Likelihood Inference for a Simple Model of Cancer Spread

by
David Hinkley
University of Minnesota
Technical Report No. 340
April, 1979

Research supported in part by NSF Grant MCS-77-00959

Likelihood Inference for A Simple Model of Cancer Spread

David Hinkley

School of Statistics, University of Minnesota

Summary

Large-sample methods of likelihood inference are examined for a simple model describing the spread of cancer cells. Attention focusses on the use of conditional methods which make appropriate use of the amount of information actually delivered by an experiment. Numerical results are given to illustrate the performances of several methods of obtaining confidence limits.

Key words: Likelihood; Ancillary statistic; Information; Pivotal; Confidence interval; Maximum likelihood estimate; Binary trials; Stopping rule; Cancer.

1. Introduction

When likelihood methods are used to make inferences about parameters of a probability model, it is important to realise that the actual amount of information obtained in an experiment may differ appreciably from what is expected. This seems to have been duly noted by Fisher (1925), Bliss and James (1966) and a few others. Recently there has been theoretical research on the role of the observed information by Efron and Hinkley (1978), who consider in particular various "curved" exponential family models. One instance of such a model is given by Downham and Green (1976) in connexion with an experiment to measure the relative division rate of cancer cells. The purpose of the present paper is to describe and illustrate various likelihood methods of setting confidence limits for the unknown division rate in Downham and Green's model. One conclusion reached is that the likelihood ratio method is preferable, in the sense that it makes good use of the amount of information obtained and that reliable confidence coefficients are often obtained via the usual chi-square approximation.

Section 2 outlines Downham and Green's sampling model and the associated likelihood. The sample information and approximate methods of using the information in obtaining confidence limits are described in Section 3. Sections 4 and 5 deal with numerical results concerning the variability of information and performance of confidence limit methods for moderate sample sizes.

2. The Cancer Model and Its Likelihood

Downham and Green (1976) describe a simple Markov model for the spread of cancer cells in a layer of competing normal cells, there being initially one cancer cell. According to the model, if N_t is the number of cancer cells after t cell divisions, irrespective of type, then

$$N_0 = 1, \quad N_t = N_{t-1} + \varepsilon_t \quad (t = 1,2,\ldots), \tag{2.1}$$

where $\varepsilon_1, \varepsilon_2, \ldots$ are independent and

$$pr(\varepsilon_t = +1) = \pi, \quad pr(\varepsilon_t = -1) = 1-\pi.$$

The ratio $\pi/(1-\pi) = \theta$ is the relative division rate of cancer cells, and is the parameter of interest. If at some point $N_x = 0$, then necessarily $N_t = 0$ for $t > x$ since no cancer cells can be produced thereafter.

The unit experiment considered by Downham and Green consists of observing $\{N_t\}$ until either $N_t = 0$ or $N_t = m$, for some pre-assigned integer $m \geq 2$. This unit experiment is repeated n times. The observations resulting from the replicated experiment are then (X_j, Y_j), $j=1,\ldots n$, where

$$X = \min\{x: N_x = 0 \text{ or } N_x = m\}$$

$$Y = \begin{cases} 1 \text{ if } N_x = m \\ 0 \text{ if } N_x = 0 \end{cases} . \tag{2.2}$$

Note that if $m = 2$ then $X = 1$.

Elementary considerations show that the joint probability function of (X,Y) is

$$f(x,y) = c(m,x,y)\pi^{\frac{1}{2}(x-1+my)}(1-\pi)^{\frac{1}{2}(x+1-my)}, \tag{2.3}$$

since there must be a net gain of my-1 when we sum the x +1's and -1's. The model (2.3) is a curved exponential family model (Efron & Hinkley, 1978,

Section 7) for m > 2. The minimal sufficient statistic from the n independent unit experiments is then found to be

$$S = (S_1, S_2) = (\Sigma X_j, \Sigma Y_j). \qquad (2.4)$$

Notice that the likelihood for π, or θ, is the same as that for a simple binomial sample with

$$\text{number of trials} = S_1 \text{ , number of +1's} = \tfrac{1}{2}(S_1 + mS_2 - n).$$

Since S_1 varies from sample to sample, the distribution of likelihoods is not the same as for simple binomial sampling, and the amount of information varies from sample to sample, as we shall see.

