Examples Questioning the Use of Partial Likelihood

C. A. de B. Pereira; D. V. Lindley

The Statistician, Vol. 36, No. 1. (1987), pp. 15-20.

Stable URL:
http://links.jstor.org/sici?sici=0039-0526%281987%2936%3A1%3C15%3AEQTUOP%3E2.0.CO%3B2-X

The Statistician is currently published by Royal Statistical Society.

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

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

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

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

http://www.jstor.org
Thu Jan 24 20:24:30 2008

The Statistician (1987) 36, pp. 15–20

Examples questioning the use of partial likelihood

C. A. DE B. PEREIRA[1] & D. V. LINDLEY[2]

[1]*Universidade de São Paulo, IME, C.P. 20570, CEP 01498, São Paulo, Brasil and*
[2]*2 Periton Lane, Minehead TA24 8AQ, U.K.*

Abstract. Examples are given which cast doubt on the general validity of the partial likelihood argument and suggest that more stringent conditions need to be placed on the method before it can be used generally. Reasons are given for thinking that no such conditions can be found.
Key words and phrases: nuisance parameter, minimal sufficiency, Borel paradox, likelihood principle, contingency table, life table.

1 Introduction

Consider data x with density $p(x|\theta,\phi)$ dependent on two parameters, θ and ϕ. In the present note θ and ϕ will be real numbers but in general they could be many dimensional or even real functions. For any statistic $t=t(x)$, we may write

$$p(x|\theta,\phi)=p(t|\theta,\phi)p(x|t,\theta,\phi) \tag{1.1}$$

in terms of the marginal distribution of t and the conditional distribution of x, given t. (Here and elsewhere, $p(\cdot|\cdot)$ is a general notation for the probability density of what precedes the vertical line, given what follows it.) If the conditional distribution does not depend on ϕ,

$$p(x|\theta,\phi)=p(t|\theta,\phi)p(x|t,\theta) \tag{1.2}$$

and there is a factorisation of the likelihood for θ and ϕ, given x, into a product of two terms one of which does not involve ϕ. The suggestion has been made that, in order to make inferences about θ,ϕ being a nuisance parameter, it might be sensible to ignore the first factor, involving ϕ, and use only the second, so removing the need to consider the nuisance parameter ϕ at all. This is of considerable practical value, especially when ϕ is a complicated parameter as it is in the first example, Cox (1972), of the method, where ϕ is a hazard function $\lambda(t)$ of time t. The factor $p(x|t,\theta)$ is called a *partial likelihood* and may be used for likelihood methods of inference or, using the sample space of x, given the observed value of t, for inferential methods based on sampling distributions. (There is another factorisation of (1.1) in which the marginal distribution of t does not depend on ϕ. Similar ideas apply then to $p(t|\theta)$ which is called a *marginal likelihood.* An example of this occurs in Section 5.)

Although the method of partial likelihood has had several apparent successes, notably in the proportional hazards model of Cox (1972), well described by Kalbfleisch & Prentice (1980), it is difficult to justify discarding the other factor $p(t|\theta,\phi)$ in (1.2) especially in the important practical case where ϕ is complicated and its removal all the more desirable. The present note collects together some counter-examples where the method of partial likelihood either does not work or possesses properties that may be thought to be undesirable. This is not done in any destructive sense but in order to clarify some pertinent matters. What are needed are easily verified conditions under which the method can be used and the nuisance parameter safely forgotten. The examples suggest that these will be rather restrictive.

If (1.2) holds then, for any known value of θ, t is sufficient for ϕ. It is then usual to demand, Cox (1975) and Godambe (1980), that t be complete before the method can be used. Then, for any known value of θ, t is a minimal sufficient statistic for ϕ. This requirement guarantees the smallest possible reduction of the full model, $p(x|\theta,\phi)$, to a model that depends only on θ, $p(x|t,\theta)$. It will be convenient to write $x=(a,t)$ where a is another statistic which, together with t, is equivalent to x. Equation (1.2) may then be written

$$p(a,t|\theta,\phi)=p(t|\theta,\phi)p(a|t,\theta) \tag{1.3}$$

and the partial likelihood is $p(a|t,\theta)$.

2 2×2 Contingency table: binomial distributions

the most familiar application of the partial likelihood argument is to the 2×2 contingency table

$$\begin{array}{ccc} a & b & m \\ c & d & n-m \\ \hline t & n-t & n \end{array}$$

where m and n are fixed. A convenient description is in terms of two coins which are independently tossed m and $n-m$ times, resulting in a and c heads, respectively; a and c having binomial distributions with parameters p and q respectively, the chances of heads for the coins. The argument is that, in testing the hypothesis that $p=q$, the other margin, t, may also be held fixed and inferences may be based on the conditional distribution of the table entries (effectively a) given the margins.

