IMPERIAL COLLEGE OF SCIENCE AND TECHNOLOGY

Department of Mathematics

TRANSFORMATIONS TO ADDITIVITY FOR BINARY DATA

by

Francisco Javier Aranda-Ordaz

Thesis submitted for the degree of

Doctor of Philosophy in the University of London

and the

Diploma of Membership of Imperial College

July, 1980

A mi esposa, mi madre y mi hermano

ABSTRACT

Binary response data occur when an observation takes one of two possible forms, e.g. success or failure. The main objective is to study how the probability of success depends on explanatory variables. One of the most common methods of analyzing such data is to fit transforms of the probabilities by linear functions of parameters.

Families of power transformations for the probabilities are considered. Three families are proposed. One treats successes and failures symmetrically, while the other two treat them asymmetrically. It is suggested that a suitable scale for a particular linear model is estimated by maximum likelihood methods. The concept of no interaction for "uni-response, multi-factor" experimental situations is related to the choice of a particular scale within the symmetric family. Two new tests for symmetric or asymmetric departures from the logistic model are proposed. The new methods are applied to several examples and comparisons with the results of previous analyses performed.

Extensions for polytomous and multivariate binary responses are outlined.

A family of transformations for probabilities is considered for the analysis of grouped survival data. Additive and multiplicative models for the hazard function are compared. A method is suggested for estimating the scale for which an additive representation of the hazard in terms of explanatory variables is appropriate. A new test for departures from the grouped proportional hazards model is proposed. Several examples are analyzed using the new methods.

ACKNOWLEDGEMENTS

I am deeply indebted to my supervisor, Professor D.R. Cox, for suggesting the topic of research and for the invaluable insights he has provided on this and many other subjects. His encouragement and patient guidance of this work are very much appreciated.

Financial support has been provided by a postgraduate scholarship from the Consejo Nacional de Ciencia y Tecnologia, Mexico and the Universidad Nacional Autonoma de Mexico (Programa de Superacion del Personal Academico). I would like to record my gratitude to both institutions for their assistance.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Chapter 1: INTRODUCTION

The use of transformations for the analysis of response data is a well established method in Statistics. The usual aim is to simplify the analysis by making applicable some standard technique which, without transformation, would not be appropriate. Transformations may be applied directly to data or indirectly via parameters used in probabilistic models. For quantitative data the most widely used transformations are powers (and the logarithm). For (0,1) data any such transformation has to be applied to the underlying parameters, i.e. probabilities, rather than to the data directly. This thesis considers families of such transformations for the analysis of binary data.

To determine the general type of transformation in a family, certain characteristics must be required from its members, e.g. for quantitative data the achievement of additivity, normality and homoscedasticity are ones commonly used. For binary data we aim to achieve additivity of representation in terms of the explanatory variables. The other two objectives are not achievable and perhaps even not desirable, because of the character of the data.

Three families of transformations for probabilities are introduced in Chapter 2. They have in common the inclusion of the logistic transformation as a simple special case. One of these families consists of transformations which treat symmetrically successes and failures. It includes also a linear transformation and approximations to probit and arcsine. The other two consist of transformations which treat successes and failures asymmetrically. They include the complementary log log and complementary log among their members.

Chapter 3 is concerned with the use of two of the families in particular to detect departures from the logistic model. The symmetric family is used to detect departures in the direction of the linear transformation and to study the relation between the choice of a certain scale for the representation of the data and definitions of no interaction between the explanatory variables. A new statistic is introduced to test for departures from the multiplicative effects assumption, implicit in the use of the logistic model, in the direction of the additive effects assumption associated with a linear model in the probabilities themselves. An asymmetric family is used to detect departures in the direction of skewed alternatives, in special the complementary log log transformation. A new statistic is also introduced to carry out a formal test.

In Chapter 4, methods based on the symmetric family introduced in Chapter 2 are applied to several sets of data. The most appropriate scale within the family, which permits a simple representation consistent with the data, is estimated by maximum likelihood methods. In a first stage a screening method, based on the maximization of the log likelihood function given a model configuration, is used. Further analysis may be performed using the GLIM computing package. The test for departures from the logistic model introduced in §3.3 is applied to one example, the result agrees with preliminary findings.

Chapter 5 considers polytomous and multivariate binary responses. An extension of the symmetric family is introduced for polytomous unordered data. For multivariate binary responses a simple extension is suggested.

Chapter 6 deals with survival data. The aim is to provide a general model for the hazard function that includes the multiplicative

and additive versions as special cases. The approach is to analyze sequences of contingency tables obtained by grouping continuous data. Then, the ideas of the first part of the thesis are applicable.
A subfamily of an asymmetric family defined in Chapter 2 is used to construct a useful comprehensive parametric model. A test statistic to detect departures from the model with proportional hazards in the direction of one with additive hazards is introduced. A scale where an additive model is appropriate for the data is estimated similarly to Chapter 4. The methods are applied to some examples.

Chapter 2: A FAMILY OF SYMMETRIC TRANSFORMATIONS FOR PROBABILITIES

2.1 Introduction

To analyze the dependence of binary response data on explanatory variables it is common to fit transforms of the probabilities by linear functions of parameters. The most important examples are linear, probit, arcsine and logistic models. The last has in particular the theoretical advantage of corresponding to the exponential family natural linear model for the binomial distribution. Also all real values for the transform are meaningful, so that there are no inevitable constraints on the model. For these and other reasons it is probably the most commonly used model for analyzing this type of data. However, as for all models, it is tentative and therefore some consideration of adequacy is needed. It is important to be aware if some nonlogistic model gives a simpler or better fit. If we can find a procedure which detects inadequacy, and that also indicates the kind of desirable modification to the model, this is of potential usefulness. A final choice of model must, however, depend partly on the ease with which the conclusions can be presented and understood.

One possibility is to compare the fit of the logistic model with that of the linear model; we shall discuss in Chapter 3 the implications of this choice. An appealing and informative way to achieve this objective is to construct an extended model which reduces to the linear and logistic models as special cases. We consider first transformations that are symmetric, in the sense of leading to essentially the same answers if successes and failures are interchanged; these include the logistic and linear transformations.

Throughout we deal with situations involving several non homogeneous sets of data. Then the basic objective is to achieve a simple summary of the variation between sets by means of additive models for the probabilities, on a scale determined by a transformation.

2.2 The general member of the family and its properties

The guidelines used to determine the form of the general member of the family of transformations were essentially

(1) to find a simple expression, depending on one or two parameters (besides those in the linear component), which reduces as special cases to the logistic and linear transformations, and

(2) to aim that the probabilities can be expressed simply in terms of the transformed values, and vice versa.

The last condition was imposed so that maximum likelihood estimation would be reasonably simple. It is not essential; see for example the solution given by Fisher (1935) to estimation in the probit model. Nevertheless this allows a computationally more flexible approach to the problem.

A transformation which fulfils the above requirements is

$$T_\lambda(\theta) = \frac{2}{\lambda} \frac{\theta^\lambda - (1-\theta)^\lambda}{\theta^\lambda + (1-\theta)^\lambda} , \qquad (2.2.1)$$

where θ denotes the probability of success, and λ denotes the transformation parameter, initially assumed to be unconstrained.

We denote the family of transformations with general form in (2.2.1) by

$$\mathcal{T} = \{T_\lambda\}_{\lambda \in S} \qquad S \subset \mathbb{R} .$$

Two important simple features of the transformation are that

P1) $$T_\lambda(\theta) = T_{-\lambda}(\theta),$$

and that

P2) $$T_\lambda(\theta) = -T_\lambda(1-\theta),$$

i.e. T_λ treats successes and failures in a symmetrical way. We call $\mathcal{T}$ a symmetric family.

Further

P3) $$T_\lambda(\theta_1) < T_\lambda(\theta_2) \qquad \lambda < \infty,$$

if $\theta_1 < \theta_2$, i.e. T_λ is monotonically increasing in θ.

Expression (2.2.1) reduces to the logistic transformation, in the limit when $\lambda = 0$, and to a linear transformation when $\lambda = 1$. Inverting (2.2.1) we obtain

$$\theta = \begin{cases} 0 & \text{if } \lambda\tau/2 \leq -1, \\ \dfrac{(1 + \lambda\tau/2)^{1/\lambda}}{(1+\lambda\tau/2)^{1/\lambda} + (1-\lambda\tau/2)^{1/\lambda}} & \text{if } |\lambda\tau/2| < 1, \\ 1 & \text{if } \lambda\tau/2 \geq 1, \end{cases} \qquad (2.2.2)$$

where τ denotes a value in the image of $T_\lambda(\cdot)$.

For our purposes, τ will be assumed to have a linear expression in terms of some parameters associated with the explanatory variables considered in a specific situation. We obtain in this way a comprehensive model for transformed probabilities. This model includes as special cases the logistic and linear ones.

If we fit by maximum likelihood a linear model for τ for a range of values of λ, we can consider the maximized log likelihood as a function of λ and hence derive not only the maximum likelihood estimate $\hat{\lambda}$, but also determine which values of λ provide an acceptable fit.

We assume that the observations are independent and that the probability of success is homogeneous within sets. Hence, the situation may be considered as one with m sets of independent binomially, $B(n_i,\theta)$, distributed observations.

2.3 Relationship with generalized linear models

We may recast the discussion in the last section within the context of generalized linear models (Nelder and Wedderburn, 1972), hereafter denoted by GLM, straightforwardly. To specify a GLM we need to identify its three components, namely the error structure of the data, the linear systematic part, or linear predictor, of the model and the linking function. These models are defined for members of the exponential family of distributions so the usual terminology for this family is used.

For the situation in which we are interested the data are represented by $\underset{\sim}{r} = (r_1,\ldots,r_m)$. The r_i $(i = 1,\ldots,m)$ are assumed independent random variables with means μ_i $(i = 1,\ldots,m)$. The components of the GLM are as follows:

(i) The error structure. The data follow binomial distributions $B(n_i;\theta_i)$ $(i = 1,\ldots,m)$ with moment parameter $\mu_i = n_i\theta_i$ where

$$\theta = \exp(\phi)\{1 + \exp(\phi)\}^{-1} ,$$

and ϕ denotes the natural parameter of the distribution.

(ii) The linear systematic part of the model has the form

$$\underset{\sim}{\tau} = X\underset{\sim}{\beta} ,$$

where $\underset{\sim}{\beta}$ is a vector of unknown parameters, and X is a known matrix determined by the values of the explanatory variables.

(iii) The linking function is defined in terms of the transformation in (2.2.1) as follows

$$\mu_i = n_i\, T_\lambda^{-1}(\tau_i) \qquad (i = 1,\ldots,m),$$

or explicitly

$$\mu_i = \begin{cases} 0 & \text{if } \lambda\tau_i/2 \le -1, \\ \dfrac{n_i(1+\lambda\tau_i/2)^{1/\lambda}}{(1+\lambda\tau_i/2)^{1/\lambda} + (1-\lambda\tau_i/2)^{1/\lambda}} & \text{if } |\lambda\tau_i/2| < 1, \\ 1 & \text{if } \lambda\tau_i/2 \ge 1, \end{cases}$$

where $\tau_i = \underset{\sim}{x}_i\underset{\sim}{\beta}$, $\underset{\sim}{x}_i$ is the ith row of X.

In this way the relationship between the mean of the ith random variable and its linear predictor is established. The linking function may be defined equivalently in terms of the natural parameter as follows:

$$\phi_i = \frac{2}{\lambda} \tanh^{-1} (\lambda\tau_i/2) \qquad |\lambda\tau_i/2| < 1.$$

For GLM estimation by maximum likelihood can be regarded as a form of the iterative weighted least squares procedure with weights

$$w_i = (d\mu/d\tau)^2/V \qquad (i = 1,\ldots,m);$$

in this particular case

$$d\mu/d\tau = [\mu^\lambda + (n-\mu)^\lambda]^2/[4n\mu^{\lambda-1}(n-\mu)^{\lambda-1}],$$

and

$$V = \mu(n-\mu)/n, \quad \text{the variance of } r.$$

This is just an application of Fisher's general method of efficient scores. It is a modified Newton-Raphson process for the solution of the likelihood equations, where expected rather than observed second order derivatives of the log likelihood function are used. Nelder and Wedderburn (1972) provide a complete description of the estimation method.

Then, the proposed model may be fitted using the GLIM package (GLIM3 release; Baker and Nelder, 1978) for fixed values of the transformation parameter λ.

2.4 Approximation to other transformations

By definition the family $\mathcal{T}$ includes as members the logistic and linear transformations. However, there are other transformations quite often used in practice and which treat symmetrically successes and failures. Perhaps the most important among these are the probit and arcsine. A natural question to examine is if there are members of $\mathcal{T}$ which approximate closely those transformations.

To carry out the search we need to define the transformations suitably. Consider the normit and the standardized arcsine (sinit) defined as follows:

i) normit $(\theta) = \Phi^{-1}(\theta)$, where $\Phi^{-1}(\cdot)$ denotes the inverse of the standard normal probability integral,

ii) sinit $(\theta) = \sin^{-1}\sqrt{\theta} - \pi/4$.

Probits are just normits plus a constant, defined to avoid negative transformed values.

To compare with T_λ we need to choose a scaling constant for each transformation. We examine two possibilities. The first is to force identity of tangents of the inverse transformation at the 50% point. This seems best when interest is on the central part of the range of proportions. A second possibility is to achieve agreement at some specific point in the extremes of the range of θ, say at $\theta = 0.8$; this is better when the interest is more in the extreme probabilities. Of course there are many other possibilities, e.g. the minimization of some suitable defined measure of distance between functions.

Without loss of generality, we assume that the 50% point is at the origin on the transformed scale. The functions to be approximated may be expressed as follows:

$$\Phi(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{x} e^{-v^2/2}\, dv \quad , \tag{2.4.1}$$

$$\text{and } \zeta(x) = \begin{cases} 0 & \text{if } x < -\pi/4, \\ \sin^2(x + \pi/4) & \text{if } |x| \leq \pi/4, \\ 1 & \text{if } x > \pi/4 . \end{cases} \tag{2.4.2}$$

Analogously the approximating function is

$$T^{-1}(t) = \begin{cases} 0 & \text{if } \lambda t/2 \leq -1, \\ \dfrac{(1+\lambda t/2)^{1/\lambda}}{(1+\lambda t/2)^{1/\lambda} + (1-\lambda t/2)^{1/\lambda}} & \text{if } |\lambda t/2| < 1, \\ 1 & \text{if } \lambda t/2 \geq 1, \end{cases} \tag{2.4.3}$$

where $t = ax$ and a is a constant to be chosen.

For the criterion of equality of tangents, a is equal to 1.5958 and 4 for normit and sinit respectively. These values are independent of λ. Table 2.1 shows the results obtained in this case. The values of λ were selected to minimize, (i) the sum of squared deviations (L^2) and (ii) the largest absolute deviation (L^0). The contact of the approximation at the origin is actually of second order because not only the tangents coincide but also, by symmetry, the second order derivatives. This means that the associated curves cross each other at the origin but remain close for a wide range about it.

For the criterion of agreement at a specified point, the comparability constants depend on the value at which agreement is required, say θ_0, and λ. They all have the following form

$$a = \frac{2}{\lambda x_0} \frac{1 - (1/\theta_0 - 1)^{\lambda}}{1 + (1/\theta_0 - 1)^{\lambda}} ,$$

where x_0 is the transformed value for θ_0.

For instance, for the normit at $\theta_0 = 0.8$, the constant a is equal to 1.6101. The selection of λ is also made according to the deviation criteria L^2 and L^0. Table 2.2 shows the λ's that provide the best approximation for several points of agreement, for each transformation and approximation criterion.

Tables 2.3 and 2.4 show the probability value for several points on the transformed scale for the normit and its approximations according to the criteria L^2 and L^0, respectively. Tables 2.5 and 2.6 show similar computations for sinit.

We observe on tables 2.1 and 2.2 a good degree of stability of the parameter λ around 0.38 for the normit and 0.67 for the sinit. Hence, these values may be used to characterize normit and sinit within the family $\mathcal{T}$.

The arrangement of the values of λ for the logistic, normit, sinit and linear transformations, in that order, within $\mathcal{T}$ is interesting. This agrees with the speed at which each transformation approaches its limits.

The similarity among the transformed values for probabilities in the range 0.25-0.75 over the different scales is striking. Even for scales as different as the logistic and linear those values are very similar. The explanation is the approximate linearity of the relationship between the probability, θ, and its logit. If we denote by ϕ the logit of θ we have that

$$\phi \doteq 4\theta - 2, \qquad (2.4.4)$$

for $\theta \in [0.25, 0.75]$. The expression above may be obtained expanding logit(θ) in Taylor series about $\theta = 1/2$ and neglecting terms of order higher than two.

To show the degree of approximation between T_0 and T_1 for $\theta \in [0.25, 0.75]$, we computed the logit and the approximate value for some θ's in the range (0, 0.5). These results are shown in Table 2.7.

TABLE 2.1

Values of the transformation parameter (λ) for approximations under L^2 and L^0 with contact of second order at the origin. The largest relative absolute deviation from the actual probability (d)* and the probability point at which it is observed (p) are also shown.

Transformation	λ	d	p
Normit	0.3955	0.0031	0.8751
	0.3869	0.0036	0.8851
Sinit	0.6755	0.0024	0.9938
	0.6698	0.0028	0.9020

First row corresponds to L^2, second to L^0.

* observed in tabulated values.

TABLE 2.2

Values of the transformation parameter (λ) for approximations under L^2 and L^0 with agreement at a specific point (θ_0) along the range for θ. The largest relative absolute deviation from the actual probability (d)* and the probability point at which it is observed (p) are also shown.

θ_0	Normit			Sinit		
	λ	d	p	λ	d	p
0.7	0.3893	0.0026	0.8851	0.6713	0.0022	0.9965
	0.3820	0.0029	0.8851	0.6663	0.0023	0.9045
0.8	0.3800	0.0018	0.9192	0.6650	0.0019	0.9965
	0.3746	0.0021	0.9332	0.6610	0.0017	0.9263
0.9	0.3663	0.0039	0.6915	0.6597	0.0027	0.6731
	0.3590	0.0043	0.6915	0.6519	0.0033	0.6913
0.95	0.3752	0.0040	0.6915	0.6670	0.0023	0.9938
	0.3513	0.0066	0.6915	0.6491	0.0045	0.7093

First row corresponds to L^2, second to L^0.

* observed in tabulated values.

TABLE 2.3

Comparison of normit and two approximations under L^2, namely (ET) forcing equality of tangents at the origin, and (A) with agreement at $\theta_0 = 0.8$, denoted below by θ^* and θ^{**} respectively.

