null hypothesis, not differences in a specified direction. If there is only one parameter of interest, a good approximation to the sample-size required by a one-sided test can be obtained by calculating the sample size required at twice the significance level of the two-sided test. Doubling the significance level above from

0.05 to 0.1 yields $n = 37.2$, a somewhat conservative value. An exact computation shows that a one-sided test with 35 trials has a significance level of 0.034 and a power of 0.805.

3 Equality Constraints

Null hypotheses that constrain parameter values to be equal necessitate problem reformulation before Result 1 can be applied. Let θ be the vectors of parameters in the original formulation. A linear transformation, A, is applied to θ so that Result 1 holds for the reparameterization, $\theta' = A\theta$.

A constraint, $\theta_j = \theta_k$, results in a reparameterization: $\theta'_l = \theta_j - \theta_k$ and $\theta'_m = \theta_k$ for some subscripts l and m. θ'_l is constrained by the null hypothesis while θ'_m is not. The reparameterization of a chain of equality constraints is similar. Let the chain be:

$$\theta_{j_1} = \theta_{j_2} = \cdots = \theta_{j_p}.$$

Then for some $l_1, \ldots, l_n$, $\theta'_{l_n} = \theta_{j_n} - \theta_{j_{n+1}}$, for $n = 1, \ldots, (p-1)$ and $\theta'_{l_p} = \theta_{j_p}$. The first $p-1$ of these θ' parameters are constrained by the null hypothesis; θ'_{l_p} is not. In addition to reparameterizing for equality constraints, A also moves θ' parameters that are constrained by the null hypothesis to the top of the θ' vector and unconstrained parameters to the bottom to arrive at the partitioning of Result 1. Constructing A manually is straightforward for many problems (see the following example); a short computer program can construct A from an arbitrary list of constraints.

To obtain the information matrix $\mathbf{i}'$ of θ', let $\Sigma = \mathbf{i}^{-1}$ be the asymptotic variance-covariance matrix of $\hat{\theta}$. By Result 2, $\mathbf{A}\Sigma\mathbf{A}^T$ is the variance-covariance matrix of $\hat{\theta}$'. Since $\mathbf{i}'$ is the inverse of this variance-covariance matrix,

$$\mathbf{i}' = (\mathbf{A}^{-1})^T \, \mathbf{i} \, \mathbf{A}^{-1}. \tag{6}$$

3.1 Example: Three-Group Poisson

Three independent Poisson processes produce events at mean rates of $\theta_1 = 1$, $\theta_2 = 2$, and $\theta_3 = 3$ per day. For how many days must the processes be observed to have an 80% chance of detecting that the means are different at an overall significance level of 0.05?

We calculate the information matrix for a single Poisson observation. The probability that j events occur in a single observation from a Poisson process with mean μ is

$$f(j) = \frac{\mu^j}{j!} \exp -\mu,$$

so

$$\frac{\partial^2 \log(f(j))}{\partial \mu^2} = -\frac{j}{\mu^2}. \tag{7}$$

Since the number of events, j, is distributed Poisson the expectation of j is μ. Hence,

$$\mathbf{i} = \frac{1}{\mu}. \tag{8}$$

If the processes are observed for d days, the information matrix will be

$$\mathbf{i} = d \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/3 \end{pmatrix}.$$

By the methods of the previous section, $\theta_1' = \theta_1 - \theta_2$, $\theta_2' = \theta_2 - \theta_3$, and $\theta_3' = \theta_3$. Hence,

$$\mathbf{A} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}.$$

$\mathbf{i}'$ is calculated, and Result 1 is applied using the θ' values. In terms of θ',

$$\tilde{\psi} = (-1, -1)$$

$$\psi_0 = (0, 0).$$

The noncentrality parameter is $1.09d$. There are two constraints and thus two degrees of freedom for the χ^2. The upper 0.95 quantile of the central χ^2_2 distribution is 5.99. We must choose d to satisfy

$$\chi^2_2(5.99|1.09d) = 0.2,$$

from which $1.09d = 9.632$ or $d = 8.84$.