Now

$$p(a,c|p,q)=\binom{m}{a}\binom{n-m}{c}p^a(1-p)^b q^c(1-q)^d. \tag{2.1}$$

The new parameterisation

$$\theta=\frac{p(1-q)}{q(1-p)}$$

and $\phi=q$ makes the hypothesis to be tested $\theta=1$. (Note that the argument would apply to any inferences about the odds-ratio such as its estimation.) Then (2.1) is

$$p(a,t|\theta,\phi)=\binom{m}{a}\binom{n-m}{t-a}\frac{\theta^a\phi^t(1-\phi)^{n-t}}{(1-\phi+\theta\phi)^m} \tag{2.2}$$

where c has been replaced by $t-a$. On summing over a

$$p(t|\theta,\phi)=\frac{\phi^t(1-\phi)^{n-t}}{(1-\phi+\theta\phi)^m}\sum_a\binom{m}{a}\binom{n-m}{t-a}\theta^a$$

so that

$$p(a|t,\theta,\phi)=\frac{\binom{m}{a}\binom{n-m}{t-a}\theta^a}{\sum_k\binom{m}{k}\binom{n-m}{t-k}\theta^k} \tag{2.3}$$

which does not depend on the nuisance parameter ϕ and hence may be used for inferences about θ, in particular to test $\theta=1$ or $p=q$. This is Fisher's exact test. Note that t is sufficient and complete (for ϕ, given θ).

3 2×2 Contingency table: negative-binomial distributions

Let the contingency table situation be modified so that b and d are fixed, rather than m and $n-m$. The distributions for the two coins will then be *negative* binomial in contrast to the (positive) binomial usually considered. Now

$$p(a,c|p,q)=\binom{m-1}{a-1}\binom{n-m-1}{c-1}p^a(1-p)^b q^c(1-q)^d \tag{3.1}$$

and in terms of θ and ϕ

$$p(a,c|\theta,\phi)=\binom{a+b-1}{a-1}\binom{t-a+d-1}{t-a-1}\frac{\theta^a\phi^t(1-\phi)^{b+d}}{(1-\phi+\theta\phi)^{a+b}} \tag{3.2}$$

where (a,t) are the data and (b,d) are fixed. An attempt to carry through the previous argument that led to (2.3) fails because the minimal sufficient statistic for ϕ given θ is the whole sample (a,t).

However, all is not lost, for consider the alternative parametrisation

$$\xi=\frac{p}{q}$$

and $\phi=q$ so that the null hypothesis $p=q$ becomes $\xi=1$ and ϕ is the nuisance parameter. (3.1) may now be written

$$p(a,t|\xi,\phi)=\binom{a+b-1}{a-1}\binom{t-a+d-1}{t-a-1}\xi^a(1-\xi\phi)^b\phi^t(1-\phi)^d$$

and summation over a gives

$$p(t|\xi,\phi)=(1-\xi\phi)^b\phi^t(1-\phi)^d\sum_a\binom{a+b-1}{a-1}\binom{t-a+d-1}{t-a-1}\xi^a$$

so that

$$p(a|t,\xi,\phi)=\frac{\binom{a+b-1}{a-1}\binom{t-a+d-1}{t-a-1}\xi^a}{\sum_k\binom{k+b-1}{k-1}\binom{t-k+d-1}{t-k-1}\xi^k} \tag{3.3}$$

which does not depend on the nuisance parameter, ϕ. Now t is minimal sufficient for θ given ξ and the partial likelihood argument can now proceed.

We notice that if this new parametrisation, (ξ,ϕ), were to be used in the usual (positive) binomial case, the minimal sufficient statistic for ϕ given ξ would be the whole data (a,t) and the argument would fail just as in the negative binomial case with (θ,ϕ). The roles of the statistics t and (a,t) with the parametrisations (θ,ϕ) and (ξ,ϕ) are reversed by changing the model from (positive) binomial to negative binomial or vice-versa.