Normit scale	Probability θ	ET		A	
		θ^*	$\theta-\theta^*$	θ^{**}	$\theta-\theta^{**}$
0	.5000	.5000	0	.5000	0
.200	.5793	.5792	.0000	.5799	-.0007
.400	.6554	.6551	.0003	.6564	-.0010
.600	.7258	.7250	.0008	.7265	-.0008
.800	.7882	.7866	.0016	.7883	-.0001
1.000	.8415	.8392	.0023	.8407	.0008
1.200	.8851	.8824	.0027	.8836	.0016
1.400	.9192	.9168	.0025	.9175	.0017
1.600	.9452	.9433	.0019	.9436	.0016
1.800	.9641	.9630	.0011	.9630	.0011
2.000	.9772	.9772	.0000	.9769	.0003
2.200	.9861	.9870	-.0009	.9866	-.0005
2.400	.9918	.9933	-.0015	.9929	-.0011
2.600	.9953	.9971	-.0018	.9967	-.0014
2.800	.9974	.9991	-.0017	.9988	-.0014
3.000	.9987	.9999	-.0012	.9997	-.0011
3.200	.9993	1.0000	-.0007	1.0000	-.0007
3.400	.9997	1.0000	-.0003	1.0000	-.0003
3.600	.9998	1.0000	-.0002	1.0000	-.0002
3.800	.9999	1.0000	-.0001	1.0000	-.0001
4.000	1.0000	1.0000	-.0000	1.0000	-.0000

TABLE 2.4

Comparison of normit and two approximations under L^0, namely (ET) forcing equality of tangents at the origin, and (A) with agreement at $\theta_0 = 0.8$, denoted below by θ^* and θ^{**} respectively.

Normit scale	Probability θ	ET		A	
		θ^*	$\theta-\theta^*$	θ^{**}	$\theta-\theta^{**}$
0	.5000	.5000	0	.5000	0
.200	.5793	.5792	.0000	.5800	-.0007
.400	.6554	.6551	.0003	.6565	-.0010
.600	.7258	.7248	.0009	.7266	-.0008
.800	.7882	.7864	.0018	.7883	-.0001
1.000	.8415	.8388	.0026	.8406	.0008
1.200	.8851	.8819	.0032	.8834	.0017
1.400	.9192	.9162	.0031	.9173	.0019
1.600	.9452	.9426	.0026	.9433	.0019
1.800	.9641	.9623	.0018	.9627	.0014
2.000	.9772	.9765	.0008	.9766	.0007
2.200	.9861	.9863	-.0002	.9862	-.0001
2.400	.9918	.9928	-.0010	.9926	-.0008
2.600	.9953	.9967	-.0014	.9965	-.0011
2.800	.9974	.9989	-.0014	.9987	-.0012
3.000	.9987	.9998	-.0011	.9997	-.0010
3.200	.9993	1.0000	-.0007	1.0000	-.0007
3.400	.9997	1.0000	-.0003	1.0000	-.0003
3.600	.9998	1.0000	-.0002	1.0000	-.0002
3.800	.9999	1.0000	-.0001	1.0000	-.0001
4.000	1.0000	1.0000	-.0000	1.0000	-.0000

TABLE 2.5

Comparison of sinit and two approximations under L^2, namely (ET) forcing equality of tangents at the origin and (A) with agreement at $\theta_0 = 0.8$, denoted below by θ^* and θ^{**} respectively.

Sinit scale	Probability θ	ET		A	
		θ*	θ-θ*	θ**	θ-θ**
0	.5000	.5000	0	.5000	0
.039	.5392	.5392	.0000	.5395	-.0003
.079	.5782	.5782	.0000	.5787	-.0005
.118	.6167	.6166	.0001	.6174	-.0007
.157	.6545	.6543	.0002	.6552	-.0007
.196	.6913	.6910	.0004	.6921	-.0007
.236	.7270	.7264	.0006	.7276	-.0006
.275	.7612	.7604	.0009	.7616	-.0004
.314	.7939	.7927	.0012	.7940	-.0001
.353	.8247	.8233	.0014	.8244	.0003
.393	.8536	.8519	.0017	.8529	.0006
.432	.8802	.8784	.0018	.8792	.0010
.471	.9045	.9027	.0018	.9033	.0012
.511	.9263	.9248	.0016	.9251	.0012
.550	.9455	.9444	.0011	.9445	.0010
.589	.9619	.9615	.0004	.9613	.0006
.628	.9755	.9760	-.0005	.9756	-.0000
.668	.9862	.9877	-.0016	.9871	-.0009
.707	.9938	.9963	-.0024	.9955	-.0017
.746	.9985	1.0000	-.0015	1.0000	-.0015
.785	1.0000	1.0000	0	1.0000	0

TABLE 2.6

Comparison of sinit and two approximations under L^0, namely (ET) forcing equality of tangents at the origin, and (A) with agreement at $\theta_0 = 0.8$, denoted below by θ^* and θ^{**} respectively.

Sinit scale	Probability θ	ET		A	
		θ*	θ-θ*	θ**	θ-θ**
0	.5000	.5000	0	.5000	0
.039	.5392	.5392	.0000	.5395	-.0003
.079	.5782	.5782	.0000	.5788	-.0005
.118	.6167	.6166	.0001	.6175	-.0007
.157	.6545	.6543	.0002	.6553	-.0008
.196	.6913	.6909	.0004	.6922	-.0008
.236	.7270	.7263	.0007	.7277	-.0007
.275	.7612	.7602	.0011	.7617	-.0004
.314	.7939	.7925	.0014	.7940	-.0001
.353	.8247	.8229	.0018	.8244	.0003
.393	.8536	.8514	.0021	.8528	.0008
.432	.8802	.8778	.0024	.8791	.0011
.471	.9045	.9020	.0025	.9031	.0014
.511	.9263	.9240	.0024	.9248	.0016
.550	.9455	.9435	.0020	.9440	.0015
.589	.9619	.9605	.0014	.9608	.0011
.628	.9755	.9750	.0005	.9750	.0005
.668	.9862	.9867	-.0006	.9865	-.0003
.707	.9938	.9954	-.0016	.9950	-.0012
.746	.9985	1.0000	-.0015	.9998	-.0014
.785	1.0000	1.0000	0	1.0000	0

TABLE 2.7

Approximation to the logit by a linear transformation

θ	$\lvert logit \rvert$	$\lvert 4\theta-2 \rvert$	difference
.1 (.9)	2.1972	1.600	0.5972
.15 (.85)	1.7346	1.400	0.3346
.2 (.8)	1.3863	1.200	0.1863
.25 (.75)	1.0986	1.000	0.0986
.3 (.7)	0.8473	0.800	0.0473
.35 (.65)	0.6190	0.600	0.0190
.4 (.6)	0.4055	0.400	0.0055
.45 (.55)	0.2007	0.200	0.0007

2.5 Extensions

The family $\mathcal{T}$ proposed above treats successes and failures symmetrically. However, there are occasions where it is desirable to treat them asymmetrically. For instance, Yates (1955) gives some examples where, for theoretical reasons, expression on an asymmetric scale is called for; see especially the discussion of the Drosophila data. In general this will be the case where there is some connection with extreme value problems. Hence, it is useful to find a family of transformations for the probabilities that includes some of the transformations usually employed in asymmetric situations.

An explicit expression of the probability parameter θ in terms of the transformed values is required. In defining the family, we want the complementary log log and the logistic transformations to be special members of it. The reason for this is that the complementary log log model provides a suitable alternative to the logistic one among the asymmetric transformations.

The general form of a possible candidate is as follows

$$V_{\lambda,\beta}(\theta) = \frac{[\log(\frac{\theta}{1-\theta} + \beta)]^{\lambda} - 1}{\lambda} , \qquad (2.5.1)$$

where $-\infty < \lambda < \infty$ and $0 \leq \beta$.

If we equate (2.5.1) to a linear function of unknown parameters, denoted by τ, we obtain

$$V_{\lambda,\beta}(\theta) = \tau \quad . \qquad (2.5.2)$$

Expression (2.5.2) reduces for certain values of the parameters λ and β, to some well known transformations, as follows.

(i) $\beta = 0$, $\lambda = 1$. Here

$$\log \theta/(1-\theta) - 1 = \tau ,$$

which is essentially the logistic transformation, so long as τ includes a constant term, which is usually the case, and is assumed through this section.

(ii) $\beta = 1$, $\lambda = 0$. Here

$$\log \log [1/(1-\theta)] = \tau,$$

or equivalently $\log[-\log(1-\theta)] = \tau$,

the complementary log log transformation.

(iii) $\beta = 1$, $\lambda = 1$. Here

$$\log(1-\theta) = \tau ,$$

the complementary log transformation.

Another possible member is the log transformation. However, in that case the value of β depends on the unknown probability. We are interested in situations with several sets of data where the probabilities vary from set to set. The value of β characterizing the log transformation would vary accordingly. Then, this characterization is not useful for the interpretation of results.

For $\beta = 0$ and $\lambda = 0$ (2.5.2) reduces to

$$\log \log(\theta/(1-\theta)) = \tau .$$

The loglogistic transformation is obtained in this case. This transformation is not defined when $\theta/(1-\theta) < 1$. Then, some care must be used if the value zero is tried for both parameters. However, these transformations will be used mainly for extreme cases. Then, if all the probabilities are greater than 1/2 there is no problem. Besides, simple recoding serves to avoid the problem if all the probabilities are less than 1/2. This suggests a simple generalization of (2.5.1), namely

$$V^*_{\lambda,\beta,\delta}(\theta) = \frac{\{\log[(\frac{\theta}{1-\theta})^{\delta} + \beta]\}^{\lambda} - 1}{\lambda} , \qquad (2.5.3)$$

where $-\infty < \lambda < \infty$, $0 \leq \beta$ and $\delta = -1,1$.

The parameters in (2.5.3) may be interpreted as follows: λ determines the scale, β the degree of asymmetry and δ the suitable coding of successes and failures. An even richer family may be defined if we let δ to take all the values in $[-1,1]$ with the exception of the trivial $\delta = 0$. This enriched family would correspond to an analysis of fractional powers of odds ratios.

Transformations that can be obtained varying the values of λ, β and δ in (2.5.3) are summarized in Table 2.8.

The inverse of the general expression for the family $\mathcal{V}^* = \{V^*_{\lambda,\beta,\delta}\}$ takes the form

$$\theta = \frac{\{\exp(1+\lambda\tau)^{1/\lambda} - \beta\}^{1/\delta}}{1 + \{\exp(1+\lambda\tau)^{1/\lambda} - \beta\}^{1/\delta}} , \qquad (2.5.4)$$

where τ is in the image of $V^*_{\lambda,\beta,\delta}$.

This reduces for $\delta = 1$, to

$$\theta = \frac{\exp(1+\lambda\tau)^{1/\lambda} - \beta}{\exp(1+\lambda\tau)^{1/\lambda} + (1-\beta)} . \qquad (2.5.5)$$

A simpler family may be defined as follows:

$$W(\theta) = \frac{[\frac{\theta}{1-\theta} + \beta]^{\lambda} - 1}{\lambda} , \qquad (2.5.6)$$

where $\beta \geq 0$, $\lambda \in \mathbb{R}$ are unknown parameters.

Expression (2.5.6) reduces to logit(θ) for $\beta = 0$ and $\lambda = 0$, to the negative of the complementary log transformation for $\beta = 1$ and $\lambda = 0$, and to the odds ratio for $\beta = \lambda = 1$.

Although (2.5.6) is not as rich as (2.5.1), it may be useful to model in a simple way certain departures from the logistic in the direction of asymmetric transformations. Consider as before that the alternative is the complementary log log transformation. Denote by $W^*(\theta)$ the case of (2.5.6) for $\beta = 1$. We assume that

$$\ln W^*(\theta) = \tau , \qquad (2.5.7)$$

where τ denotes a value in the image of the composed transformation in (2.5.7). This expression serves to define a GLM if τ is the linear systematic part of the model. For $\lambda = 1$ the logistic model is obtained and for $\lambda = 0$ the complementary log log.

Hence, these two models may be compared in terms of just one parameter if we restrict our attention to the subfamily $\mathcal{W}$ of transformations with general form in (2.5.6) with $\beta = 1$ and $\lambda \geq 0$. The inverse of (2.5.7) is easily obtained, namely

$$\theta(\tau) = \begin{cases} 1 - (1+\lambda e^{\tau})^{-1/\lambda} & \lambda e^{\tau} > -1, \\ 1 & \text{otherwise.} \end{cases} \qquad (2.5.8)$$

This expression is used in the next chapter to test for asymmetric departures from the logistic model in the direction of models in the complementary log log scale.

TABLE 2.8

Some members of the family $\mathcal{U}^*$

Parameters			Associated transformation
λ	β	δ	
1	0	1	logistic
0	0	1	log logistic
1	1	1	complementary log
0	1	1	complementary log log
0	1	-1	logistic(1)
0	0	-1	log logistic(1)
1	1	-1	log
0	1	-1	log log

(1) Transformation for a recoding of the data.

Chapter 3: TESTS FOR DEPARTURES FROM THE LOGISTIC MODEL

3.1 Introduction

The family $\mathcal{T}$ of symmetric transformations for binary data was introduced in chapter 2. We now consider its application to contingency tables and in particular to representations with no interaction. We distinguish sharply between response and explanatory variables, as stressed explicitly by Bhapkar and Koch (1968), who give a comprehensive discussion on the formulation of the hypotheses to be tested, and their interpretation according to the different kinds of experimental situation which might generate the data. We consider in this chapter "uni-response, multi-factor" experiments, where the interest lies in the way in which the explanatory variables affect the response, and the problems of analysis are analogous to those in analysis of variance for continuous observations.

Two forms of no interaction commonly considered are the so-called additive and multiplicative definitions. These names refer to the way factors are assumed to affect the response. They may be considered as special cases of a more general family. The concept of no interaction is related to statistical independence in the case of experimental situations with multiple responses, where one of the objectives is to study the relationships among the different responses. That case is not considered in this chapter.

It is also interesting to test for departures from the logistic model in the direction of asymmetric alternatives. The subfamily $\mathcal{W}$ defined in §2.5 is used for such objective.

3.2 Transformations in $\mathcal{T}$ and definitions of no interaction

To discuss the relation between the transformation parameter and possible definitions of no interaction, we consider first the interpretation that may be given to the additive and multiplicative definitions cited above. In the first case the probabilities are represented by a linear expression in terms of parameters associated with the explanatory variables. Then for a factorial arrangement absence of interaction between two factors, say A and B, means that the difference between probabilities at two arbitrary levels of A is the same for all levels of B, i.e. the effects of A and B on the probabilities themselves are additive. In the second case, the multiplicative definition, the underlying probabilities are assumed instead to be decomposable as a product of parameters associated with the explanatory variables. Thus, the difference of logits at two arbitrary levels of A remains constant for all levels of B.

In terms of the general family of transformations, we may consider models of the form

$$T_\lambda(\theta_{ij}) - T_\lambda(\theta_{i'j}) = \Delta_{ii'} \quad \text{for all } j, \tag{3.2.1}$$

where i,i' are arbitrary, $\lambda = 0,1$ and $\Delta_{ii'}$ is a constant; θ_{ij} denotes the probability of positive response being in the ith category of A and the jth category of B.

Expression (3.2.1) reduces to a difference between logits for $\lambda = 0$, whereas for $\lambda = 1$ it is a difference between probabilities. Hence as an immediate generalization of (3.2.1) we suppose that for some value of λ a mode of no interaction holds. In particular the inclusion of both logistic and linear transformations enables us to model, by means of members of $\mathcal{T}$, departures from the assumption of multiplicative influence of the explanatory variables on the response in the direction of additive influence. The results in §2.4 make possible to interpret models fitted in probit or arcsine scales, as those adequate for values of λ intermediate between zero and one.

Note that there is no reason to restrict λ to the interval [0,1]. An informal discussion of the effect of choosing values of λ greater than one may be based on Table 3.1, which shows the absolute values of transformed probabilities in the range [0, 0.49], and by symmetry [0.51, 1], for several values of λ. It is convenient to define the intervals R_1, R_1' and R_2 as follows: $R_1 = [0, 0.25)$, $R_1' = (0.75, 1]$ and $R_2 = (0.25, 0.75)$. Note the increasing importance given to changes in probabilities within R_2 for increasing values of λ. The weight put on similar changes in probabilities within R_1 or R_1' decreases accordingly; see Lewis* (1962), who points out the ordering of certain transformations with respect to the weight given to probabilities at the extremes of the range [0,1]. He also comments on the similarity of results from tests of no second-order interaction, associated with linear, logistic and probit transformations, when the probabilities lie within the range [0.2, 0.8]. This is caused by the near equivalence of those transformations in that range; see for example Cox (1970, pp. 27-29).

* Lewis, B.N. (1962). On the analysis of interaction in multidimensional contingency tables. J. R. Statist. Soc. A125 88-117

Table 3.1 Absolute value of transformed probabilities under $T_\lambda(\cdot)$ for several values of the transformation parameter λ.

Probability	T_0	$T_{.1}$	$T_{.5}$	$T_{.7}$	$T_{.9}$	T_1	$T_{\sqrt{2}}$	T_2	T_3	T_4
0 (1)	∞	20.0	4.0	2.857	2.222	2.000	1.414	1.000	0.667	0.500
0.05 (0.95)	2.944	2.923	2.507	2.212	1.929	1.800	1.371	0.994	0.666	0.500
0.1 (0.9)	2.197	2.188	2.000	1.847	1.682	1.600	1.293	0.976	0.665	0.500
0.15 (0.85)	1.735	1.730	1.633	1.549	1.451	1.400	1.190	0.940	0.659	0.499
0.2 (0.8)	1.386	1.384	1.333	1.287	1.231	1.200	1.065	0.882	0.646	0.496
0.25 (0.75)	1.099	1.098	1.072	1.047	1.017	1.000	0.920	0.800	0.619	0.488
0.3 (0.7)	0.847	0.847	0.835	0.823	0.808	0.800	0.759	0.690	0.589	0.467
0.35 (0.65)	0.619	0.619	0.614	0.610	0.604	0.600	0.582	0.550	0.487	0.422
0.4 (0.6)	0.405	0.405	0.404	0.403	0.401	0.400	0.395	0.385	0.362	0.335
0.45 (0.55)	0.201	0.201	0.201	0.200	0.200	0.200	0.199	0.198	0.195	0.191
0.47 (0.53)	0.120	0.120	0.120	0.120	0.120	0.120	0.120	0.120	0.119	0.118
0.49 (0.51)	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040	0.040

3.3 Test of the assumption of multiplicative effects

We develop below a test of the appropriateness of the logistic scale for linear models with a given configuration. The test is based on the characterization of the logistic transformation as an element of $\mathfrak{T}$; it detects departures in the direction of other symmetric transformations which give less weight to extreme probabilities.

The problem is then to test the hypothesis H_M: $\lambda = 0$. We consider the parameter vector $\underset{\sim}{\beta}$, in the linear systematic expression of the model, a nuisance parameter. An exact similar test is not available. One possible alternative is to use a maximum likelihood ratio test, see Box and Cox (1964) for an application in a similar context. However, in this case it is preferable for computational simplicity, to employ a score test which is asymptotically equivalent and locally most powerful (see Cox and Hinkley, 1974 §9.3; Atkinson, 1973).