4 Comprehensive Examples

In many cases, specifying a family of alternative hypotheses is difficult because of the large number of parameters influencing power. Wherever possible, this number is reduced by imposing constraints. The first and third examples (Sections 4.1 and 4.3) illustrate this point.

4.1 Detecting a Quadratic Term in Logistic Regression

The model examined posits that the probability of an event at dose d, $p(\eta)$, is quadratic in the logistic link, i.e., the model is

$$p(\eta) = \frac{e^\eta}{1 + e^\eta},$$

where $\eta = a + bd + cd^2$. The experiment tests whether $c = 0$.

The case arose in radiobiology during an examination of whether regrowth of tissue is accelerated during a course of treatment; if so, the logistic model would contain a quadratic term, otherwise it would not. The three parameters of the model and the doses of d chosen (of which there must be at least three to allow the fit of a quadratic model) all contribute to the information matrix. A list of power values for many combinations of these six or more values would overwhelm the investigator.

To reduce the dimensionality of this problem, we will assume that the probability of response is known at two doses: $p = 0.1$ at dose d_1 and $p = 0.9$ at dose d_2. The dose scale can be shifted and scaled so that d_1 and d_2 are transformed to 0 and

1 respectively. Specifically, let $z = (d - d_1)/(d_2 - d_1)$, then $p(z = 0) = 0.1$ and $p(z = 1) = 0.9$. The linear part of the model, η, is still quadratic in z. In terms of this formulation, $a = \text{logit}(0.1) = -2.1972$ and $b + c = \text{logit}(0.9) - a = 4.3944$.

It is reasonable to assume that the probability of responses will increase with dose, which implies that $p(z)$ is monotone in d. Monotonicity implies that the derivative of the quadratic, $b + 2cz$, is positive at all values of z, which in turn implies both that b is positive and that $b+2c$ is positive. Monotonicity limits values of c to the range -4.3944 to 4.3944. Curves satisfying these constraints are shown in Figure 1.

Figure 1. Family of logistic transforms of quadratic curves for alternative hypotheses. The family was determined by requiring all curves to intersect at two points and by requiring monotonicity. The logistic curve without a quadratic term is the heavy curve in the center.

The design uses z values of 0, 0.5, and 1 with equal numbers of subjects at each of the three values.

The information matrix for this model is a special case of that for the general multivariate logistic model in which $\eta = \sum_{i=1}^{p} \theta_j x_j$.

By two applications of the chain rule,

$$\frac{\partial^2 \log(p(\eta))}{\partial\theta_i \partial\theta_j} = \frac{\partial^2 \log(p(\eta))}{\partial\eta^2} \frac{\partial\eta}{\partial\theta_i} \frac{\partial\eta}{\partial\theta_j} \tag{9}$$

because

$$\frac{\partial^2 \eta}{\partial\theta_i \partial\theta_j} = 0.$$

Since

$$\frac{\partial^2 \log(p(\eta))}{\partial\eta^2} = \frac{\partial^2 \log(1 - p(\eta))}{\partial\eta^2} = -p(\eta)\left[1 - p(\eta)\right], \tag{10}$$

the $(j.k)$ component of the information matrix is

$$p(u)\left[(1 - p(u)\right] x_j x_k$$

summed over all observations.

Computing the sample size required for a particular power at any value of c is straightforward. Interpretation of c may not be easy for the investigator, so sample size is presented as a function of the difference in proportion between the quadratic and linear logistic models at $z = 0.5$. The sample sizes needed to provide a power of 0.8 using a test with significance level 0.05 are shown in Figure 2.

4.2 Interaction/Synergy

A power determination was needed in planning a study of the effectiveness of behavioral counseling and nicotine replacement in cessation of smokeless tobacco use. The outcome variable is the proportion of subjects who have abstained from tobacco use for the month preceding the first anniversary of entering treatment. Estimates of the proportion who are expected to refrain under the four treatment

Figure 2. Sample size required to provide a power of 0.8 for detecting a quadratic term in the logistic regression. The abscissa values are the absolute difference between the proportion of responses predicted by the linear and quadratic models at $x = 0.5$.

regimes, shown in Table 1, were obtained as a weighted average of the results of several studies.