It is known (Lindley & Phillips, 1976) that with a *single* coin, significance tests with the positive (m fixed) and negative (b fixed) binomials give different results; so it is not surprising that with two coins (2.3) and (3.3) should differ. What is interesting is that in order to obtain the tests or estimates different parametrisations have to be used:

$$\theta = \frac{p(1-q)}{q(1-p)}$$

for the binomial and

$$\xi = \frac{p}{q}$$

for the negative binomial. Note that, although different, there is a one-to-one correspondence between any two of the parametrisations (p,q), (θ,ϕ), and (ξ,ϕ). However, θ and ξ are quite different; for example, one is not a function of the other. Even the hypotheses $\theta=1$ and $\xi=1$ are in a sense different, as Borel's paradox shows (Lindley, 1980). Why should a scientist be forced to shift his interest from θ to ξ just because the same data were obtained in different ways? This raises doubts concerning the value of the partial likelihood approach. The next examples concerns another feature of the argument.

4 Information in the marginal distribution

It is clear that the method of partial likelihood will not be satisfactory if the omitted factor of the original likelihood (1.2), $p(t|\theta,\phi)$ contains substantial information about θ, despite the presence of ϕ. The next example concerns a situation where this is true.

Suppose the data are (a,t) and that the marginal distribution of t is normal with mean θ and variance $1+\delta[2F(\phi)-1]$. Here θ and ϕ are real, unknown parameters; δ is a known, positive real number (to be thought of as small); and $F(\cdot)$ is a known distribution function. The idea is that as ϕ, the nuisance parameter, varies over the whole real line, the variance of t changes a little from $1-\delta$ to $1+\delta$. To complete the description, suppose that the conditional distribution of a, given t, is normal with mean $t+\theta$ and unit variance. This does not involve ϕ and the factorisation (1.2) required for the partial likelihood argument is available. Using it, we have that $(a-t)$ is the best estimate of θ with unit variance.

However, if all the data are considered we may write

$$\begin{aligned} t&=\theta+\varepsilon_1,\\ a&=t+\theta+\varepsilon_2, \end{aligned} \tag{4.1}$$

where ε_1 and ε_2 are independent normals with zero means and variances about one and exactly one respectively. Hence $a=2\theta+\varepsilon_1+\varepsilon_2$ and $\frac{1}{2}a$ is an estimate of θ with variance about $\frac{1}{2}$ and, if $\delta<2$, certainly less than 1, the variance of the estimate based on partial likelihood. Consequently the partial likelihood method is only about 50% efficient.

Another feature of the example is that although the partial likelihood approach leads to the use of $(a-t)$ whenever $\delta\neq 0$, it gives $\frac{1}{2}a$ when $\delta=0$, when t is no longer sufficient. Consequently the argument has an abrupt discontinuity at $\delta=0$, which is unsatisfactory.

Note that for a known value of θ,t, although sufficient, is not complete for ϕ, a point we had overlooked above. With θ known $u=(t-\theta)^2$ is both sufficient and complete for ϕ and the partial likelihood argument should be based on u, not t. Write $t=(u,z)$ where $z=+1(-1)$ if $t>(<)\theta$. Clearly $p(z=+1)=1/2$, irrespective of the values of u, θ and ϕ, and the partial likelihood is $p(a,z|u,\theta)$. This is equal to

$$p(a|z,u,\theta)p(z|u,\theta)= \frac{1}{2}p(a|z,u,\theta)= \frac{1}{2}p(a|t,\theta)$$

since z has a known distribution and $t=(u,z)$. We are therefore led to the same partial likelihood, $p(a|t,\theta)$ originally used.

Finally we notice that the use of $u=(t-\theta)^2$ is awkward because although it is a statistic, given θ, it is not a statistic in the original problem.

5 Life-table example

This example is due to Basu (1977) and Lindley (1979) where it arises in the analysis of life-tables. Consider a trinomial distribution with chances in three cells

$$\frac{1}{2}(1+\theta)\,(1-\phi),\ \frac{1}{2}(1-\theta)\,(1+\phi),\ \theta\phi$$

depending on two parameters, θ and ϕ, both in the unit interval. For data (a,b,c) the likelihood is

$$(1+\theta)^a(1-\theta)^b\theta^c.\ (1-\phi)^a(1+\phi)^b\phi^c \tag{5.1}$$

which factors into a function of θ times one of ϕ. Since, in the life-table context, θ and ϕ refer to different features (θ to death and ϕ to withdrawal) it seems reasonable to base inferences for θ entirely on that factor of (5.1) that contains it. Yet there are no statistics a and t that enable it to be written as a product of probability distribution as (1.2) does. Consequently the sample space for a significance test or an unbiased estimate is unclear.