Now consider inference about the parameter of interest, $\theta = \pi/(1-\pi)$. The various methods we shall consider involve the maximum likelihood estimate $\hat{\theta}$, which is the unique stationary point of the likelihood, given by

$$\hat{\theta} = \frac{S_1 + mS_2 - n}{S_1 - mS_2 + n} \cdot = \frac{T}{U}$$

$S_1 + mS_2 - n = 2\#(\text{Successes}) = 2T$

$S_1 - mS_2 + n = 2\#(\text{Failures}) = 2U$

Reduction of the sufficient data (S_1, S_2) to $\hat{\theta}$ alone must involve a loss of information, unless m = 2, but we shall see that some methods recover this lost information in a useful way. An essential point is that different samples with the same $\hat{\theta}$ contain different amounts of information, and it seems sensible to condition inference on the amount of information actually observed when computing the standard error of $\hat{\theta}$.

3. Information and Informative Inference.

In order to condition our inference about θ on the actual information obtained, we first need to define information quantitatively. Next we need

to find an ancillary statistic A, with distribution approximately independent of θ, which measures information. Conditional inferences will then be drawn from the distribution of $\hat{\theta}$ given the sample value of A.

The log-likelihood obtained from the replicated experiment described in Section 2 is, from (2.3) and (2.4)

$$\ell_\theta = \ell_\theta(S) = \text{constant} + \tfrac{1}{2}S_1\{\log\theta - 2\log(1+\theta)\} + \tfrac{1}{2}mS_2\log\theta - \tfrac{1}{2}n\log\theta \quad . \tag{3.1}$$

Successive derivatives with respect to θ will be denoted by $\dot{\ell}_\theta, \ddot{\ell}_\theta$, etc. For the unit experiment, corresponding derivatives will be denoted by $\dot{\ell}_\theta(X,Y)$, etc. Then the observed Fisher information for (3.1) is defined as

$$I = -\ddot{\ell}_{\hat{\theta}}(S) = \frac{S_1 - mS_2 + n}{2\hat{\theta}(1+\hat{\theta})} = \frac{v^3}{\tau(\tau+v)} = \frac{v^3}{\tau S_1} \tag{3.2}$$

in contrast to the expected, or average, Fisher information

$$ni_\theta = E\{-\ddot{\ell}_\theta(S)\} = nE\{-\ddot{\ell}_\theta(X,Y)\}. \tag{3.3}$$

For our particular model one can see quite easily that the pair $(\hat{\theta},I)$ is equivalent to S, and so contains all the experimental information about θ. Further, if we standardize the observed information I we obtain an approximately ancillary measure of the informativeness of the experiment, namely

$$A = \sqrt{n}\left(1 - \frac{I}{ni_{\hat{\theta}}}\right)/\gamma_{\hat{\theta}} \quad , \tag{3.4}$$

where

$$i_\theta^2\gamma_\theta^2 = \mathrm{Var}\{\ddot{\ell}_\theta(X,Y)\} - \frac{[\mathrm{Cov}\{\dot{\ell}_\theta(X,Y), \ddot{\ell}_\theta(X,Y)\}]^2}{\mathrm{Var}\{\dot{\ell}_\theta(X,Y)\}} \quad . \tag{3.5}$$

Efron and Hinkley (1978) show that for large n, A is approximately a standard normal variable. As the standardization in (3.4) suggests, the coefficient of variation of I is approximately $\gamma_{\hat{\theta}}/\sqrt{n}$, which as we shall see can be quite large.