The natural model considers the probabilities of failing treatment, because to fail treatments A and B combined, one must first fail A and then fail B. If the probabilities of failure are independent, then the probability of joint failure is the product of the probabilities. The μs are the logarithms of these failure probabilities.

The probabilities of failure are calculated from the proportions in Table 1. The probability of failure with no treatment is $1 - 0.03 = 0.97$, so $\mu = \log(0.97)$. The probability of failure with behavioral treatment alone, 0.88, is the product of the probabilities of failure with no treatment and with behavioral treatment

Table 1: Proposed Smokeless Tobacco Trial

Treatment	N. Subjects	Propn. Quit	Model Terms
None	100	0.03	μ
Behavioral	100	0.12	$\mu + \mu_B$
Patch	100	0.12	$\mu + \mu_P$
Behavioral + Patch	100	0.25	$\mu + \mu_B + \mu_P + \mu_{BP}$

alone. Hence, the probability of failure with behavioral treatment alone is 0.907, so $\mu_B = \log(0.0907)$. Similarly, the probability of failure with the nicotine patch alone is 0.907 and $\mu_P = \log(0.907)$. Finally, the probability of failure with both treatments is the failure rate for both treatments, 0.75, divided by the product of the probabilities of failure with no treatment, behavioral counseling alone, and the patch alone. This probability is 0.940; $\mu_{BP} = \log(0.940)$.

The log-linear model posits that

$$p(u) = \exp(u),$$

where $p(u)$ is the probability of treatment failure and

$$u = \mu x + \mu_B x_B + \mu_P x_P + \mu_{BP} x_{BP}.$$

The x variables are zero or one-valued indicators of effects. Thus, x is identically 1, and x_B is 1 when behavioral therapy is applied, i.e., the second and fourth treatment. The indicator x_{BP} is 1 only if both behavioral therapy and the patch are applied, i.e., treatment four. A discussion of parameterization is found in Sections 3.1 through 3.5 of McCullagh and Nelder (1989).

The expected information matrix is readily obtained. If the observation is a treatment failure, then the contribution to the log-likelihood is u and the second derivative with respect to u is zero. If the observation is a treatment success, then

the contribution to the log-likelihood is $\ln(1 - \exp(u))$, and the second derivative of this expression with respect to u is $-p(u)/(1-p(u))^2$. Consequently, by (9), the component of the information matrix corresponding to μ_i and μ_j for subscripts i and j is

$$\frac{p(u)}{(1-p(u))} x_i x_j.$$

With 100 subjects in each of the four treatment groups, using the effects as estimated, the probability of detection of the interaction, μ_{BP}, is only 0.123 using a significance level of 0.05. Only when the success rate of behavior and patch jointly applied rises to 39% do we have a 0.8 probability of showing the reality of the interaction.

4.3 Effect of a Covariate on Time to Event

The use of biological markers to predict treatment outcome in cancer receives a great deal of attention. We examine the size of a study necessary to detect the effect of a marker in a time-to-event study using a parametric proportional hazards model as descibed in Kalbfleish and Prentice (1980)

The time might be the disease-free period or the time to cancer progression. For simplicity, we term the event "death" and speak of survival time. The time is modeled as a function of the covariate, x, by the proportional hazards assumption. The hazard at time t for covariate value x is

$$h(t|x) = \exp(a + bx) h(t|0),$$

where $h(t|0)$ is the hazard at covariate value 0. For brevity, let w denote $\exp(a+bx)$. We consider the simplest case in which $h(t|0)$ takes the constant value 1 for all t. Survival for fixed x will thus be exponential. The case of positive values for b is considered, so hazard will increase and survival will decrease as x increases.

The distribution of x is assumed to be normal; by translating and rescaling, x can be assumed to have mean 0 and standard deviation 1.

The proposed study will last L years and during this period will accrue A subjects. Each subject will enter the study at a random time between 0 and L, so the subject's follow-up time, U, will be uniformly distributed between 0 and L.

For a subject with covariate x and follow-up time U, there are two possible outcomes: the subject can die at time t between 0 and U, or the subject can be withdrawn alive at time U.