To carry out the procedure we need to compute the efficient score $U(\lambda) = \partial\ell/\partial\lambda$ and the Fisher's information matrix when $\lambda = 0$. In principle we must make allowance for the presence of the nuisance parameter (Moran, 1970). Thus we need to consider the statistic

$$A = (\partial\ell/\partial\lambda)_{\underset{\sim}{\omega}} - \sum_{j=1}^{p} \gamma_j (\partial\ell/\partial\beta_j)_{\underset{\sim}{\omega}} ,$$

where $\underset{\sim}{\omega} = (\lambda,\underset{\sim}{\beta})$, γ_j $(j = 1,\ldots,p)$ are the regression coefficients of the first term on the others, and p is the dimension of $\underset{\sim}{\beta}$.

To obtain a usable form of the statistic A, some estimate of $\underset{\sim}{\beta}$ should be substituted instead of the unknown parameter. Substitution of $\hat{\underset{\sim}{\beta}}_0$, the m.l.e. of $\underset{\sim}{\beta}$ when $\lambda = 0$, adds only an $o_p(1)$ term and because $\hat{\underset{\sim}{\beta}}_0$ is assumed to satisfy $U_{\underset{\sim}{\beta}}(0,\underset{\sim}{\beta}) = \underset{\sim}{0}$, the statistic A is greatly simplified. Besides, the parametric model under H_M needs to be fitted only. A minor problem arises because $U(\lambda)$ vanishes identically when

$\lambda = 0$. One possible solution is to reparameterize the problem. In this case the reparameterization we choose is in terms of $\phi = \lambda^2$, and the efficient score for ϕ takes the form

$$U_{.}(\phi) = \sum_{i=1}^{m} U_i(\phi) \quad , \qquad (3.3.1)$$

where for $i = 1,\ldots,m$

$$U_i(\phi) = (r_i - n_i\hat{\theta}_i)\left[\frac{\phi^{-1}\tau_i}{2(1-\phi\tau_i^2/4)} - \phi^{-3/2}\tanh^{-1}(\phi^{1/2}\tau_i/2)\right],$$

$$\hat{\theta}_i = (1+\phi^{1/2}\tau_i/2)^{\phi^{-1/2}}\left[(1+\phi^{1/2}\tau_i/2)^{\phi^{-1/2}} + (1-\phi^{1/2}\tau_i/2)^{\phi^{-1/2}}\right]^{-1} ,$$

τ_i is the linear systematic part of the model,

r_i is the number of positive responses in group i,

n_i is the number of individuals in group i.

Expression (3.3.1) takes the following limit form as ϕ tends to zero,

$$U_{.}(0) = \sum_{i=1}^{m} (r_i - n_i\hat{\theta}_i^{o})\tau_i^{3}/12 \quad , \qquad (3.3.2)$$

where $\hat{\theta}_i^{o} = [1 + \exp(-\tau_i)]^{-1}$.

The test may be carried out assuming asymptotic normality of a standardized form of (3.3.2). Large values of this statistic will lead to rejection of H_M. To standardize (3.3.2) we need the value of its variance which may be expressed as

$$I_{\phi_0\phi_0} - I_{\phi_0\underset{\sim}{\beta}}\, I^{-1}_{\underset{\sim}{\beta}\underset{\sim}{\beta}}\, I_{\underset{\sim}{\beta}\phi_0} \ , \qquad (3.3.3)$$

where

$$I = \left(\begin{array}{c|c} I_{\phi_0\phi_0} & I_{\phi_0\underset{\sim}{\beta}} \\ \hline I_{\underset{\sim}{\beta}\phi_0} & I_{\underset{\sim}{\beta}\underset{\sim}{\beta}} \end{array} \right) ,$$

is Fisher's information matrix for $(\phi_0, \underset{\sim}{\beta})$.

The components of I are simple to compute and have the following expressions

$$I_{\phi_0\phi_0} = \sum_{i=1}^{m} n_i d_i \tau_i^{6}/144 ,$$

$$I_{\phi_0\beta_s} = \sum_{i=1}^{m} n_i d_i x_s \tau_i^{3}/12 \qquad (s = 1,\ldots,p),$$

$$I_{\beta_r\beta_s} = \sum_{i=1}^{m} n_i d_i x_s x_r \qquad (r,s = 1,\ldots,p),$$

where $d_i = \hat{\theta}_i^{o}(1 - \hat{\theta}_i^{o})$.

All the values required to perform the test may be computed from the output of a logistic fit. For a GLIM fit, the τ_i $(i = 1,\ldots,m)$ are given directly by the 'linear predictor vector'.

It is interesting that the factor $\tau_i^{3}/12$ which appears in the expression of the efficient score $U_{.}(0)$, weighting the discrepancies between observed and expected positive responses, may be obtained by other means as a measure of disagreement with the logistic model. Consider that the appropriate transformation in a certain situation is $T_{\phi}(\cdot)$ with $\phi \neq 0$. Then the following expression holds

$$\text{logit}(\theta_i) = \phi^{-1/2} \ln \frac{1 + \phi^{1/2}\tau_i/2}{1 - \phi^{1/2}\tau_i/2} \quad (i = 1,\ldots,m). \qquad (3.3.4)$$

Expanding the right hand side of (3.3.4) in series we obtain

$$\text{logit}(\theta_i) = 2\phi^{-1/2} \sum_{k=1}^{\infty} \frac{1}{2k-1} (\phi^{1/2}\tau_i/2)^{2k-1} ,$$

neglecting terms of order higher than 3 we have that

$$\text{logit}(\theta_i) \equiv \tau_i + \phi\tau_i^3/12 \quad .$$

Hence, $\tau_i^3/12$ may be interpreted as a measure of inadequacy of the logit model in this case.

An analogous test may be devised to assess the plausibility of the linear scale. The quantities required to carry out an asymptotically optimal test of H_A: $\lambda = 1$, are given by the following expressions,

$$U_{.}(1) = \sum_{i=1}^{m} (r_i - n_i\hat{\theta}_i')u_i ,$$

$$I_{\lambda\lambda} = \sum_{i=1}^{m} n_i e_i u_i^2 ,$$

$$I_{\lambda\beta_s} = \sum_{i=1}^{m} n_i e_i x_s u_i/(1 - \tau_i^2/4),$$

$$I_{\beta_r\beta_s} = \sum_{i=1}^{m} n_i e_i x_r x_s/(1 - \tau_i^2/4)^2 ,$$

where $e_i = \hat{\theta}_i'(1 - \hat{\theta}_i')$, $\quad \hat{\theta}_i' = (2 + \tau_i)/4$

$$u_i = [\frac{\tau_i}{1 - \tau_i^2/4} - 2 \tanh^{-1}(\tau_i+2)] .$$

Large negative values of the standardized test statistic will lead to rejection of H_A: $\lambda = 1$. Asymptotic normality of this standardized form is assumed.

3.4 Test of asymmetry

The test in the last section is aimed to detect departures from the logistic model within a family of symmetric transformations. Here we develop a test based on members of the family $\mathcal{W}$ defined in §2.5. The objective is to detect departures from the logistic model in the direction of asymmetric transformations. The complementary log log transformation was chosen as a plausible alternative.

We assume the data are as above. Then the loglikelihood function, substituting the expression for θ_i in (2.5.8), may be expressed as

$$\ell = \sum_{i=1}^{m} [r_i \ln(c_i^{1/\lambda} - 1) - n_i \ln(c_i)/\lambda] ,$$

where $c_i = 1 + \lambda \exp(\tau_i)$.

The efficient score for λ takes the form

$$U_{.}(\lambda) = \partial\ell/\partial\lambda = \sum_{i=1}^{m} \frac{r_i - n_i\hat{\theta}_i}{\hat{\theta}_i} \left[\frac{\exp(\tau_i)}{\lambda c_i} - \frac{\ln(c_i)}{\lambda^2} \right] .$$

Within the family $\mathcal{W}$, the logistic transformation is characterized by $\lambda = 1$. We develop below an asymptotically optimal test for the hypothesis H_L: $\lambda = 1$. This test is based on $U_{.}(1)$ and the alternatives of interest are $\lambda < 1$. The test statistic, having substituted the m.l.e. of the nuisance parameters under $\lambda = 1$, is basically the efficient score of $\lambda = 1$ with expression

$$U_{\cdot}(1) = \sum_{i=1}^{m} \frac{r_i - n_i\hat{\theta}_i}{\hat{\theta}_i} [\hat{\theta}_i + \ln(1 - \hat{\theta}_i)] \quad ,$$

whose standard deviation may be computed as in the last section. In this case the components of I, Fisher's information matrix, have the form

$$I_{\lambda\lambda} = \sum_{i=1}^{m} n_i[\hat{\theta}_i + \ln(1 - \hat{\theta}_i)]^2/e_i,$$

$$I_{\beta_s\lambda} = \sum_{i=1}^{m} [\hat{\theta}_i + \ln(1 - \hat{\theta}_i)]n_i x_s(1 - \hat{\theta}_i),$$

$$I_{\beta_r\beta_s} = \sum_{i=1}^{m} n_i x_s x_r \hat{\theta}_i(1 - \hat{\theta}_i) \quad ,$$

where $e_i = \exp(\tau_i), \quad \hat{\theta}_i = e_i/(1 + e_i).$

Just as for the test in §3.3, the one suggested here may be carried out almost directly using the results of a logistic fit. For a GLIM fit the values of τ_i $(i = 1,\ldots,m)$ are given by the 'linear predictor' vector. We shall reject the hypothesis $\lambda = 1$ when large negative values of the standardized test statistic are observed. An asymptotic standard normal distribution of the statistic is assumed.

Chapter 4. APPLICATIONS OF THE SYMMETRIC FAMILY

4.1 Introduction

This chapter comprises the analysis of some sets of data which have been examined before by other authors. One object is comparison with previous results and solution of some open questions about previous findings. The method used consists, essentially, in fitting generalized linear models based on members of the family Υ proposed in Chapter 2. Our aim is to show some advantages of this approach, especially the possibility of making a quantitative assessment of the suitability of a particular scale to obtain a simple decomposition or representation of the underlying probabilities.

We borrow freely from the terminology of analysis of variance. Thus, throughout the chapter, we refer to explanatory variables as factors; models including only terms associated with the factors and a general mean are called main effects models. Interaction between two factors is denoted by a sequence of two digits, e.g. 13, which in this case means that the interaction is between factors 1 and 3, for a consistent classification of the factors. Interactions of higher order may be defined similarly.

The analyses were carried out by means of a step-up procedure comparing at each step across scales and from step to step between different model configurations. A main effects model was fitted initially to assess the significance of the inclusion of further terms. Comparisons are based on twice the difference between the maximized loglikelihood achieved for different models. This quantity is assumed to have an approximate chi-squared distribution with the corresponding number of degrees of freedom. We were unable to compute the correction factor needed to improve the approximation (Lawley, 1956), but it is very likely that it is larger than one. This procedure was complemented by the comparison of the maxloglikelihood value attained by a particular model with the overall maxloglikelihood achievable, i.e. the one for the saturated model; in this way a goodness-of-fit test was readily obtained. One d.f. is reduced because of the estimation of φ.

Approximate confidence intervals for the transformation parameter may be obtained from the loglikelihood curve as in Box and Cox (1964).

4.2 An example from data in The American Soldier

The following results correspond to the analysis of data originally presented by Stouffer et al. (1949) and analyzed subsequently by several other authors.

Table 4.1 shows a cross-classification of 8036 soldiers with respect to four dichotomized variables, namely (1) race, (2) region of origin, (3) location of present camp, and (4) preference as to camp location. The last is considered below as a response to the first three.

The last column of table 4.1 suggests that the logistic scale may be the most suitable for a simple decomposition of the probabilities in terms of the explanatory variables. However, in view of the discussion of these data in Coleman (1964, pp. 198-199), it seems interesting to assess the plausibility of an additive representation on alternative scales for which the interpretation may be easier or more practical.

A summary of the results is shown in Table 4.2, all of them correspond to the logistic scale because that was the member of $\mathcal{T}$ which consistently provided the best fit for these particular data.

We see that the inclusion of either interaction (13), between race and location of present camp, or interaction (12), between race and region of origin, does not improve upon the fit of model S1, that includes main effects only (see models S3 and S2 respectively). Those interaction effects may seem required to obtain a simple representation of the transformed proportions, but in fact are not necessary. In the results obtained for different values of the transformation parameter, interactions 12 and 13 do not improve the fit for a broad range of that parameter. We may conclude that, practically, those factors do not interact.

Model S4, including main effects and interaction between region of origin and location of present camp (23), provides a great improvement upon model S1 as may be observed in Table 4.3. Thus, a simple representation on the logistic scale may be based on the original explanatory variables and the interaction between region of origin and location of present camp. This assertion is reinforced by the results for the more complex models S5 and S6 which do not improve significantly upon model S4.

Our results agree with those obtained by Goodman (1972) but disagree with the ones in Coleman (1964). This is because a substantial improvement in the maxlikelihood was also achieved in the linear scale, as may be observed comparing figures 4.1 and 4.2, for model S4. Coleman attributes the lack of fit of his main effects model in a linear scale to a supposed interaction between region of origin and present location of camp which applies to blacks but not to whites. We detected an interaction which applies homogeneously to both, blacks and whites. It is not possible to make a direct comparison with Coleman's method.

We may say that in this particular case a qualitative discussion of the effects of the explanatory variables on the response may be based

TABLE 4.1

Number of soldiers who prefer a northern camp location classified with respect to three dichotomized variables, namely (1) race, (2) region of origin, and (3) location of present camp.

Race	Region of origin	Location present camp	Positive response	Total	Proportion
B	N	N	387	423	0.915
B	N	S	876	1126	0.778
B	S	N	383	653	0.587
B	S	S	381	2093	0.182
W	N	N	955	1117	0.855
W	N	S	874	1384	0.632
W	S	N	104	280	0.371
W	S	S	91	960	0.095

Source: Goodman (1972). B stands for black, W for white, N for North and S for South.

TABLE 4.2

Results for several logistic models fitted to data in Table 4.1

	Model configuration	Max-loglikelihood	Improvement over m.e. model	Goodness of fit
S1	Main effects, m.e.	-4026.605	---	24.962 (3)
S2	m.e. + 12	-4026.582	0.046 (1)	24.916 (2)
S3	m.e. + 13	-4022.810	7.590 (1)	17.372 (2)
S4	m.e. + 23	-4014.847	23.516 (1)	1.446 (2)
S5	m.e. + 13 + 23	-4014.783	23.644 (2)	1.317 (1)
S6	m.e. + 12 + 23	-4014.466	24.278 (2)	0.683 (1)
	Saturated model	-4014.124		

N.B. Degrees of freedom of corresponding asymptotic χ^2 appear in parentheses.

TABLE 4.3

Results from the fit of model S4 in Table 4.2.

The scale used is the logistic.

Race	Region of origin	Location of present camp	Observed value	Expected value
B	N	N	387	390.636
B	N	S	876	879.786
B	S	N	383	376.786
B	S	S	381	380.265
W	N	N	955	951.364
W	N	S	874	870.687
W	S	N	104	110.214
W	S	S	91	91.735

Pearson chi-squared : 1.4552 2 d.f.

L.R. chi-squared : 1.4458

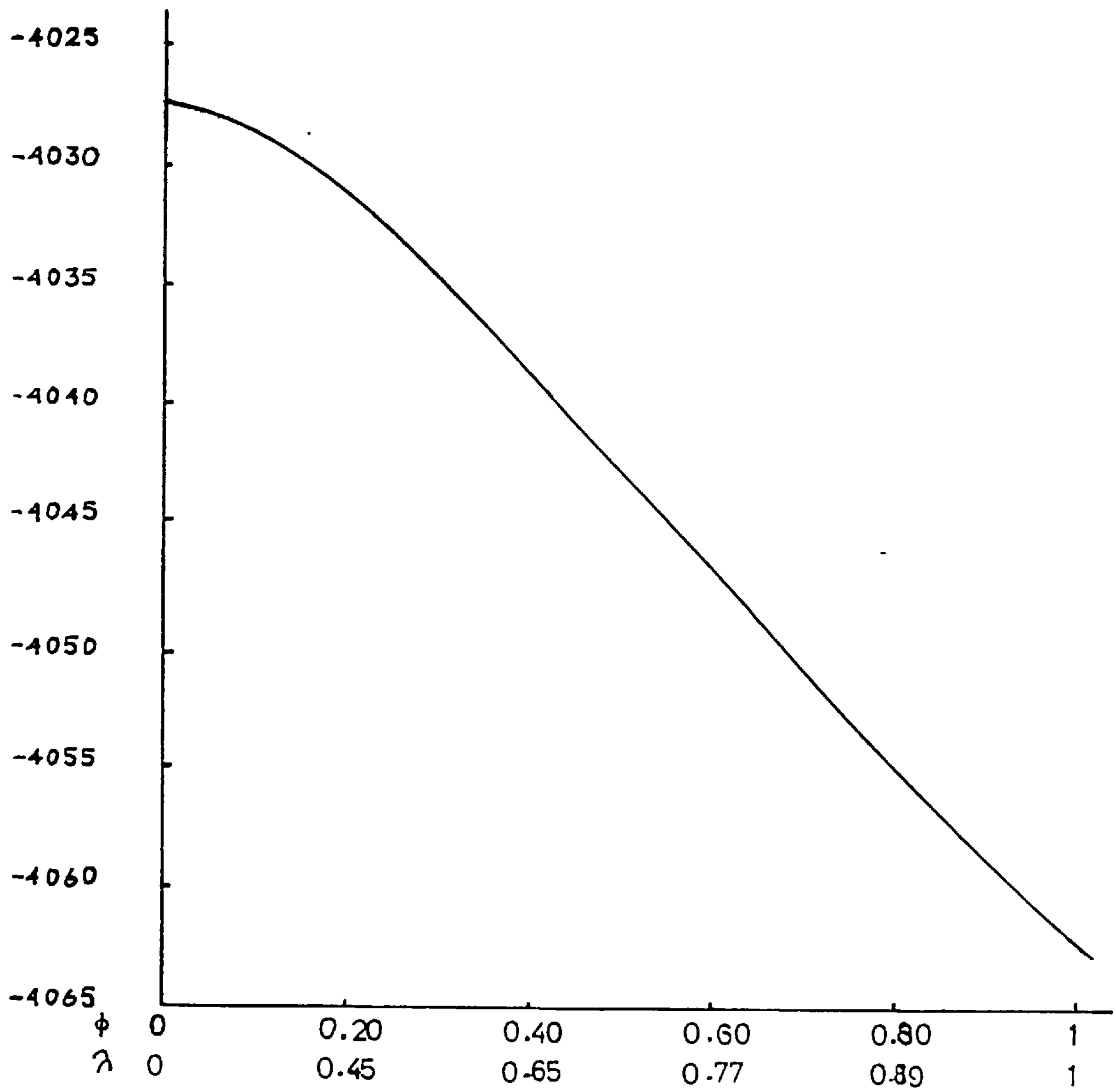

Figure 4.1

Loglikelihood curve for the transformation parameter. Model configuration: main effects; data in Table 4.1.