One may claim that this example exactly fits the definition of likelihood factor given by Lindsay (1980). But Lindsay's paper contains an example that shows his definition will not work. It has been discussed in detail by Lindley (1985). The essence of the argument is that again there is a trinomial with probabilities

$$\phi\left(\frac{1}{2}-\theta\right)^2+(1-\phi)\left(\frac{1}{2}+\theta\right)^2,\ 2\left(\frac{1}{2}-\theta\right)\left(\frac{1}{2}+\theta\right),\ \phi\left(\frac{1}{2}+\theta\right)^2+(1-\phi)\left(\frac{1}{2}-\theta\right)^2$$

with $\leq\theta\leq 1$ and $0\leq\theta\leq\frac{1}{2}$. For data $(a,t,n-a-t)$ t has a distribution that depends only on θ and is a marginal likelihood (see Section 1). This distribution is binomial with parameter

$$2\left(\frac{1}{2}-\theta\right)\left(\frac{1}{2}+\theta\right)=\frac{1}{2}-2\theta^2$$

and t/n is accordingly the best estimate of it. Lindsay shows that in the original trinomial situation it has optimality properties. Yet it can easily happen that t/n is greater than $\frac{1}{2}$, whereas the parameter is necessarily less than or equal to $\frac{1}{2}$. Consequently the 'best' estimate can be an impossible value for the parameter. (On the average it is all right.)

6 Discussion

A very common and important problem in statistical practice is how to make inferences about a parameter of interest, θ, in the presence of nuisance parameters, ϕ. Outside the Bayesian method, there is a no general way to do this, and even within it the necessary integration with respect to ϕ may be difficult to perform. The partial (or marginal) likelihood approach seems to offer an excellent solution because, when the factorisation is possible, the nuisance parameter disappears completely. This is especially appealing to the applied statistician because he can forget ϕ, thinking that the theoretician has justified the procedure. It is for these reasons that we originally became interested in the method. Now it is immediately clear that it will not work in complete generality: Cox & Hinkley (1974, Example 5.12) provide a strong example. So it is necessary to delineate those conditions when it will work. A case usually cited is the 2×2 contingency table, but our discussion in Sections 2 and 3 shows that even this has unsatisfactory features. The example in Section 4 shows how much information can exist in the discarded part of the likelihood. That in Section 5 shows that it will not work even when the factorisation is especially straightforward. Yet the method continues to be used, most often with the proportional hazards model. Here there is an asymptotic justification, Efron (1977), but the small-sample properties seem to be unknown. And the failure of asymptotic results in small samples is well illustrated by Lindsays example in Section 5. Thus the evidence suggests that this very promising procedure will not work.

Acknowledgement

This work was partially supported by FAPESP, a state of São Paulo Foundation for Science.

References

BASU, D. (1977) On the elimination of nuisance parameters, *Journal of the American Statistical Association*, 72, pp. 355–366.

BASU, D. (1979) Discussion of Joseph Berkson's paper 'In Dispraise of the Exact Test'. *Journal of Statistical Planning and Inference*, pp. 189–192.

COX, D.R. (1972) Regression models and life tables (with discussion), *Journal of the Royal Statistical Society Series B*, 30, pp. 248–275.

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

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

EFRON, B. (1977) Efficiency of Cox's likelihood function for censored data, *Journal of the American Statistical Association*, 72, pp. 557–565.

GODAMBE, V.P. (1980) On sufficiency and ancillarity in presence of a nuisance parameters, *Biometrika*, 67, pp. 155–162.

KALBFLEISCH, J.D. & PRENTICE, R.L. (1980) *The Statistical Analysis of Failure Time Data* (New York, John Wiley).

LINDLEY, D.V. (1979) Analysis of life tables with grouping and withdrawals, *Biometrics*, 35, pp. 605–612.

LINDLEY, D.V. (1980) The Bayesian approach to statistics, in: J. TIAGO DE OLIVEIRA (Ed.) *Some Recent Advances in Statistics*, pp. 65–87 (London, Academic Press).

LINDLEY, D.V. & PHILLIPS, L.D. (1976) Inference for a Bernoulli process (a Bayesian view), *The American Statistician*, 30, pp. 112–119.

LINDLEY, D.V. (1985) Bernoulli pairs with invariant reversals (an example of partial likelihood), *Proceedings of Symposium in honour of V. M. Joshi* (in press).

LINDLEY, B.G. (1980) Nuisance parameters, mixture models, and the efficiency of partial likelihood estimators, *Philosophical Transactions of the Royal Society of London Series A*, 296, pp. 639–662.

PEREIRA, C.A. DE B. (1983) Stopping rules and conditional inference in 2×2 contingency tables, *São Paulo, Technical Report RT-MAE-8301*, 7 pp.