First, consider the case of death. The density of time to event for an exponential distribution with hazard w is

$$w \exp(-wt), \tag{11}$$

and the logarithm of this value is the contribution to the log-likelihood of a death at time t. The matrix of second derivatives of the log-likelihood is

$$-wt \begin{pmatrix} 1 & x \\ x & x^2 \end{pmatrix}, \tag{12}$$

where the upper-left entry of the matrix is the second derivative with respect to a twice, the lower-right entry is with respect to b twice, and the off-diagonals are the cross derivatives.

To obtain the expected contribution of the failure to the information, each element of the matrix in (12) must be integrated twice. First, the element is multiplied by the probability of death at t from (11), and then it is integrated with respect to t from 0 to U. Since U is uniformly distributed on 0 to L, this result is then integrated over U from 0 to L and multiplied by $1/L$ (the uniform density). The result, negated to be a contribution to the information matrix, is

$$\frac{\exp(-wL)(wL+2)+wL-2}{wL} \begin{pmatrix} 1 & x \\ x & x^2 \end{pmatrix}. \tag{13}$$

In the second case failure does not occur, and the subject will be censored at the end of the study, i.e., at time U after his entry. The probability of this is

$$S(U|x) = \exp(-wU).$$

The logarithm of this expression is the contribution of the censored observation to the log-likelihood. To calculate the contribution to the information matrix the contribution to the log-likelihood is differentiated twice with respect to a and b. These derivatives are multiplied by $S(U|x)$, and the result is integrated over U from 0 to L then multiplied by $1/L$. The negated result is

$$-\frac{\exp(-wL)(wL+1)-1}{wL}\begin{pmatrix} 1 & x \\ x & x^2 \end{pmatrix}. \tag{14}$$

The sum of (13) and (14) is the expected information given x. This must be multiplied by the normal density for x and integrated over x. The integral was numerically evaluated over x from -4 to 4 using a 10-point Gaussian formula.

Calculations are performed for a 3-year study, i.e., $L = 3$. To reduce the dimensionality of the problem from two parameters to one, we suppose that 70% of all patients fail by 3 years. Then,

$$\int_{-\infty}^{\infty} \exp(-3e^{(a+bx)})\mathrm{dnor}(x)dx = 0.3,$$

where $\mathrm{dnor}(x)$ is the unit normal density at x. This can be solved numerically to obtain a as a function of b as shown in Figure 3.

The results of these calculations could be presented in several ways. The obvious presentation shows the requisite sample size for a particular power as a function of b. The investigator may not, however, have an intuitive grasp of the import of b. Hence, we present sample size as a function of the 3-year expected survival of a subject whose x-value is at the upper-quartile; this survival figure is a monotone function of b. The results appear in Figure 4.

A sample size of 209 is required for an 80% probability of detecting a nonzero b if the expected survival of the upper-quartile observation is 0.22. This survival corresponds to $b = 0.3$. A sample size of 40 is required when the expected survival is 0.13, corresponding to $b = 0.7$.

The reader may notice a lack of realism in this example. Calculations have been made for a fixed sample size, whereas in most trials patients are accrued as they

Figure 3. Plot of a (intercept of proportional hazards model) versus b (coefficient of covariate). The relation between a and b was determined by the condition that overall 3-year survival was 30%.

appear and thus the sample size is random. To accommodate this consideration, the number of patients accruing to the trial can be considered Poisson. Power can be calculated for various numbers of subjects near the center of the Poisson distribution and these power values can be averaged with Poisson weights.

5 Implementation

Likelihood ratio, power, and sample size calculations should be implemented in a high-level scientific/statistical language such as Gauss, Matlab, S, or SAS, since these languages allow easy incorporation of additional models. Adding C or Fortran

Figure 4. Sample size required for 0.8 power for a test with significance level 0.05 as a function of the three year expected survival of a subject whose covariate is at the upper-quartile. The null value (corresponding to $b = 0$ is 0.30).

code to a package is more difficult.

Our implementation is in S. It contains several routines that calculate the information matrix for particular generalized linear models. One program calculates the noncentrality parameter from the information matrix, the alternative hypothesis parameter values, and a list of constraints defining the null hypothesis. Other routines calculate sample size or power from the noncentrality parameter and the number of constraints.