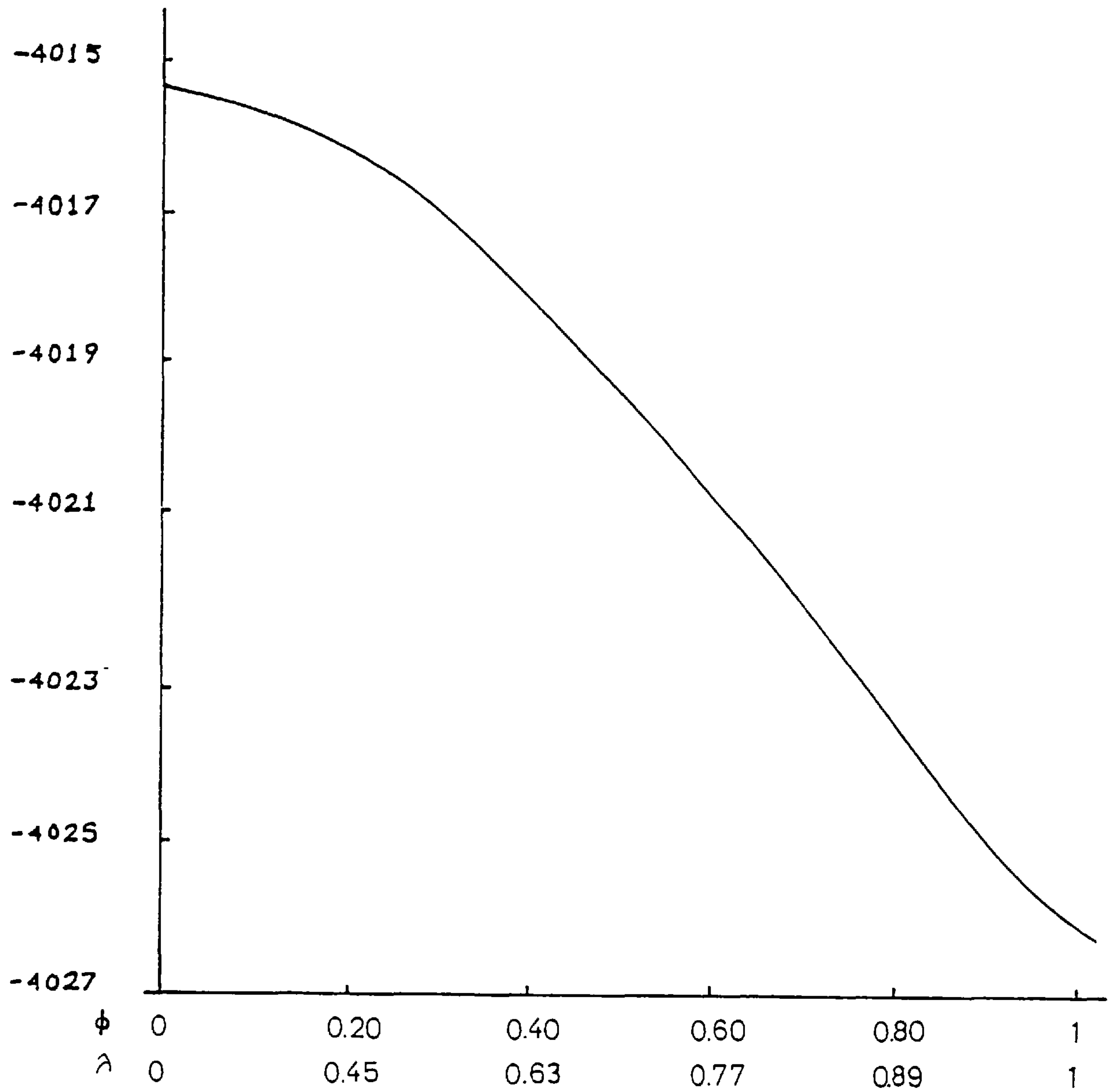

Figure 4.2

Loglikelihood curve for the transformation parameter. Model configuration: main effects plus interaction 23; data in Table 4.1.

either on the logistic or linear scales. However, if a quantitative discussion is needed, the logistic scale is the most appropriate (see especially figures 4.1 and 4.2).

4.3 Two examples of data on deviant behaviour

Goodman (1975) compares methodically two ways of analyzing systems of dichotomous variables, namely fitting models where the effects of the factors are assumed to be either multiplicative or additive. He comments on the similarity of results for both approaches when the proportions lie in the interval $I = [0.25, 0.75]$. See the discussion in §2.4 in this respect.

Goodman presents a detailed exposition of the correspondences of the parameters for the additive and multiplicative representations, he gives also the simple relationships between the parameters for different codings for the explanatory variables. This is useful for researchers who want to compare their results with those of others but find that the coding was different.

The parameters of models fitted in the logistic and in the linear scales differ little when the probabilities lie in the interval I. In this case, models in a scale associated with $\phi \geq 1$ will tend to give a better fit. This is because of the relative weight put on probabilities in I by these scales. However, if most of the proportions lie outside the interval I, that behaviour will reverse and values of the parameter ϕ near to zero will often be more adequate. This is because of the relative weight given to probabilities on the extremes by scales with small ϕ. See the discussion in §2.4.

Knoke (1975) provides an example of each of the above situations. His paper is closely related to the one by Goodman.

Knoke's data consist of two sets taken from the 1973 General Social Survey conducted by the National Opinion Research Center (U.S.). Two items on preference for legalizing deviant behaviours were chosen for analysis. The response variable in the first set is agreement with legalization of abortion. The factors are: (1) church attendance, (2) education, and (3) religion, dichotomized as shown in Table 4.4. In this case most of the proportions lie in the central range. The second set is an example of extreme proportions, the response variable is agreement with legalization of use of marijuana and the dichotomized factors are: (1) church attendance, (2) education, and (3) age. These data appear in Table 4.5.

Our results are summarized in Tables 4.6 for data on legalization of abortion, and 4.7 for data on legalization of use of marijuana.

For the data on abortion model A4, including main effects and interaction between education and religion (23), is the one which represents the data most simply. The improvement over the main effects model (A1) is significant at 5% level, and, as may be observed in Table 4.6, the inclusion of further terms brings very little improvement in the maximized loglikelihood achieved. This is reinforced by the results shown in Table 4.8 for a fit of model A4 to the data. An approximate confidence region for the transformation parameter at the 90% level excludes the value $\phi = 0$, i.e. the logistic transformation is not appropriate for that model configuration. The linear scale, $\phi = 1$, is plausible (see figure 4.3) judging by its inclusion in approximate confidence intervals for ϕ at the usual levels. For simplicity we may use the linear scale for representing the data rather than the one for $\hat{\phi}$. Our results differ from those suggested by Knoke (1975), who proposes model A6 consisting in main effects plus interactions between church attendance and education (12), and between education and religion (23).

We may observe in Table 4.6 that the improvement of model A6 upon A4, in the scale suggested by our method, is not significant at any usual level. Nevertheless, the inclusion of (12) is significant in the logistic scale which is equivalent to the one used by Knoke. However, we prefer the simpler decomposition provided by model A4.

The fluctuation in the value of $\hat{\phi}$ for different models in Table 4.6 is a striking feature of these data. This is because the proportions are not extreme so that a broad range of values of ϕ provide essentially equivalent scales. For instance, although the model proposed by Knoke, A6, differs in configuration from ours, A4, just in the inclusion of the interaction term 12, $\hat{\phi}$ changes from 1.6 to 0, suggesting the logistic scale as appropriate. However, an approximate confidence interval for ϕ includes the indicator of the linear and other more extreme transformations as well. Besides, the inclusion of the term 12 improves the max-loglikelihood achieved for the configuration A4 in the logistic scale but not in others (see discussion in §3.2). The results of the fit of model A4 appear in Table 4.8.

With respect to the data on legalization of marijuana the results were more predictable. In general the logistic scale was chosen by our method as the most suitable, i.e. $\hat{\phi}$ was near or equal to zero, to obtain a simple representation of the data. It is clear from Table 4.7 that a model in the logistic scale including only main effects explains the data very reasonably, and the inclusion of further terms is superfluous. Besides, the best models including one two-factor interaction term plus main effects only are those with interactions 13 or 23. These are related with age, probably the most important determinant of attitude towards legalization of marijuana.

Knoke's results in relation to what he calls the regression approach (linear scale) are confirmed by our analyses ($\phi = 1$). In this case the model with main effects and interaction 13, between church attendance and age, does not differ significantly from the main effects model. The interaction 13 is not relevant on the linear scale, but it is so in the logistic. Results of the fit of model M1 appear in Table 4.9.

Knoke discusses the difference between the effects identified as relevant on the logistic and linear scales and the "correctness" of the findings. Our results suggest that, in this case, the logistic scale is suitable for an additive decomposition of the probabilities in terms of the factors (see Fig. 4.4). The suggested method allows us to make a direct and homogeneous quantitative comparison.

Knoke compares models fitted in different scales. These models are equivalent only in their configuration, the meaning of the interaction terms is different on different scales. This has to be taken into account whenever a comparison is attempted.

TABLE 4.4

Observed frequencies on willingness to legalize abortion. The individuals are classified according to: (1) church attendance, (2) education and (3) religion.

		Church attendance			
		Low		High	
Religion	Education	No	Yes	No	Yes
Catholic	College	47	23	62	92
	No college	110	43	211	136
Non-catholic	College	11	24	24	168
	No college	55	61	150	229

Source: Knoke (1975)

TABLE 4.5

Observed frequencies on willingness to legalize the use of marijuana. The individuals are classified according to: (1) church attendance, (2) education and (3) age.

		Church attendance			
		Low		High	
Age	Education	No	Yes	No	Yes
Old	College	79	34	101	15
	No college	157	25	319	12
Young	College	67	74	55	24
	No college	168	57	234	34

Source: Knoke (1975)

TABLE 4.6

Results for several models fitted to data on legalization of abortion in Table 4.4.

	ϕ	Model configuration	Maximized log-likelihood	Improvement over m.e. model	Goodness of fit
A1	2.8	Main effects, m.e.	-903.171	--	6.071 (3)
A2	1.5	m.e. + 12	-902.817	0.709 (1)	5.362 (2)
A3	3.0	m.e. + 13	-902.984	0.374 (1)	5.696 (2)
A4	1.6	m.e. + 23	-901.090	4.144 (1)	1.916 (2)
A5	1.5	m.e. + 12 + 13	-902.564	1.213 (2)	4.857 (1)
A6	0	m.e. + 12 + 23	-900.323	5.696 (2)	0.375 (1)
A7	2.0	m.e. + 13 + 23	-900.702	4.937 (2)	1.134 (1)
		Saturated	-900.136		

N.B. Degrees of freedom of corresponding asymptotic χ^2 appear in parentheses.

TABLE 4.7

Results for several models fitted to data on legalization of use of marijuana in Table 4.5.

	ϕ	Model configuration	Maximized log likelihood	Improvement over m.e. model	Goodness of fit
M1	0	Main effects, m.e.	-614.559	--	2.018 (3)
M2	0.5	m.e. + 12	-614.304	0.510 (1)	1.507 (2)
M3	0	m.e. + 13	-614.861	1.396 (1)	0.624 (2)
M4	0.5	m.e. + 23	-613.960	1.198 (1)	0.819 (2)
M5	0	m.e. + 12 + 13	-613.819	1.480 (2)	0.539 (1)
M6	0.5	m.e. + 12 + 23	-613.592	1.934 (2)	0.085 (1)
M7	0	m.e. + 13 + 23	-613.858	1.402 (2)	0.617 (1)
		Saturated	-613.549		

N.B. Degrees of freedom of corresponding asymptotic χ^2 appear in parentheses.

TABLE 4.8

Results from the fit of model A4, data on legalization of abortion

Church attendance	Education	Religion	Observed	Expected
High	No C	No Ca.	229	232.924
High	No C	Ca.	61	60.598
High	C	No Ca.	168	167.783
High	C	Ca.	24	21.097
Low	No C.	No Ca.	136	132.106
Low	No C.	Ca.	43	43.290
Low	C	No Ca.	92	92.696
Low	C	Ca	23	25.816

Pearson chi-squared: 1.876 2 d.f.

L.R. chi-squared: 1.910

N.B. C - college; Ca - catholic.

TABLE 4.9

Results from the fit of model M1. Data on legalization of use of marijuana.

Church attendance	Education	Age	Observed	Expected
High	No C	O	12	15.420
High	No C	Y	34	31.124
High	C	O	15	15.409
High	C	Y	24	23.047
Low	No C	O	25	21.612
Low	No C	Y	57	59.843
Low	C	O	34	33.558
Low	C	Y	74	74.986

Pearson chi-squared: 1.987 2.d.f.

L.R. chi-squared: 2.019

N.B. C - college; O - old; Y - young.

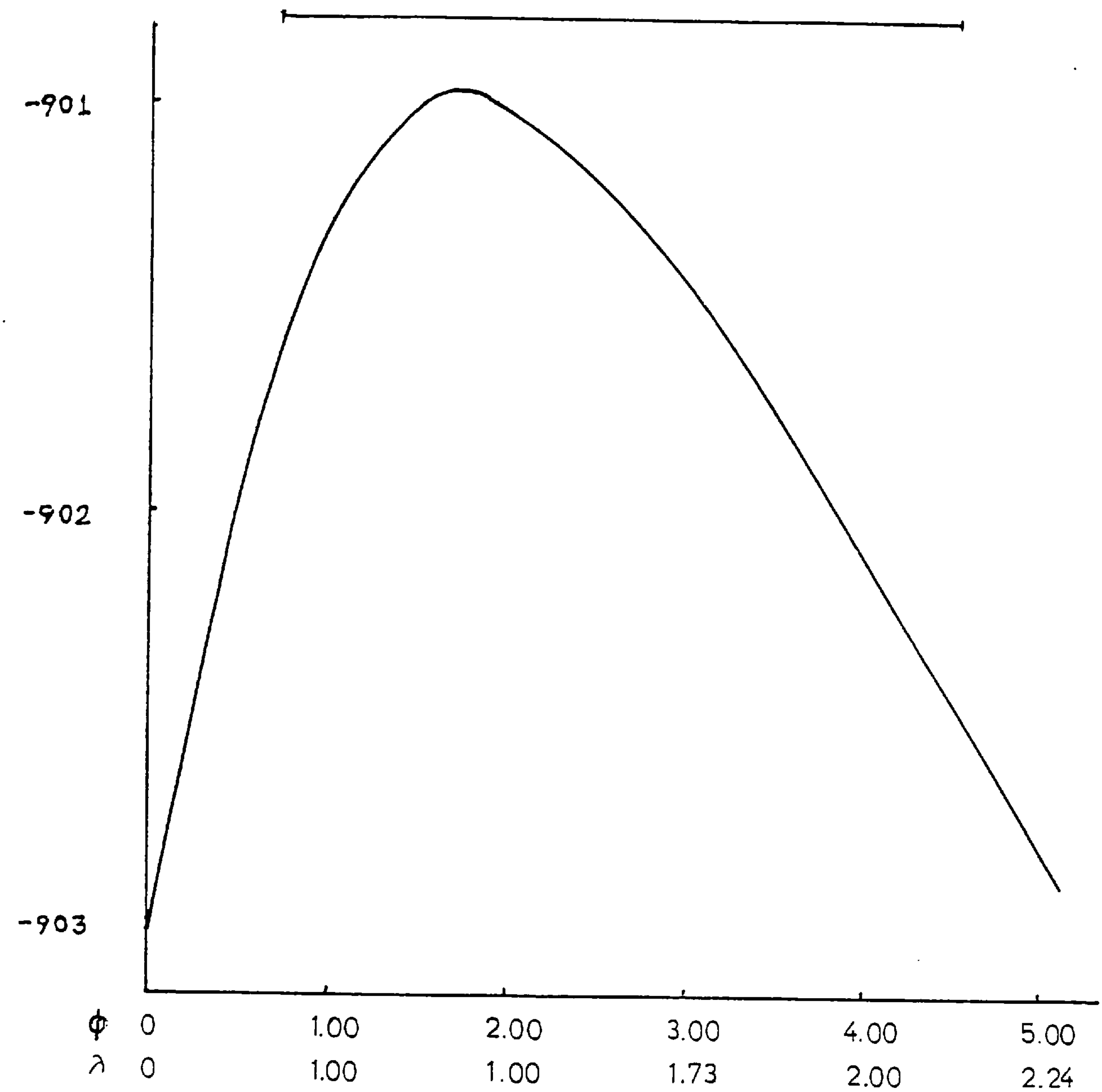

Figure 4.3

Loglikelihood curve for the transformation parameter. Model configuration: main effects plus interaction 23; data in Table 4.4. An approximate 90 per cent confidence interval for $\phi(\lambda)$ is shown on the top of the figure.

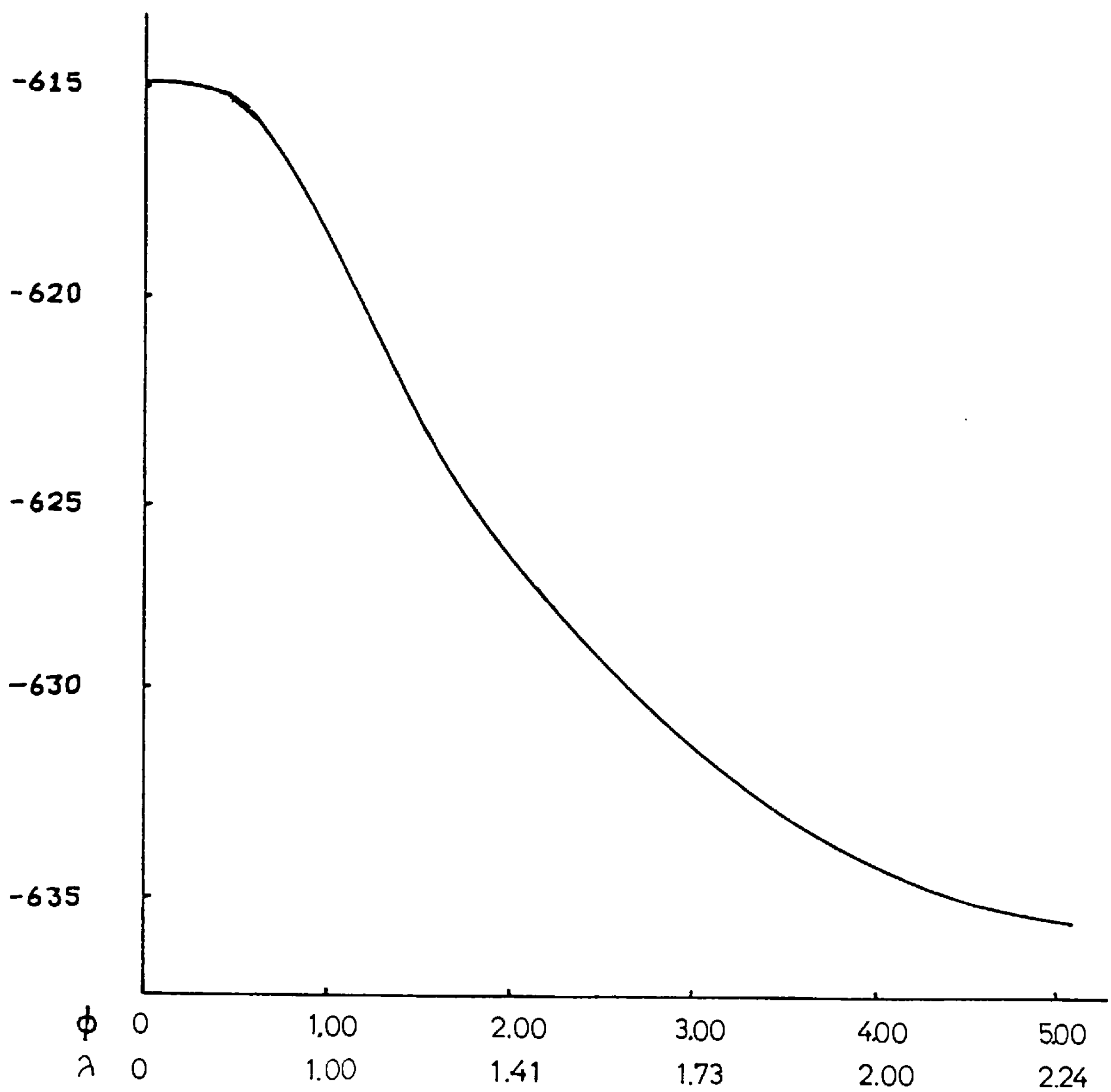

Figure 4.4

Loglikelihood curve for the transformation parameter.
Model configuration: main effects; data in Table 4.5.