Computing the second derivatives of the log-likelihood can be tedious. Symbolic manipulation programs, such as MacSyma, Maple, Mathematica, or Reduce accurately perform analytic differentiation. If analytic forms of the information matrix

exist, the same tools can also perform the summation or integration required to obtain the expectation of the second derivatives. Several of these systems will write C or Fortran code that is readily converted to the scientific/statistical computer language.

If the outcome of the trial is continuous, calculation of the information matrix requires integration, possibly over several dimensions. The proportional hazards example above required three nested integrals. Because the underlying hazard was chosen to be constant in the example, two of the integrals could be evaluated analytically. Were a more complex form of hazard chosen, all three integrations would be evaluated numerically. Numeric multi-dimensional integration is not standard in all scientific/statistical languages, and interested readers might investigate Berntsen (1991). Fortran code for this algorithm is available on netlib.

Computations involving the noncentral χ^2 may also not be available in the scientific/statistical language. A recent review by Boomsma and Molenaar (1994) compares four commercial products that perform these calculations. Routines are also available by anonymous ftp from the current authors, both as Fortran and C subroutines and as a stand-alone package.

Finding a zero of a monotone function is used to handle constraints on model parameters. For this purpose, we use Algorithm R of Bus and Dekker (1975).

The authors' code can be obtained by anonymous ftp to odin.mda.uth.tmc.edu (129.106.3.17), and it will be submitted to statlib. An index to routines available by ftp is in file ./pub/index, which can also be obtained by executing the Unix command
'finger software@odin.mda.uth.tmc.edu'.

References

Berntsen, J., Espelid, T. O., and Genz, A. (1991), "Algorithm 698: DCUHRE – An Adaptive Multidimensional Integration Routine for a Vector of Inte-

grals," *ACM Transactions on Mathematical Software*, 17, 452-456.

Bus, J. C. and Dekker, T. J. (1975), "Two Efficient Algorithms with Guaranteed Convergence for Finding a Zero of a Function," *ACM Transactions on Mathematical Software*, 1, 330-345.

Boomsma, A. and Molenaar, I. W. (1994), "Four Electronic Tables for Probability Distributions," *The American Statistician*, 48, 153-162.

Cox, D.R. and Hinkley, D.V. (1974), *Theoretical Statistics.* London: Chapman and Hall.

Heitjan, D. (1993), "Power Package for S." To obtain this package, send e-mail to statlib@lib.stat.cmu whose message is 'send power from s'.

Kalbfleisch, J.D. and Prentice, R.L. (1980), *The Statistical Analysis of Failure Time Data,* New York: John Wiley & Sons.

McCullagh, P. and Nelder, J.A. (1989), *Generalized Linear Models: Second Edition.* London: Chapman and Hall.

Muller, K.E. and Peterson, B.L. (1984), "Practical Methods for Computing Power in Testing the Multivariate General Linear Hypothesis," *Computational Statistics and Data Analysis*, 2, 143-158.

O'Brien, R. G. (1986a), "Power Analysis for Linear Models," *Proceedings of the Eleventh Annual SAS Users Group International Conference*, Cary, NC: SAS Institute, 915-922.

O'Brien, R. G. (1986b), "Using the SAS System to Perform Power Analyses for Log-linear Models," *Proceedings of the Eleventh Annual SAS Users Group International Conference*, Cary, NC: SAS Institute, 778-784.

Self, S. G. and Mauritsen, R. H. (1988), "Power Calculations for Likelihood Ratio Tests in Generalized Linear Models," *Biometrics*, 48, 31-39.

Self, S. G., Mauritsen, R. H., and Ohara, J. (1992), "Power Calculations for

Likelihood Ratio Tests in Generalized Linear Models," *Biometrics*, 48, 31-39.

Tong, Y. L. (1990), *The Multivariate Normal Distribution,* New York: Springer-Verlag.

Wilks, S. S. (1938), "The Large-sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses," *Annals of Mathematical Statistics*, 9, 60-62.