4.4 Data on knowledge about cancer

Lombard and Doering (1947) reported and analyzed the data in Table 4.10 which consists of records of 1729 individuals who were allotted a good or poor score on cancer knowledge, classified according to the presence or absence of four variables, namely (1) exposure to radio addresses, (2) solid reading, (3) newspapers, and (4) exposure to lectures. These variables are considered below as explanatory for the expected proportions of individuals with good scores. Several authors have analyzed these data, we shall refer especially to the papers by Dyke and Patterson (1952) and Cox and Snell (1968). Some comparisons are made with their results.

Table 4.10 suggests that a linear transformation is likely to be adequate because most of the observed proportions lie in the range [0.25, 0.75]. This is confirmed below by our results. An interesting characteristic of this example is the use of an approximate screening technique. This is useful when the number of factors is moderate to large and the data form a structured sample.

The results are summarized in Table 4.11. From the fitting of a main effects model (C1), the plausibility of a simple representation on a logistic scale is nearly rejected at the 10% significance level. The main effects configuration fits the data well in the scale associated with $\phi = 1.8$, the value at which the approximate maximum of the loglikelihood curve for ϕ is attained, see figure 4.5.

Table 4.12 shows some results from the fit of model C1. The values of the commonly used Pearson's and likelihood ratio chi-squared statistics were not significant at the usual levels. However, further analysis was carried out to take account of possibly relevant higher-order effects. To have an indication of plausible terms to include in the model we took advantage of the structure of the problem, namely that of a 2×2^4 factorial system. Crude generalized residuals, as defined by Cox and Snell (1968) were computed as well as chi-squared residuals defined in this case as $(o_i - e_i)/\{e_i(1-\hat{\theta}_i)\}^{1/2}$, where o_i and e_i denote observed and expected responses respectively, $\hat{\theta}_i$ is the estimated probability of positive response. For model C1 these residuals appear in Table 4.12. We consider them as if they were observations from a 2^4 design and compute the sum of squares of the residuals (SSR) for the different effects as usual, using this to detect effects which seem worth including in future models. This method is used by Cox and Snell. It must be borne in mind that in this case the technique is just approximate, and to judge the actual relevance of a certain effect we need to fit a model including it. Table 4.13 shows the resulting SSR's after applying this procedure to the residuals in Table 4.12.

Results in Table 4.13 suggest the inclusion of the interaction effect 24, i.e. interaction between solid reading and exposure to lctures. This is an interesting feature of the data because the SSR for the factors 2 and 4 are not large, and usually the contrary happens. A possible explanation is that the interaction term needs to be included to account for duplication of knowledge obtained from the two sources of information. This seems to be confirmed by the actual fit of the suggested model and by the sign of the estimated parameter.

The fit of model C2, including main effects and interaction 24, results in a modest increment in the maximized loglikelihood value, although the fit to the data seems to have improved (compare Tables 4.12 and 4.14). Models C4 and C5 are fitted to compare with the results of other authors. Cox and Snell identify interactions 14, 24, 34 and 23, as worth including in the model, using the screening procedure mentioned above. They base their computations on two types of generalized residuals calculated from the original ones resulting from the fit of Dyke and Patterson's main effects models. Dyke and Patterson themselves suggest the inclusion of interactions 14, 24 and 34. Models C5 and C4 correspond respectively to the two configurations suggested above. From Table 4.11 it seems that model C4 is worth being considered but model C5 improves little upon C4. However, this is true for the scales shown in Table 4.11, which are approximately the most suitable for the associated model configuration. The inclusion of interaction 23 does improve the value of the maximized loglikelihood in the logistic scale, the scale where it is identified by Cox and Snell, but not in other scales. The identification of relevant interaction effects depends on the scale used to analyze the data.

Our results suggest that model C1 should be used if simplicity of representation and interpretation is wished. This is because models with main effects only have several advantages if they provide a reasonable fit to the data. There is, in principle, no reason to continue including more parameters, unless a complete explanation of the data is intended. With respect to the choice of scale, in this case the linear scale is plausible and, on some practical grounds, preferable. This leads to the adoption of an additive definition of no interaction. This is not the additive definition discussed by Darroch (1974), in the context of multiple response experimental situations.

TABLE 4.10

Lombard-Doering's data on cancer knowledge. Individuals are classified according to presence (+) or absence (-) of four variables: (1) exposure to radio addresses, (2) solid reading, (3) newspapers, and (4) exposure to lectures.

1	2	3	4	Good score	Total	Proportion
+	+	+	+	23	31	0.742
+	+	+	-	102	169	0.604
+	+	-	+	1	4	0.250
+	+	-	-	16	32	0.500
+	-	+	+	8	12	0.677
+	-	+	-	35	94	0.372
+	-	-	+	4	7	0.571
+	-	-	-	13	63	0.206
-	+	+	+	27	45	0.600
-	+	+	-	201	378	0.532
-	+	-	+	3	11	0.273
-	+	-	-	67	150	0.447
-	-	+	+	7	13	0.538
-	-	+	-	75	231	0.325
-	-	-	+	2	12	0.167
-	-	-	-	84	477	0.176

Source: Lombard and Doering (1949).

TABLE 4.11

Results for several models fitted to data in Table 4.10

	ϕ	Model configuration	Maximized loglikelihood	Improvement over m.e. model	Goodness of fit
C1	1.8	Main effects (m.e.)	-1050.774	---	11.019 (10)
C2	1.2	m.e. + 24	-1049.820	1.908 (1)	9.111 (9)
C3	1.5	m.e. + 14 + 24	-1048.693	4.163 (2)	6.856 (8)
C4	2.5	m.e. + 14 + 24 + 34	-1046.655	8.239 (3)	2.780 (7)
C5	1.0	m.e. + 14 + 24 + 34 + 23	-1046.544	8.460 (4)	2.559 (6)
C6	0.5	m.e. + 23 + 123	-1050.496	0.557 (2)	10.462 (8)
C7	0	m.e. + 23 + 24	-1049.137	3.275 (2)	7.744 (8)
C8	3.6	m.e. + all two-factor interactions	-1046.354	8.842 (6)	2.177 (4)
		Saturated	-1045.265		

N.B. Degrees of freedom of corresponding asymptotic χ^2 appear in parentheses.

TABLE 4.12

Results from the fit of model C1 to data in Table 4.10

Observed	Expected	Chi-squared residuals	Generalized residuals
23	21.557	0.563	0.546
102	101.205	0.125	0.165
1	2.261	-1.272	-1.314
16	15.207	0.281	0.207
8	5.758	1.296	1.318
35	36.528	-0.323	-0.357
4	2.478	1.203	1.199
13	15.849	-0.827	-0.840
27	28.420	-0.439	-0.402
201	203.917	-0.301	-0.245
3	5.572	-1.551	-1.588
67	62.429	0.757	0.729
7	5.471	0.859	0.853
75	75.330	-0.046	-0.020
2	3.468	-0.935	-0.934
84	82.766	0.149	0.186

Pearson's chi-squared:	10.845	10 d.f.
L.R. chi-squared:	11.019	

TABLE 4.13

Sum of squares of residuals from the fit of model C1

Effect	Sum of squares	
	Chi-squared residuals	Generalized residuals
Radio (1)	0.4069	0.2811
Solid reading (2)	0.6451	0.5968
Newspapers (3)	0.9645	1.2334
Lectures (4)	0.0005	0.0081
(12)	0.0005	0.0071
(13)	0.0243	0.0100
(14)	1.6639	1.8993
(23)	0.0133	0.0002
(24)	3.0893	3.3345
(34)	2.0590	1.8760
(123)	0.4311	0.3521
(124)	0.3895	0.3599
(134)	0.4157	0.3596
(234)	0.4171	0.3048
(1234)	0.3082	0.3386

TABLE 4.14

Results from the fit of model C2 to data in Table 4.10

Observed	Expected	Chi-squared residuals	Generalized residuals
23	20.406	0.982	0.982
102	102.852	-0.134	-0.135
1	2.092	-1.093	-1.129
16	15.171	0.294	0.293
8	6.804	0.697	0.701
35	35.728	-0.155	-0.178
4	3.029	0.741	0.748
13	15.222	-0.654	-0.733
27	26.777	0.068	0.053
201	206.349	-0.553	-0.554
3	5.067	-1.250	-1.279
67	61.730	0.874	0.861
7	6.559	0.245	0.237
75	73.109	0.268	0.241
2	4.437	-1.457	-1.532
84	83.091	0.110	0.124

Pearson's chi-squared:	8.660	9 d.f.
L.R. chi-squared:	9.111	

TABLE 4.15

Sum of squares of residuals from the fit of model C2

Effect	Sum of squares	
	Chi-squared residuals	Generalized residuals
Radio (1)	0.3520	0.3594
Solid reading (2)	0.0230	0.0166
Newspapers (3)	0.9278	0.9970
Lectures (4)	0.0780	0.0809
(12)	0.0193	0.0181
(13)	0.0078	0.0094
(14)	1.6067	1.7213
(23)	0.0377	0.0394
(24)	0.3695	0.4026
(34)	2.4401	2.5090
(123)	0.6245	0.6178
(124)	0.4232	0.4872
(134)	0.3391	0.3788
(234)	1.1273	1.1214
(1234)	0.2129	0.2616

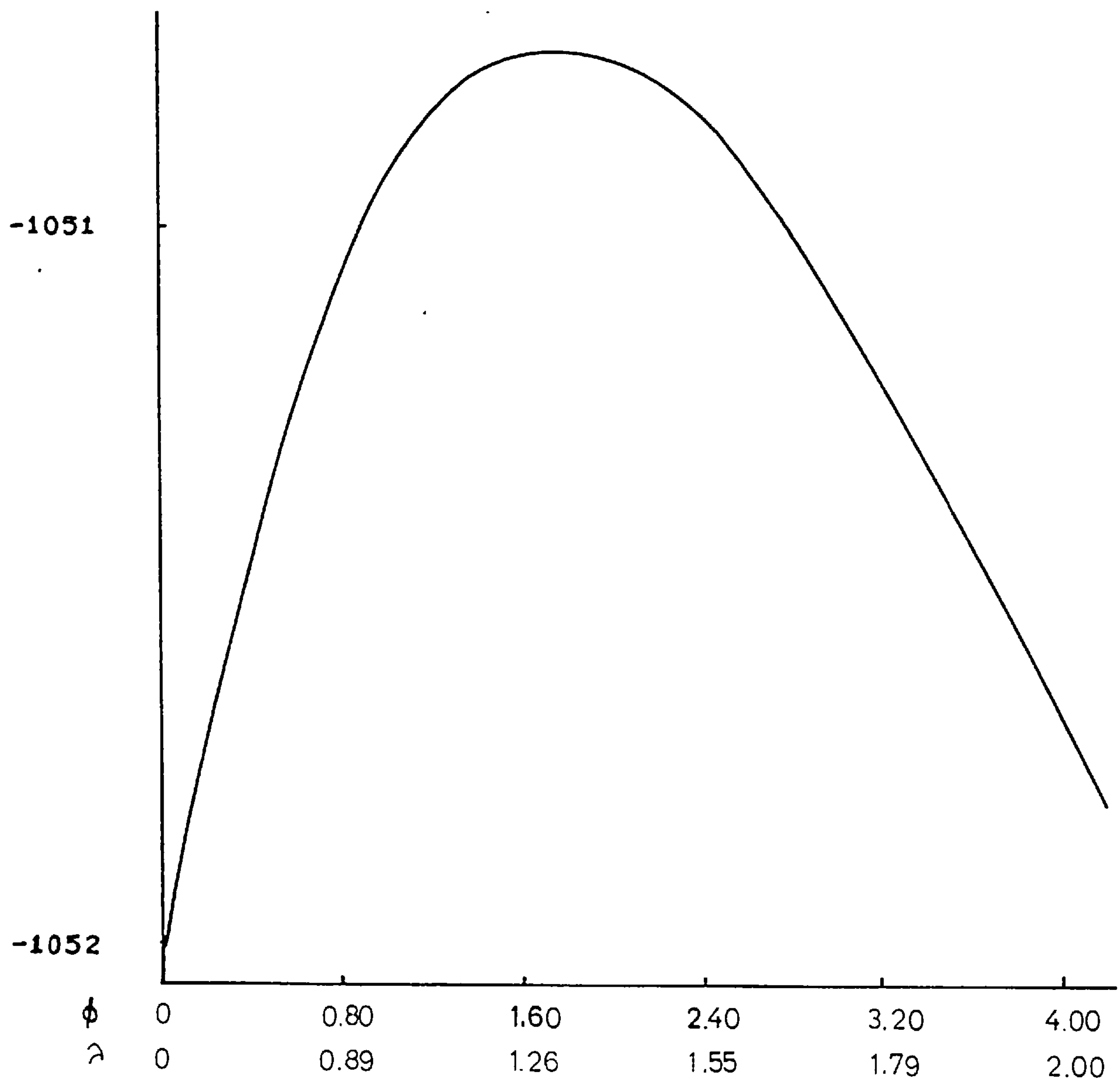

Figure 4.5

Loglikelihood curve for the transformation parameter.
Model configuration: main effects; data in Table 4.10.

4.5 One example of bioassay data

Data in Table 4.16 are due originally to Martin (1940) and have been analyzed by Finney (1952, p.94) and Ashton (1972, p.54). The data correspond to a test of toxicity of derris root in relation to the grain beetle Oryzaephilus surinamensis. We have not made adjustments to allow for natural mortality because we want to illustrate the use of the method of estimation of the transformation parameter, and to apply the test statistic introduced in §3.3, rather than to analyze thoroughly the data.

Our preliminary results suggested departures from the logistic model. Figure 4.6 shows the approximate loglikelihood curve for $\lambda = \phi^{1/2}$. Direct overall maximization provided $\hat{\lambda} = 0.5849$ which agrees with what we observe in Fig.4.6. The relation between the probability of response θ, and the log concentration of derris, x, that determines the dose is given by

$$T_{\lambda}(\theta) = \alpha + \beta x .$$

The estimates of the parameters from a GLIM fit with $\lambda = 0.5849$ are

$$\hat{\alpha} = -3.873\ (0.3595) \quad \text{and} \quad \hat{\beta} = 3.445\ (0.1897),$$

standard deviations appear in parentheses. The fit obtained is very good, the values for the deviance and Pearson's chi-squared statistics with one d.f. are 0.0154 and 0.0176, respectively (fitted values and residuals appear in Table 4.17).

An approximate confidence interval for λ at 90% level (shown in Fig. 4.6) does not include the logistic transformation. This agrees with the result from the GLIM fit for a logistic model that has a deviance of 4.72, which even considering 2 d.f. is significant at 10% level.

Given these results the test introduced in §3.3 was applied obtaining a value of 1.8566. From tables of the standard normal distribution this value is significant at 4% level; thus there is agreement with the previous discussion.

TABLE 4.16

Data for a toxicity test on derris

log concentration of derris (x)	number of insects killed	total exposed
1.08	58	126
1.68	115	128
2.00	126	127
2.17	142	142

Source: Ashton (1972)

TABLE 4.17

Results for model with $\lambda = 0.5849$

fitted to data in Table 4.16

Observed	Fitted	Residual
58	58.200	-0.036
115	114.801	0.058
126	125.898	0.098
142	141.999	0.024

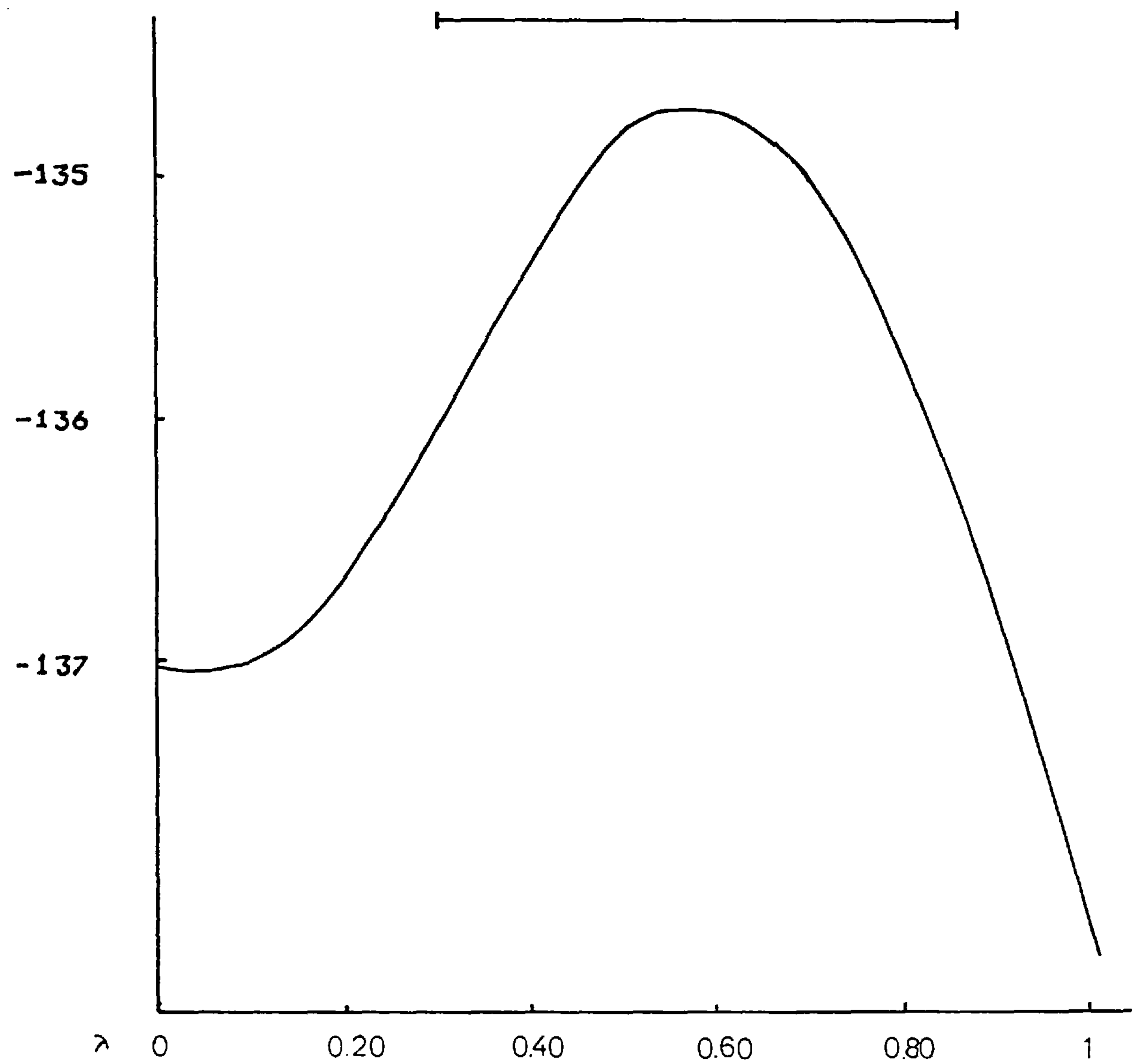

Figure 4.6

Loglikelihood curve for the transformation parameter. Simple model configuration; data in Table 4.16. An approximate 90 per cent confidence interval for λ is shown on the top of the figure.

Chapter 5. POLYTOMOUS AND MULTIVARIATE RESPONSES

5.1 Introduction

We have treated until now the case of binary or dichotomous response. The two natural extensions, namely polytomous and multivariate responses are discussed here, though we shall not go into great detail.

For polytomous responses it is usually necessary to distinguish between ordered and unordered responses. For a recent discussion of models for ordered responses see McCullagh (1980). We shall concentrate on unordered responses.

The multivariate case introduces new aspects into the analysis, namely the relationships among the responses and the way they are affected by the explanatory variables. If there are several of these, it is also of interest to study the marginal and joint effects on the responses.

Several of the models suggested for the two situations above are, either originally or under an equivalent formulation, based on extensions of the concept of an underlying distribution of tolerances used in the dichotomous case. Although for (0, 1) data this concept is often unnecessary, for some generalizations it serves to suggest models. For the multivariate case, in particular, the assumption of a multinormal distribution has been used, mainly because of the analytical simplicity it provides to model situations with several variables. However, the computation of multinormal probabilities in dimensions higher than 2 is not easy, it may be preferable to work directly with the original probabilities or transformations of them.

In general, the objective is to achieve a simple representation in terms of explanatory variables of the probabilities of getting particular responses.

5.2 Unordered polytomous responses

Cox (1966, 1970 ch. 7) extends the logistic model to polytomous unordered responses. If the possible responses are labelled arbitrarily 1, ..., k, the probability of the jth response (j = 1,...,k) may be represented as

$$\Pr\{Y = j\} \quad = \quad C^{-1} \exp(\underset{\sim}{\beta}'_j \, x) \qquad (5.2.1)$$

where $\underset{\sim}{\beta}_j$ is the vector of parameters for the jth response, Y is the response variable, $\underset{\sim}{x}$ is the vector of explanatory variables and C^{-1} is a normalizing constant. Cox suggests imposing the constraint $\underset{\sim}{\beta}_1 = 0$ to make the parameters unique; the first response may, for example, be taken as the relatively most frequent. This model is considered independently by Mantel (1966a).

Bock (1970) treats essentially the same model in more general terms providing a Newton-Raphson procedure to carry out the estimation by maximum likelihood methods.

Another type of model is based on a sequential argument, e.g. in the case of a response, Y, with three categories we may determine first whether Y = 3 or Y ≠ 3, and then, given Y ≠ 3 determine whether Y = 2 or Y = 1. This is advantageous for estimation because the likelihood function can be maximized by maximizing the likelihood of the binary case repeatedly.

Other ways of assigning the probabilities to obtain a certain response have been suggested; see for example Aitchison and Bennett (1970), where the response is determined from the outcome of some hypothetical auxiliary experiments. In general, there is implicit an exhaustive and exclusive partition of a probability space associated with the values taken by the polytomous response.

A natural generalization of the results on the symmetric family of transformations in Chapter 2 is to consider the expression

$$\theta_j = \frac{(1 + \lambda\tau_j)^{1/\lambda}}{\sum_{g=1}^{k} (1 + \lambda\tau_g)^{1/\lambda}} \qquad -1 < \lambda\tau_j, \qquad (5.2.2)$$

where θ_j denotes the probability of getting response j ($j = 1,\ldots,m$), and $\tau_j = \underset{\sim}{\beta}'_j x$.

Expression (5.2.2) reduces for $\lambda = 0$ to

$$\theta_j = \frac{\exp(\tau_j)}{\sum_{g=1}^{k} \exp(\tau_j)} ;$$

then we obtain the generalized logistic model in (5.2.1). For $\lambda = 1$ (5.2.2) takes the form

$$\theta_j = \frac{1 + \tau_j}{\sum_{g=1}^{k} (1 + \tau_g)} ,$$

which, subject to the constraint $\sum_{g=1}^{k} \tau_g = 0$, is

$$\theta_j = \frac{1 + \tau_j}{k} ,$$

or equivalently

$$k\theta_j - 1 = \tau_j ,$$

i.e. a linear model for the cell probabilities. The constraint on the parameters suggested above is automatically fulfilled in the binary case.

We have a natural extension of the family $\mathcal{T}$ for polytomous response. To obtain the form of the transformation associated with (5.2.2) we consider the relative odds θ_j/θ_k. Thus we obtain

$$\theta_j/\theta_k = [(1 + \lambda\tau_j)/(1 + \lambda\tau_k)]^{1/\lambda} \qquad (j = 1,\ldots,k-1),$$

from which we get

$$\frac{1}{\lambda} \frac{k\theta_j^{\lambda} - \sum_{g=1}^{k} \theta_g^{\lambda}}{\sum_{g=1}^{k} \theta_g^{\lambda}} = \tau_j , \tag{5.2.3}$$

for $j = 1,\ldots,k$.

For $\lambda = 1$ (5.2.3) reduces to

$$k\theta_j - 1 = \tau_j ,$$

whereas for $\lambda = 0$

$$\ln \frac{\theta_j}{(\prod_g \theta_g)^{1/k}} = \tau_j .$$

The last expression is invariant under rescaling of the θ's, thus it is equivalent to working with $\theta_j^* = c\theta_j$ where $\prod_{g=1}^{k} \theta_g^* = 1$. Hence, we may use the expression

$$\ln \theta_j^* = \tau_j^*,$$

the actual parameterization for θ's differs from the one above only by an additive constant which may be obtained using $\sum_{j=1}^{k} \theta_j = 1$ and then $\sum_{j=1}^{k} c\theta_j = c$. In our case the constant is

$$c = \left(\prod_{g=1}^{k} \theta_g\right)^{-1/k} .$$

The implication of the above result is that the linear systematic part of a loglinear fit will differ from the one needed in our model only by an additive constant. Then the fitting of the model may be carried out by a modified loglinear fit.

In principle it is possible to develop a score test, analogous to the one in Chapter 3, for departures from the generalized logistic model in the direction of the linear model. The details are not given because of their algebraic complexity.

An alternative is to construct auxiliary variables by considering the response in each category of the polytomous variable as binary in relation with its complement. Different models may be fitted to the variables so constructed. Comparison of the most suitable models, configuration and scale, for each individual variable may shed some light on the way the explanatory variables influence the polytomous response.

5.3 Multivariate responses

Cox (1966, 1970 §7.6) suggests also one possible way to analyze multivariate responses. He discusses, in particular, a pair of binary responses. In this case a variable with four levels of response is created and then the problem reduces to one of polytomous response. This approach may be generalized to more complex situations.

The objective may be stated as to get a description of the regression of the responses on the explanatory variables as concisely as possible. This may be compared with the aim in canonical regression or canonical correlation analysis for quantitative responses.

Nerlove and Press (1973) present a comprehensive discussion of models for jointly dependent qualitative response variables. They interpret the resulting conditional probabilities as analogues of structural equations in systems of simultaneous equations used in econometrics. A general model for the analysis where quantitative and qualitative explanatory variables are used is suggested. The listing of a computer program to fit the model by maximum likelihood methods is provided.

Ashford and Sowden (1970) use the idea of an underlying distribution of tolerances. They construct a jointly normal model from marginal distributions conveniently specified. Their treatment is for binary responses only. Mantel and Brown (1973) consider an alternative logistic analysis of the previous example. It seems that the model based on the normal distribution provides better results.

For multivariate binary data one possible alternative might be to apply marginally transformations of the type suggested in Chapter 2. The values of the transformation parameters appropriate for an additive decomposition of each response may be compared. These values may serve to study similarities and differences in how the explanatory variables affect the responses.

Another possibility would be to construct a polytomous response as suggested by Cox (1966). The generalization of the family $\mathcal{T}$ suggested in the last section could be applied then. This procedure may be considered as a joint transformation of the original variates.

The last two suggested possibilities would provide complementary information about the data. Thus, it seems that they might both be tried.

Chapter 6. TRANSFORMATIONS FOR PROBABILITIES AND SURVIVAL DATA ANALYSIS

6.1 Introduction

The analysis of survival data arises in diverse disciplines, e.g. medical and actuarial studies, industrial life testing, social sciences, etc. The interest is centred on failure time, T. Two intimately related ways of describing the behaviour of T are the survivor function, $\mathcal{F}(t)$, and the hazard function, h(t), defined respectively as follows:

$$\mathcal{F}(t) = \Pr[T > t] = \int_t^\infty f_T(t)dt,$$

where $f_T(t)$ denotes the probability density function of T, and

$$h(t) = \lim_{\delta \downarrow 0} \delta^{-1} \Pr\{t < T \leq t + \delta \mid T > t\}$$

$$= -d \ln \mathcal{F}(t)/dt.$$

These definitions are for continuous time. There are corresponding discrete versions (Cox, 1972).

One of the main problems in the analysis of survival data is to study the effect of explanatory variables on T. The most convenient way to do so is perhaps to establish simple models for the hazard function in terms of the explanatory variables. We focus our attention on two classes of model, namely additive and multiplicative. The approach taken here is to analyze sequences of contingency tables constructed from the data. This is the essence of the common life table analysis. In his pioneering paper Mantel (1966b) uses a similar approach; though the way we determine the sequences parallels Holford (1976). Certain similarities with the problems studied above are exploited. In particular, a family of transformations for probabilities is defined to include both additive and multiplicative models as special cases. Generalized linear models associated with the transformations are tried on the data and their fits compared. Two simple estimates of the survivor function are computed, and a test for departures from the multiplicative model in the direction of the additive model is suggested.

6.2 Additive and multiplicative models for the hazard function

6.2.1 The models

We consider two classes of model for the hazard function. These models differ in the way the explanatory variables are assumed to influence the underlying hazard, and may be expressed as follows:

I) The Additive model where

$$h(t;\ \underset{\sim}{z}) = \rho(t) + k(\underset{\sim}{z}), \qquad \text{and}$$

II) The Multiplicative model where

$$h(t;\ \underset{\sim}{z}) = \rho(t)g(\underset{\sim}{z}),$$

where $h(t;\ \underset{\sim}{z})$ denotes the hazard function given the vector of explanatory variables $\underset{\sim}{z}$, $\rho(t)$ is the underlying hazard when $k(\underset{\sim}{z}) \equiv 0$ or $g(\underset{\sim}{z}) \equiv 1$ depending on the model, and $k(\underset{\sim}{z})$ and $g(\underset{\sim}{z})$ are parametric functions of the explanatory variables. We use throughout two more specialized versions of (I) and (II), namely

$$h(t;\ \underset{\sim}{z}) = \rho(t) + \underset{\sim}{\alpha}'\underset{\sim}{z}, \tag{6.2.1}$$

and

$$h(t;\ \underset{\sim}{z}) = \rho(t) \quad \exp(\underset{\sim}{\beta}'\underset{\sim}{z}), \tag{6.2.2}$$

where $\underset{\sim}{\alpha}$ and $\underset{\sim}{\beta}$ are vectors of unknown parameters.

The use of additive models may be motivated as follows, suppose that the surviving process is controlled by several factors. Assume there is a finite (possibly large) number, n, of these factors. Suppose further that each factor, j, has its own lifetime, T_j, and a corresponding hazard function $h_j(t) > 0$. Suppose also that a parallel system of failure holds. Then the actual failure time is $T = \min(T_1,\ldots,T_n)$ and the overall hazard function is the sum of the hazard functions of the controlling factors. The factors themselves might not be directly observed. Consider instead p explanatory variables $z_1,\ldots,z_p$ which are associated with, or represent, the controlling factors. It seems appealing, but not compulsory, to represent the hazard function additively in terms of the z's, particularly if say each z affects only one or two of the individual hazard functions.

Multiplicative models have received a great deal of attention recently, one of the reasons being the analytical simplicity they provide. Nevertheless this convenience must not be the sole guide for choosing a model for the hazard function.

The appropriateness of the assumption of proportional hazards, implicit in the use of multiplicative models, deserves examination, particularly when strong regression effects are present. When proportionality does not seem to hold, one possible alternative is to include a time dependent explanatory variable in the model. Thus, the effect of this variable on the hazard may change smoothly over time. This procedure increases the difficulty of computation. We consider below a different way to deal with this situation.

6.2.2 Grouping continuous data

It is usually reasonable to consider survival time as essentially a continuous random variable, thus ruling out the possibility of tied failure times. However, ties frequently result from the way data are recorded. To cope with situations with an appreciable number of ties, Cox (1972) proposes a linear logistic model for the analysis in discrete time. This model provides a first order approximation to model (6.2.2). Kalbfleisch and Prentice (1973) point out some disadvantages of such a model, namely the heavy computation arising when many ties are present and the fact that the estimated parameters do not correspond exactly with the ones for the underlying continuous model. They propose to group the model (6.2.2) obtaining, essentially, a generalized linear model with a complementary log log link function. This model retains the parameters to be estimated in the continuous case.

Although in this instance grouping has been introduced for technical reasons, there are situations where it cannot be avoided. For example, consider certain types of medical follow-up studies where failure is not death, but rather time until occurrence of certain infection, tooth decay, or in general a "soft" end-point which may not be immediately obvious or clearly defined. Besides, the level of measurement used for recording certain events, e.g., weeks instead of days for re-engagement in remunerated work, may be such that with sufficiently large samples the data can be regarded as grouped and analyzed accordingly. Even when the exact failure times are known, the times of occurrence of other events, e.g. follow-up loss, may only be known on an interval basis forcing the data to be reported as grouped.

We introduce the partition $\{I_i\}$ $(i = 1,\dots,m)$ of the time scale where $I_i = (t_{i-1}, t_i]$, $t_0 = 0$ and t_m is defined such that no failure is observed after t_m. We obtain below an expression for the conditional probability, say θ_i, that an individual fails during the ith study interval, I_i, given that he was alive at time t_{i-1}. This may be expressed as follows

$$\theta_i = \Pr\{t_{i-1} < T \leq t_i \mid T > t_{i-1}\}. \qquad (6.2.3)$$

where for simplicity the dependence on $\underset{\sim}{z}$ has not been made explicit.

In terms of the survivor function (6.2.3) is expressed as

$$\theta_i = 1 - \mathcal{F}(t_i)/\mathcal{F}(t_{i-1}) \ . \qquad (6.2.4)$$

This expression takes different forms depending on the representation assumed for the hazard function. We use the following relationship between the hazard and the survivor functions

$$\mathcal{F}(t) = \exp\{-\int_0^t h(u)du\} \ .$$

We assume that the explanatory variables do not vary with time within a partition interval, though they may vary from one interval to another. Hence for a model with hazard given by (6.2.2) expression (6.2.4) becomes

$$\theta_i(\underset{\sim}{z}) = 1 - \exp\{-\exp(\underset{\sim}{\beta}'\underset{\sim}{z}) \int_{t_{i-1}}^{t_i} \rho(u)du \} \ , \qquad (6.2.5)$$

where we now make explicit the supposed dependence on $\underset{\sim}{z}$. The last expression may be rewritten as

$$\ln\{1 - \theta_i(\underset{\sim}{z})\} = - \exp(\underset{\sim}{\beta}'\underset{\sim}{z}) \int_{t_{i-1}}^{t_i} \rho(u)du \ , \qquad (6.2.6)$$

from which we obtain

$$\ln\{1 - \theta_i(\underset{\sim}{z})\} = \exp(\underset{\sim}{\beta}'\underset{\sim}{z})\ln\{1 - \theta_i(\underset{\sim}{0})\} ,$$

and taking logarithms again we get

$$\ln[-\ln\{1 - \theta_i(\underset{\sim}{z})\}] = \ln[-\ln\{1 - \theta_i(\underset{\sim}{0})\}] + \underset{\sim}{\beta}'z , \tag{6.2.7}$$

Analogously for (6.2.1) we have

$$-\ln\{1 - \theta_i(\underset{\sim}{z})\} = -\ln\{1 - \theta_i(\underset{\sim}{0})\} + \underset{\sim}{\alpha}'\underset{\sim}{z}(t_i - t_{i-1}). \tag{6.2.8}$$

These are just special cases of more general expressions corresponding to models (I) and (II). Simple linear expressions for the dependence on z are convenient to use, and may be considered as approximations to the actual ones when these are more complex. Hence, we restrict our attention to linear expressions.

6.2.3 A family of asymmetric transformations

Expressions (6.2.7) and (6.2.8) involve asymmetric transformations of the probability $\theta_i(z)$. Suppose there is a family of transformations that includes those expressions as special cases. We suggest basing on such family the assessment of a scale where an additive model for the hazard function is consistent with the data. One possible candidate is a subfamily of $\mathcal{V}$, defined in §2.5, when we fix $\beta = 1$ and let λ to vary arbitrarily, namely

$$V(\theta) = \frac{\{-\ln(1 - \theta)\}^{\lambda} - 1}{\lambda} . \tag{6.2.9}$$

This expression reduces to the complementary log log transformation for $\lambda = 0$, and to the negative of the complementary log for $\lambda = 1$. Hence, it may be used to define the comprehensive model suggested above.

6.2.4 A general model for grouped data

We assume that $(t_i - t_{i-1}) = \Delta_i$ $(i = 1,\ldots,m)$ is constant, i.e. we consider partition intervals of equal width. Then (6.2.8) may be written as follows

$$-\ln\{1 - \theta_i(\underset{\sim}{z})\} = -\ln\{1 - \theta_i(\underset{\sim}{0})\} + \underset{\sim}{\nu}'\underset{\sim}{z} , \qquad (6.2.10)$$

where $\underset{\sim}{\nu}$ incorporates the common factor, say Δ.

Thus, we may write formally that

$$V_\lambda\{\theta_i(\underset{\sim}{z})\} = V_\lambda\{\theta_i(\underset{\sim}{0})\} + \underset{\sim}{\gamma}'\underset{\sim}{z} ,$$

which incorporates (6.2.8) and (6.2.10) as special cases. To carry out the estimation procedure it is more convenient to use the expression

$$V_\lambda\{\theta_i(\underset{\sim}{z})\} = \zeta_i + \underset{\sim}{\gamma}'\underset{\sim}{z} , \qquad (6.2.11)$$

which may be easily inverted obtaining

$$\theta_i(\underset{\sim}{z}) = 1 - \exp\{-(1 + \lambda\eta_i)^{1/\lambda}\}, \qquad (6.2.12)$$

that holds as long as $\lambda\eta_i > -1$; otherwise $\theta_i(\underset{\sim}{z}) = 0$, $\eta_i = \zeta_i + \underset{\sim}{\gamma}'\underset{\sim}{z}$, $(i = 1,\ldots,m)$.

We use (6.2.12) to determine the link function of a GLM for this situation. A scale where an additive expression for the data is appropriate may be estimated from an approximate log likelihood curve obtained by fitting models for several values of λ.

If we do not assume equal width intervals, the treatment given above is not possible. One way to overcome this problem is to consider the underlying hazard function as constant within each interval. If $\rho(u) = k_i$ for $t_{i-1} < u \leq t_i$, expressions (6.2.7) and (6.2.8) may be written respectively as follows

$$-\ln\{1 - \theta_i(\underset{\sim}{z})\} = \exp(k_i' + \underset{\sim}{\beta}'\underset{\sim}{z})\Delta_i , \qquad (6.2.13)$$

and

$$-\ln\{1 - \theta_i(\underset{\sim}{z})\} = (k_i + \underset{\sim}{\alpha}'\underset{\sim}{z})\Delta_i , \qquad (6.2.14)$$

where $\Delta_i = t_i - t_{i-1}$, $k_i' = \ln k_i$.

These expressions are special cases of

$$-\ln\{1 - \theta_i(\underset{\sim}{z})\} = \{1 + \lambda(\zeta_i + \underset{\sim}{\gamma}'\underset{\sim}{z})\}^{1/\lambda}\Delta_i,$$

or equivalently

$$\theta_i(\underset{\sim}{z}) = 1 - \exp\{-[1 + \lambda(\zeta_i + \underset{\sim}{\gamma}'\underset{\sim}{z})]^{1/\lambda}\Delta_i\} ,$$

which holds for $\lambda(\zeta_i + \underset{\sim}{\gamma}'\underset{\sim}{z}) > -1$, otherwise $\theta_i(\underset{\sim}{z}) = 0$.

This expression may be used instead of (6.2.12) to define a GLM. However, it is normally best to use intervals of equal width, with the exception of situations where certain time periods are of special interest and a non-homogeneous partition is required.

The assumption about the constancy of $\rho(t)$ over prespecified intervals has been made by several authors, see for instance Kalbfleisch and Prentice (1973) and Kay (1977). Although a bit artificial it is a useful device. Another way to obtain similar results is to apply the mean value theorem to the integral

$$\int_{t_{i-1}}^{t_i} \rho(u)du ,$$

obtaining

$$\int_{t_{i-1}}^{t_i} \rho(u)du = \rho(u')\Delta_i , \quad (i = 1,\ldots,m)$$

for $u' \in [t_{i-1}, t_i]$.

Expressions (6.2.13) and (6.2.14) are obtained without assuming constancy of $\rho(t)$ within intervals. This assumption is needed if a valid step function estimate of $\rho(t)$ is required, otherwise it is just a computational device.

One possible way to make the assumption more reasonable is to use very fine partitions. However, the number of parameters in the model grows accordingly and computational problems may be faced because of the infeasibility of optimizing functions that depend on a large number of variables. More importantly, the inference about the parameters of main interest becomes less precise. We recommend to use a moderate number of intervals. It is not possible to suggest specific numbers but a compromise should be made between the amount of data available and the number of time intervals considered.

6.3 Applications of the family of transformations

6.3.1 Estimation of the survivor function

We proceed to estimate the survivor function for a vector $\underset{\sim}{z}$ of explanatory variables. Its expression in terms of the conditional probabilities of survival is

$$\mathcal{F}(t) = \prod_{I(t)} \{1 - \theta_i(\underset{\sim}{z})\} \quad ,$$

where $I(t) = \{i | t_i < t\}$ and t_i is the upper bound of the ith time interval. The survivor function at t may be estimated via the values of $\hat{\theta}_i(\underset{\sim}{z})$ which depend on $\hat{\underset{\sim}{\zeta}}(\lambda)$ and $\hat{\underset{\sim}{\gamma}}(\lambda)$, the m.l.e. of $\underset{\sim}{\zeta}$ and $\underset{\sim}{\gamma}$ for a fixed value of the parameter λ. There are two cases of special interest, namely $\lambda = 0$ and $\lambda = 1$.

For $\lambda = 0$ we have

$$\hat{\mathcal{F}}(t) = \prod_{I(t)} \exp\{-\exp(\hat{\zeta}_i(0) + \hat{\underset{\sim}{\gamma}}'(0)\underset{\sim}{z})\},$$

which for $\underset{\sim}{z}$ not depending on time may be written

$$\ln \hat{\mathcal{F}}(t) = \exp(\hat{\underset{\sim}{\gamma}}'(0)\underset{\sim}{z}) \ln \hat{\mathcal{F}}_0(t) \ , \tag{6.3.1}$$

where $\hat{\mathcal{F}}_0(t)$ is the estimator of the underlying survivor function under the proportional hazards assumption.

Analogously for $\lambda = 1$ we have

$$\ln \hat{\mathcal{F}}(t) = -\hat{\underset{\sim}{\gamma}}'(1)\underset{\sim}{z} + \ln \hat{\mathcal{F}}_0(t) \ , \tag{6.3.2}$$

in this case $\hat{\mathcal{F}}_0(t)$ is the estimator of the underlying survivor function under the assumption of hazards with constant difference.

In (6.3.1) and (6.3.2) we may observe the different way in which the explanatory variables affect the log survivor function.

The expression for $\hat{\mathcal{F}}(t)$ provides a step function estimate of the survivor function. A continuous estimate may be preferable, for instance to communicate results or for assessing the form of the underlying hazard. We derive a continuous estimator for arbitrary t as follows. Consider the survivor function at t ($t < t_m$) written in the following way

$$\mathcal{F}(t) = \exp\{-\int_0^{t_j} h(u)du - \int_{t_j}^{t} h(u)du\} ,$$

where t_j is the largest partition bound less than t. Then,

$$\mathcal{F}(t) = \mathcal{F}(t_j)\exp\{-\int_{t_j}^{t} h(u)du\},$$

if we assume $h(u) = k_{j+1}$ for $u \in I_{j+1}$, we obtain

$$\begin{aligned}\exp\{-\int_{t_j}^{t} h(u)du\} &= \exp\{-k_{j+1}(t-t_j)\} \\ &= (1 - \theta_{j+1})^{(t-t_j)/(t_{j+1}-t_j)} ,\end{aligned}$$

because $1 - \theta_{j+1} = \exp\{-\int_{t_j}^{t_{j+1}} h(u)du\}$.

Then,

$$\mathcal{F}(t) = \mathcal{F}(t_j)(1 - \theta_{j+1})^{(t-t_j)/\Delta_{j+1}} ,$$

$\Delta_{j+1} = t_{j+1} - t_j$,
substituting $\hat{\mathcal{F}}(t)$ for $\mathcal{F}(t)$ and taking logs we have

$$\ln \hat{\mathcal{F}}(t) = \ln \hat{\mathcal{F}}(t_j) + [(t-t_j)\ln(1 - \theta_{j+1})]/\Delta_{j+1} .$$

This expression provides an estimator consisting in a connected sequence of straight lines.

6.3.2 Test for departures from the multiplicative model, the two-sample case

For simplicity we treat first the two-sample problem. As before the comparison of models is based on the maximized loglikelihood achieved for different values of the transformation parameter λ.

Suppose there are initially $n = n_1 + n_2$ individuals at risk of failure, n_j in group j (j = 1,2). The individuals at risk during the ith period may fail, be censored or survive to the start of the following period. Here we work with the frequencies of the various categories and follow Cox (1975) in the discussion.

Let n_{ij} denote the observed number at risk in the ith interval for the jth group, the observed number of individuals failing in each group is denoted by f_{ij} and the corresponding number of censored individuals by c_{ij}. Thus $n_{(i+1)j} = n_{ij} - f_{ij} - c_{ij}$ $(1 < i+1 \leq m;\ j = 1,2)$ and $n_{1j} = n_j$. Censoring is assumed to take place instantaneously at the end of the interval. Thus, the number of individuals at risk throughout the ith interval, jth group is n_{ij}. This assumption is relaxed later. It may be considered alternatively that censoring occurs instantaneously at the beginning of the interval. If there is a large discrepancy in the results obtained under the two assumptions another method may be used, e.g. the so-called actuarial method.

We denote by θ_{ij} the probability that an individual, at risk in the jth group, fails during the ith interval having survived until the beginning of that interval. It is supposed that failures occur independently.

The probabilities θ_{ij} are assumed to be relevant to the whole population of individuals under study, i.e. if an individual in sample j censored in an earlier period had survived to the start of the ith interval, his probability of failure would have been θ_{ij}. Thus, censoring and failure are assumed to be determined by independent mechanisms.

Let the random variable C_{ij} represent the number of individuals censored in the jth sample just before the end of the (i-1)th interval ($C_{1j} = 0$), and F_{ij} represent the number of failures in the ith interval for sample j. Then the partial likelihood, hereafter referred to as the likelihood, as defined by Cox (1975), based on F_1, F_2 in the sequences $\{C_{i1}, F_{i1}\}$ and $\{C_{i2}, F_{i2}\}$ is

$$\prod_{i=1}^{m} \prod_{j=1}^{2} \binom{n_{ij}}{f_{ij}} \theta_{ij}^{f_{ij}} (1 - \theta_{ij})^{n_{ij}-f_{ij}} . \tag{6.3.3}$$

Our objective is to represent the probabilities $\{\theta_{11}, \theta_{12}, \ldots, \theta_{m1}, \theta_{m2}\}$ in terms of a parameter which takes account of the difference between groups and some auxiliary parameters for the intervals. An indicator variable z (1 for sample 2, 0 otherwise) may be used in a model based on (6.2.11). Then, except for a constant, the loglikelihood may be written as

$$\ell = \sum_{i=1}^{m} \sum_{j=1}^{2} \{f_{ij} \ln \theta_{ij} + (n_{ij}-f_{ij}) \ln(1 - \theta_{ij})\} ,$$

where

$$\theta_{ij} = \begin{cases} 1 - \exp\{-(1+\lambda\eta_{ij})^{1/\lambda}\} & \lambda\eta_{ij} \geq -1 \\ 0 & \text{otherwise} \end{cases} ,$$

and

$$\eta_{ij} = \zeta_i + \gamma z, \qquad z = \begin{cases} 0 & j = 1 \\ 1 & j = 2 \end{cases} .$$

In this particular case a test of the multiplicative model assumption may be obtained examining a statistic based on the efficient score $U(\lambda) = \partial\ell/\partial\lambda$ at $\lambda = 0$. Here

$$U(\lambda) = \sum_{i=1}^{m} \sum_{j=1}^{2} \left\{ \frac{f_{ij} - n_{ij}[1 - \exp(-s_{ij}^{1/\lambda})]}{1 - \exp(-s_{ij})^{1/\lambda}} \right\}$$

$$\times \left\{ \frac{\eta_{ij}}{\lambda s_{ij}} - \frac{1}{\lambda^2} \ln s_{ij} \right\} s_{ij}^{1/\lambda} ,$$

where $\eta_{ij} = \zeta_i + \gamma z$, $s_{ij} = (1 + \lambda\eta_{ij})$, z as above.

For $\lambda = 0$ the expression above takes the limiting value

$$U(0) = - \sum_{i=1}^{m} \sum_{j=1}^{2} \frac{f_{ij} - n_{ij}\theta_{ij}^{o}}{\theta_{ij}^{o}} \frac{\eta_{ij}^{2}}{2} \exp(\eta_{ij}) ,$$

where $\theta_{ij}^{o} = 1 - \exp\{-\exp(\eta_{ij})\}$.

In principle we must make allowance for the presence of the nuisance parameters $\underset{\sim}{\zeta}$ and γ. If these are replaced by their m.l.e. when $\lambda = 0$, by the same argument used in §3.3, we just need to consider the efficient score U(0) and its standard deviation to compute the test statistic. Thus, we need the expressions of elements of the information matrix I for the limiting case $\lambda = 0$, these are given below

$$I_{\zeta_s \zeta_r} = \begin{cases} \sum_{j=1}^{2} n_{sj}(1 - \hat{\theta}_{sj}^{o})\exp(2\eta_{sj})/\hat{\theta}_{sj}^{o} & r = s, \\ 0 & r \neq s , \end{cases}$$

$$I_{\gamma\gamma} = \sum_{i=1}^{m} n_{i2}(1 - \hat{\theta}^{o}_{i2})\exp(2\eta_{i2})/\hat{\theta}^{o}_{i2} ,$$

$$I_{\gamma\zeta_s} = \exp(2\eta_{s2})(1 - \hat{\theta}^{o}_{s2})n_{s2}/\hat{\theta}^{o}_{s2} ,$$

$$I_{\gamma\lambda} = \sum_{i=1}^{m} n_{i2}(1 - \hat{\theta}^{o}_{i2})\exp(2\eta_{i2})\eta^{2}_{i2}/(2\hat{\theta}^{o}_{i2}) ,$$

$$I_{\lambda\lambda} = \sum_{i=1}^{m} \sum_{j=1}^{2} \{n_{ij}(1 - \hat{\theta}^{o}_{ij})\exp(2\eta_{i2})\eta^{4}_{ij}/(4\hat{\theta}^{o}_{ij}) \} .$$

The variance of the test statistic A is given by

$$\mathrm{Var}(A) = I_{\lambda\lambda} - I_{\lambda\underset{\sim}{\phi}} I^{-1}_{\underset{\sim}{\phi}\underset{\sim}{\phi}} I_{\underset{\sim}{\phi}\lambda} ,$$

where $\underset{\sim}{\phi} = (\underset{\sim}{\zeta},\gamma)$. The inversion of $I_{\underset{\sim}{\phi}\underset{\sim}{\phi}}$ is greatly simplified because the submatrix $I_{\underset{\sim}{\zeta}\underset{\sim}{\zeta}}$, corresponding to the interval effects, $\underset{\sim}{\zeta}$, is diagonal. Then applying the formula for inversion of partitioned symmetric matrices (Rao, 1973, p.33), we obtain the required inverse in a simple way.

An asymptotic normal distribution is assumed for the standardized form of A. Because we are interested in alternatives $\lambda > 0$, we reject $\lambda = 0$ if we observe large values of A.

6.3.3 The general case

The treatment above may be generalized to allow variation in the explanatory variables over individuals. We denote by S_i, F_i, C_i the sets of individuals who respectively survive, fail or are censored in the ith interval; m partition intervals are considered as above. An individual, say j, contributes to the likelihood in the ith interval, I_i, one of the following factors,

i) $1 - \mathcal{F}_j(t_i)/\mathcal{F}_j(t_{i-1}) = \theta_{ij}$ if he fails in I_i,

ii) $\mathcal{F}_j(t_i)/\mathcal{F}_j(t_{i-1}) = 1 - \theta_{ij}$ if he survives I_i, or

iii) $\mathcal{F}_j(c_j)/\mathcal{F}_j(t_{i-1})$ if he is censored at $t_{i-1}<c_j<t_i$.

We consider the hazard function as constant over I_i to obtain an approximation of the factor in (iii). Hence we have

$$\mathcal{F}_j(c_j)/\mathcal{F}_j(t_{i-1}) = \exp\{-k_i(c_j-t_{i-1})\} \quad ,$$

where k_i is the supposed constant value for the hazard over I_i. Using (6.2.4) we may write the above expression as follows

$$\mathcal{F}_j(c_j)/\mathcal{F}_j(t_{i-1}) = (1 - \theta_{ij})^{(c_j-t_{i-1})/(t_i-t_{i-1})} \quad ,$$

and assuming a uniform distribution for c_j, the contribution of the censored individual, j, in the ith interval is approximated by

$$(1 - \theta_{ij})^{1/2} \ .$$

Then we may write the loglikelihood

$$\ell = \sum_{i=1}^{m} \{ \sum_{j \in F_i} \ln \theta_{ij} + \sum_{j \in S_i} \ln(1-\theta_{ij}) + \frac{1}{2} \sum_{j \in C_i} \ln(1-\theta_{ij})\} \ ; \qquad (6.3.4)$$

this expression may be maximized for different values of λ to obtain an approximate estimate $\hat{\lambda}$.

For the two-sample case (6.3.4) reduces to the well known actuarial rule, that is to approximate the number of individuals at risk in a specific time interval as the initial total minus half the number of censored observations during that interval.

It is convenient to work with the separate contribution to the likelihood from an individual who is censored or fails during I_k $(k = 1,\ldots,m)$, namely

$$\{\theta_k(\underset{\sim}{z}_j)\}^{\delta}\{1 - \theta_k(\underset{\sim}{z}_j)\}^{(1-\delta)/2} \prod_{i=1}^{k-1} \{1 - \theta_i(\underset{\sim}{z}_j)\} , \qquad (6.3.5)$$

where δ is zero for censored individuals and one for failures, $j = 1,\ldots,N$, and N is the total number of individuals at the start of the study.

The analysis is similar to the one used above in the two-sample case. The asymptotically optimal test introduced above has its counterpart under this set up. The expression for the score statistic for $\lambda = 0$ is

$$U(0) = -\sum_{i=1}^{m} \{ \sum_{j\in F_i} \frac{r_{ij}t_{ij}}{1-r_{ij}} (\frac{\eta_{ij}^2}{2}) + \sum_{j\in S_i} t_{ij}(\frac{\eta_{ij}^2}{2}) + \frac{1}{2} \sum_{j\in C_i} t_{ij}(\frac{\eta_{ij}^2}{2}) \} ,$$

where $r_{ij} = \exp\{-\exp(\eta_{ij})\}$ and $t_{ij} = \exp(\eta_{ij})$, $\eta_{ij} = \zeta_i + \underset{\sim}{\gamma}'\underset{\sim}{z}_j$.

The variance of the statistic is obtained from the information matrix when $\lambda = 0$. In general, it is not possible to obtain the expected value of the second order derivatives of the loglikelihood, but the observed values may be used instead, i.e., the empirical Fisher's information matrix may be employed. We show below the form of the individual contributions to the components of this matrix, namely

$$-\frac{\partial^2 \ell_{.j}}{\partial \zeta_h^2} = \begin{cases} d_h & h < k \\ \delta g_h c_h - (1-\delta) d_h/2 & h = k \\ 0 & h > k\,, \end{cases}$$

$$-\frac{\partial^2 \ell_{.j}}{\partial \gamma_s \partial \zeta_h} = \left(-\frac{\partial^2 \ell_{.j}}{\partial \zeta_h^2}\right) z_{hjs} \quad ; \quad \frac{\partial^2 \ell_{.j}}{\partial \zeta_q \partial \zeta_h} = 0 \qquad q \neq h,$$

$$-\frac{\partial^2 \ell_{.j}}{\partial \gamma_r \partial \gamma_s} = A_{ksr} - B_{ksr} + \sum_{i=1}^{k-1} z_{ijr} z_{ijs}\,,$$

where $d_h = \exp(\eta_{hj})$, $b_h = 1 - \exp\{-\exp(\eta_{hj})\}$, $c_h = d_h/b_h - 1$,

$g_h = (1 - b_h)d_h/b_h$, z_{ijs} is the sth component of $\underset{\sim}{z}_{ij}$,

$$A_{ksr} = \delta g_k (d_k - b_k) b_k^{-1} z_{kjs} z_{kjr}, \quad \text{and}$$

$$B_{ksr} = (1 - \delta) d_k z_{kjs} z_{kjr}\,.$$

These expressions are for the jth individual, with explanatory variables given by $\underset{\sim}{z}_{ij}$ in the ith time interval, who is censored or fails in the kth interval. When the explanatory variables do not vary across intervals the expressions are simplified.

The computation of the variance of the test statistic is simplified because of the form of the submatrix that corresponds to interval effects as discussed above for the two-sample case.

6.4 Examples

6.4.1 Data on remission of leukemia patients

We consider first data of Freireich et al. used by Gehan (1965) and several subsequent authors. Table 6.1 shows the ordered times for two samples of individuals; censored values are denoted with asterisks. Because of the small sample sizes, only five partition intervals were used. Different grouping schemes were tried, the difference being in the treatment of censored observations; the results were similar. Here we present the results obtained assuming that censoring occurs instantaneously at the end of each period. The time intervals were chosen of equal width, namely 5 weeks. Table 6.2 shows the resulting 2×2 tables so obtained.

Preliminary computations suggested fitting a model with $\lambda = 1$. Some results from GLIM fits are shown in Table 6.3. Twice the difference between the max loglikelihood achieved at $\lambda = 0$ and $\lambda = 1$ is approximately 3.47 with 1 d.f.; this is significant at 10% level. Hence, a model with additive hazard function seems rather suitable to describe the data than the multiplicative model. Two different estimates of the log-survivor function are shown in figure 6.1, they correspond to the estimates under the additive (solid line) and the multiplicative (dashed line) models.

Values of λ larger than one were tried and a slight improvement in the deviance observed, however it was not significant. Hence, we do not present those results.

TABLE 6.1

Times of remission (weeks) of leukemia patients

Sample 1 (drug 6-mp)	6*, 6, 6, 6, 7, 9*, 10*, 10, 11*, 13, 16, 17*, 19*, 20*, 22, 23, 25*, 32*, 32*, 34*, 35*
Sample 2 (control)	1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23

Source: Gehan (1965) *censored

TABLE 6.2

Grouped times of remission of leukemia patients. R denotes individuals with remission, N denotes total individuals at risk.

	Time intervals									
	1		2		3		4		5	
	R	N	R	N	R	N	R	N	R	N
Sample 1	0	21	5	21	1	13	1	11	2	7
Sample 2	9	21	4	12	5	8	1	3	2	2

TABLE 6.3

Parameter estimates and deviances for several simple models fitted to data in Table 6.2.

λ	approximate maximized loglikelihood	parameters of the linear systematic part of the model		
0	-56.105 (8.666)	$\hat{\zeta}_1 = -2.537$	$\hat{\zeta}_2 = -2.114$	$\hat{\zeta}_3 = -1.949$
		$\hat{\zeta}_4 = -2.511$	$\hat{\zeta}_5 = -0.873$	$\hat{\gamma} = 1.747$
0.3	-55.893 (8.242)	$\hat{\zeta}_1 = -1.978$	$\hat{\zeta}_2 = -1.456$	$\hat{\zeta}_3 = -1.514$
		$\hat{\zeta}_4 = -1.789$	$\hat{\zeta}_5 = -0.745$	$\hat{\gamma} = 1.239$
0.5	-55.424 (7.305)	$\hat{\zeta}_1 = -1.693$	$\hat{\zeta}_2 = -1.177$	$\hat{\zeta}_3 = -1.306$
		$\hat{\zeta}_4 = -1.442$	$\hat{\zeta}_5 = -0.692$	$\hat{\gamma} = 1.012$
0.8	-54.578 (5.612)	$\hat{\zeta}_1 = -1.247$	$\hat{\zeta}_2 = -0.900$	$\hat{\zeta}_3 = -1.039$
		$\hat{\zeta}_4 = -1.076$	$\hat{\zeta}_5 = -0.617$	$\hat{\gamma} = 0.694$
1	-54.367 (5.190)	$\hat{\zeta}_1 = -1.000$	$\hat{\zeta}_2 = -0.781$	$\hat{\zeta}_3 = -0.897$
		$\hat{\zeta}_4 = -0.912$	$\hat{\zeta}_5 = -0.571$	$\hat{\gamma} = 0.522$

N.B. Deviances appear in parentheses. The number of d.f. of approximate χ^2 is 3.

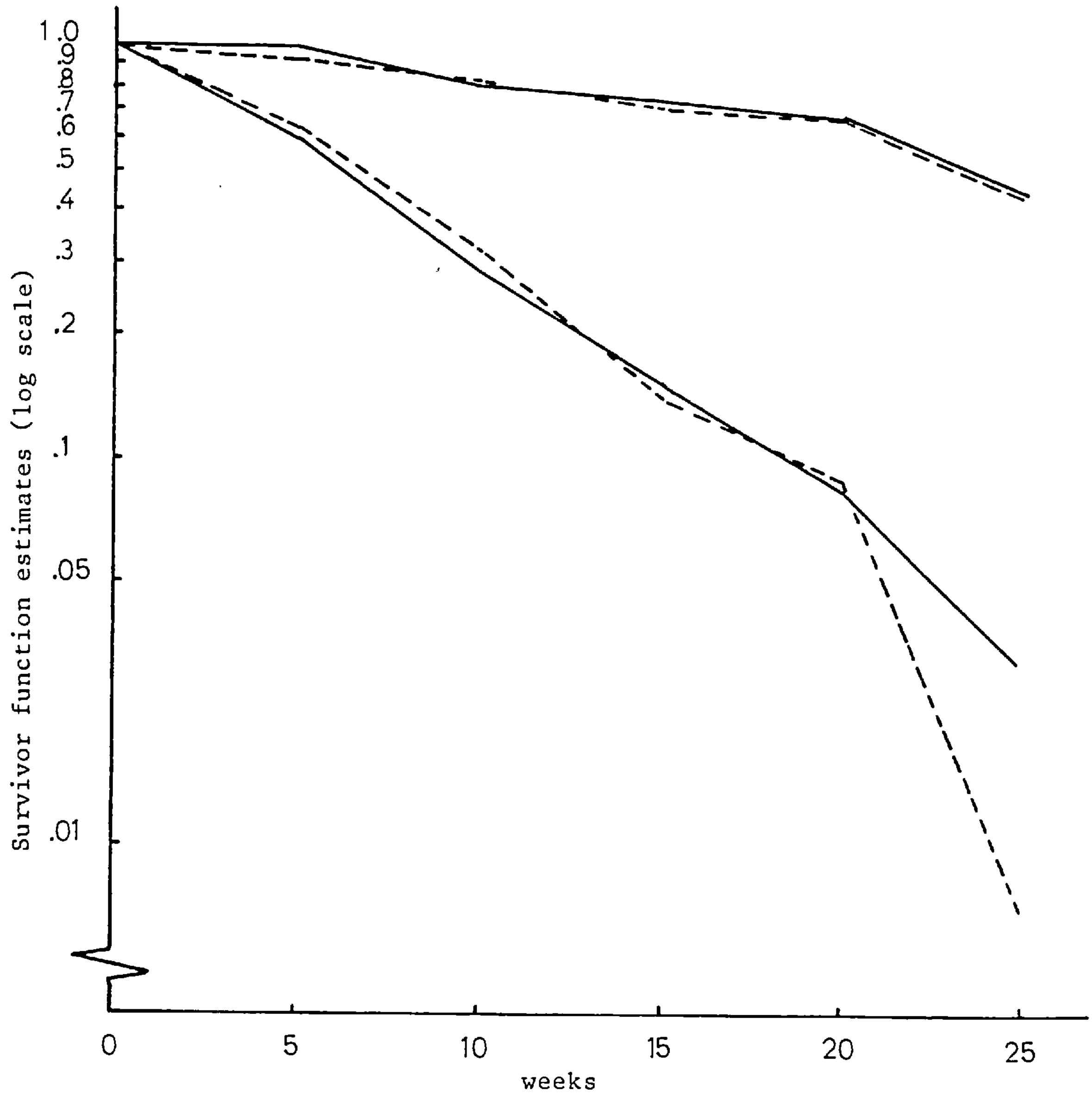

Figure 6.1

Comparison of continuous survivor function estimates under additive (——) and multiplicative (---) models for data in Table 6.2. Upper lines correspond to sample 1, lower lines to sample 2.

6.4.2 Simulated data

Several sets of simulated data were generated and analyzed. Simple patterns of failure and censoring were chosen. In general, the additive model showed a better performance.

We present here the results obtained for two of the mentioned sets. The first example corresponds to a simple situation where the probabilities of failure were (0.01, 0.13, 0.15, 0.25, 0.3) and (0.2, 0.25, 0.35, 0.4, 0.45) for samples 1 and 2 respectively. A uniform probability of censoring, 0.1, was applied for sample 1 while the observations in sample 2 were not censored. The samples were generated of equal size, 200. The data are shown in Table 6.4. Censoring was assumed to occur instantaneously at the end of the corresponding time interval.

Table 6.5 shows the results of fitting multiplicative, $\lambda = 0$, and additive, $\lambda = 1$, models to data in Table 6.4. The values of the L.R. and Pearson's chi-squared statistics show a striking difference of fit between the two models. The additive model provides a very good fit whereas the multiplicative provides a poor one. Besides, the estimates of the parameters associated to interval effects have small correlation among themselves. The correlations with the estimate of the "treatment" or group effect are also small, the biggest being -0.25. This is not so for the multiplicative model.

The results obtained fitting models with other values of the transformation parameter suggest that a maximum in the loglikelihood curve for λ is achieved at $\lambda = 1$, then $\hat{\lambda} = 1$.

The second set has probabilities of failure similar to the example above, namely (0.001, 0.125, 0.175, 0.2, 0.3) and (0.3, 0.35, 0.4, 0.45, 0.5) for samples 1 and 2, respectively. The censoring probabilities were a uniform value of 0.125 for sample 1, and (0.075, 0.1, 0.1, 0.125, 0.15) for sample 2. Hence, the difference between the failure patterns is stronger and censoring occurs in both samples. The initial size of each sample was 100. The data are shown in Table 6.6.

Table 6.7 shows the results of fitting additive and multiplicative models to data in Table 6.6. The difference in goodness of fit is striking here as well.

For this example the value of the test statistic introduced in §6.3 was computed obtaining a value of 2.9059. From tables of the standard normal distribution, this value is highly significant ($p < 0.002$). This agrees with our previous results.

A common feature of these two examples is that the multiplicative model provides a poor fit for very small probabilities of failure. The fitted values overestimate the actual ones in those cases as may be observed in Tables 6.5 and 6.7. We have chosen two extreme examples to illustrate the use and effectiveness of the suggested procedures.

TABLE 6.4

Simulated data. First set.

Sample 1			Sample 2		
Failures	Censorings	Survivals	Failures	Censorings	Survivals
1	12	187	37	0	163
20	21	146	38	0	125
19	20	107	41	0	84
25	8	74	36	0	48
22	5	47	19	0	29

TABLE 6.5

Results from fitting additive and multiplicative models to data in Table 6.4

		Additive model		Multiplicative model	
Sample	Observed	Fitted	Residual	Fitted	Residual
	1	0.996	0.399 E-2	10.875	-3.078
	20	18.267	0.427	14.494	0.369
1	19	20.781	-0.422	19.863	-0.208
	25	27.150	-0.478	22.041	0.707
	22	21.294	0.181	16.548	1.521
	37	37.148	-0.269 E-1	27.574	1.933
	38	42.641	0.827	39.385	-0.254
2	41	37.268	0.730	40.232	0.147
	36	32.702	0.738	38.478	-0.543
	19	20.018	-0.299	23.493	-1.297
L.R. chi-squared statistic:			2.479		23.799
Pearson's chi-squared statistic:			2.473		18.259

TABLE 6.6

Simulated data. Second set.

Sample 1			Sample 2		
Failures	Censorings	Survivals	Failures	Censorings	Survivals
0	10	90	29	6	65
8	10	72	26	8	31
18	10	44	11	8	12
5	6	33	5	2	5
12	4	17	4	0	1

TABLE 6.7

Results from fitting additive and multiplicative models to data in Table 6.6

		Additive model		Multiplicative model	
Sample	Observed	Fitted	Residual	Fitted	Residual
	0	0.711 E-12	-0.843 E-6	6.019	-2.531
	8	8.603	-0.216	9.451	-0.499
1	18	16.411	0.446	11.280	2.179
	5	5.137	-0.644 E-1	5.074	-0.351 E-1
	12	12.783	-0.280	11.590	0.151
	29	29.753	-0.165	23.560	1.283
	26	23.704	0.592	24.782	0.311
2	11	14.187	-1.149	16.171	-1.859
	5	4.555	0.265	4.939	0.360 E-1
	4	2.848	1.040	4.231	-0.286
L.R. chi-squared statistic:			3.298		22.090
Pearson's chi-squared statistic:			3.178		16.708

REFERENCES

AITCHISON, J. and BENNETT, J.A. (1970). Polychotomous quantal response by maximum indicant. _Biometrika_ 57 253-262.

ASHFORD, J.R. and SOWDEN, R.D. (1970). Multivariate probit analysis. _Biometrics_ 26 535-546.

ASHTON, W.D. (1972). _The Logit Transformation_. Griffin, London.

ATKINSON, A.C. (1973). Testing transformations to normality. _J.R. Statist. Soc._ B 35 473-479.

BAKER, R.J. and NELDER, J.A. (1978). The GLIM Manual-Release 3. Distributed by Numerical Algorithms Group, Oxford.

BHAPKAR, V.P. and KOCH, G.G. (1968). Hypotheses of 'no interaction' in multidimensional contingency tables. _Technometrics_ 10 107-123.

BOCK, R.D. (1970). Estimating multinomial response relations. Essays in Probability and Statistics, pp.453-479. Edited by R.C. Bose _et al_. Univ. of North Carolina Press, Chapel Hill.

BOX, G.E.P. and COX, D.R. (1964). An analysis of transformations (with discussion). _J.R.Statist.Soc._ B 26 211-252.

COLEMAN, J.S. (1964). _Introduction to Mathematical Sociology_. The Free Press of Glencoe, New York.

COX, D.R. (1966). Some procedures connected with the logistic qualitative response curve. Research Papers in Statistics: Essays in Honour of J. Neyman's 70th Birthday, pp. 55-71. Edited by F.N. David. Wiley, London.

COX, D.R. (1970). _The Analysis of Binary Data_. Methuen, London.

COX, D.R. (1972). Regression models and life tables (with discussion). _J.R.Statist.Soc._ B 34 187-220.

COX, D.R. (1975). Partial Likelihood. _Biometrika_ 62 269-276.

COX, D.. and SNELL, E.J. (1968). A general definition of residuals (with discussion). J.R.Statist.Soc. B 30 248-275.

COX, D.R. and HINKLEY, D.V. (1974). Theoretical Statistics. Chapman and Hall, London.

DARROCH, J.N. (1974). Multiplicative and additive interaction in contingency tables. Biometrika 32 587-598.

DYKE, G.V. and PATTERSON, H.D. (1952). Analysis of factorial arrangements when the data are proportions. Biometrics 8 1-12.

FINNEY, D.J. (1952). Probit Analysis. Cambridge University Press, Cambridge.

FISHER, R.A. (1935). The case of zero survivors. Appendix to Bliss, C.I. The calculation of the dosage-mortality curve. Ann. Appl.Biol.22 164-167.

GEHAN, E.A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. Biometrika 52 203-223.

GOODMAN, L.A. (1972). A modified multiple regression approach to the analysis of dichotomous variables. Amer.Sociol.Rev. 37 28-46.

GOODMAN, L.A. (1975). The relationship between modified and usual multiple-regression approaches to the analysis of dichotomous variables. Social Methodology. 1976; pp.83-110. Edited by D. Heise. Jossey-Bass, San Francisco.

HOLFORD, T.R. (1976). Life tables with concomitant information. Biometrics 32 587-598.

KALBFLEISCH, J.D. and PRENTICE, R.L. (1973). Marginal likelihoods based on Cox's regression and life model. Biometrika 60 267-278.

KAY, R. (1977). Proportional hazard regression models and the analysis of censored survival data. J.R.Statist.Soc. C 26 227-237.

KNOKE, D. (1975). A comparison of dummy variable regression and log-linear models for dichotomous variables. Sociol.Met. and Res. 3 416-434.

LAWLEY, D.N. (1956). A general method for approximating to the distribution of likelihood ratio criteria. Biometrika 43 295-303.

LOMBARD, H.L. and DOERING, C.R. (1947). Treatment of the fourfold table by partial association and partial correlation as it relates to public health problems. Biometrics 3 123-128.

MANTEL, N. (1966a). Models for complex contingency tables and polychotomous dosage response curves. Biometrics 22 83-95.

MANTEL, N. (1966b). Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemoter. Rep. 50 690-700.

MANTEL, N. and BROWN, C. (1973). A logistic reanalysis of Ashford and Sowden's data on respiratory symptoms in British coal miners. Biometrics 29 649-665.

MARTIN, J.T. (1940). The problem of the evaluation of rotenone-containing plants. V: The relative toxicities of different species of derris. Ann.Appl.Biol. 27 274-294.

McCULLAGH, P. (1980). Regression models for ordinal data (with discussion) J.R.Statist.Soc. B 42. To appear.

MORAN, P.A.P. (1970). On asymptotically optimal tests of composite hypotheses. Biometrika 57 47-55.

NELDER, J.A. and WEDDERBURN, R.W.M. (1972). Generalized linear models. J.R.Statist.Soc. A 135 370-384.

NERLOVE, M. and PRESS, S.J. (1973). Univariate and multivariate log-linear and logistic models. Rand Corporation Technical Report R-1306-EDA/NIH, Santa Monica.

RAO, C.R. (1973). Linear Statistical Inference and its Applications 2nd ed. Wiley, New York.

STOUFFER, S.A., SUCHMAN, E.A., DEVINNEY, L.C., STAR, S.A., and WILLIAMS, R.M.Jr. (1949). The American Soldier. Vol. 1, Princeton Univ. Press, Princeton.

YATES, F. (1955). The use of transformations and maximum-likelihood in the analysis of quantal experiments involving two treatments. Biometrika 42 382-403.