The sufficient data (S_1,S_2) has now been successively transformed first to $(\hat{\theta},I)$ and then to $(\hat{\theta},A)$, no information having been lost. If we dropped I or A, we would lose approximately $\gamma_{\hat{\theta}}^2 i_{\hat{\theta}}$ units of Fisher information. To condition inference about θ on the value of the ancillary "information indicator" A requires, in principle, that we find the exact distribution of $\hat{\theta}$ given A. However, for moderately large n there are simple approximations available, as described by Efron and Hinkley (1978).

First, the pivot

$$P_c(\hat{\theta},\theta) = \sqrt{I}(\hat{\theta}-\theta) \qquad (3.6)$$

is approximately standard normal given A. Second, the pivot

$$LR(\hat{\theta},\theta) = 2(\ell_{\hat{\theta}} - \ell_{\theta}) \qquad (3.7)$$

is approximately χ_1^2 given A. Both pivots may be used to set confidence limits in the usual way; for example, approximate 95% limits are

$$\text{values of } \theta \text{ such that } P_c(\hat{\theta},\theta) = \pm 1.96 \quad .$$

Notice that methods based on P_c and LR do not require calculation of A, and that both use only characteristics of the observed likelihood.

The conditional pivot P_c is to be contrasted with the unconditional pivot

$$P_u(\hat{\theta},\theta) = \sqrt{(ni_{\hat{\theta}})}\ (\hat{\theta}-\theta) \quad , \qquad (3.8)$$

which is approximately standard normal unconditionally for large n. This pivot does not take account of the observed information I.

The likelihood ratio pivot LR has a potential advantage in that the inference will not depend on the particular parametrization employed, whereas P_c would produce different results if we worked with π rather that θ, for example. The particular choice of working parameter can, in principle,

be chosen to make the standard normal approximation for P_c as accurate as possible. If the working parameter is denoted by ψ, then two suggestions for choice of ψ are (i) to make $I^{(\psi)} = -\ddot{\ell}_{\hat{\psi}}$ a constant (Efron and Hinkley, 1978), (ii) to make $E(-\dddot{\ell}_{\psi})$ or $-\dddot{\ell}_{\hat{\psi}}$ zero (Sprott, 1975). These suggestions correspond to variance stabilization and symmetrization techniques. Note that (ii) would tend to produce best agreement between LR and P_c, because $LR - P_c^2$ is approximately proportional to $-\dddot{\ell}$.

4. Variation of I in the Example

To return to our example, we wish to show how well the various pivots work in the context of obtaining confidence limits for θ. In order to do this we need to be able to compute the ancillary A defined by (3.4), which involves both i_θ and γ_θ. These two quantities are determined from standard properties of the random walk $\{N_t\}$, as described in the Appendix. Table 1 shows the values of i_θ and γ_θ when $\theta = 1$ for various values of m, from which it is evident that the standard error for I can be large. For example, with $\theta = 1$, $m = 4$ and $n = 20$, an approximate 95% range for I is

$$20i_1(1 \pm 2\gamma_1/\sqrt{20}) = 20 \pm 9 \quad .$$

This implies possibly large differences between P_c and P_u. Figure 1 illustrates γ_θ as θ varies for m = 3, 5 and 10 showing that γ tends to be largest near $\theta = 1$, whence P_c and P_u will disagree most often when $\theta \doteq 1$. By way of comparison, the value of γ for a Cauchy error model is about 1.6.

Table 1. Unit information i and curvature γ when $\theta = 1$

m	2	3	4	5	7	10	20	50
i	0.25	0.50	0.75	1.00	1.50	2.25	3.75	12.25
γ	0	0.67	0.86	1.00	1.22	1.47	2.66	3.33

Figure 1 here

5. Numerical Performance of Confidence Limit Methods Based on Pivots

To illustrate the properties of confidence limits based on P_c, LR and P_u, we have generated 10,000 samples for several combinations of θ, m and n. In each case samples have been grouped by interval values of A, and for each interval we have computed the frequency with which confidence limits covered the true θ. This then gives a crude picture of the conditional coverage frequency as A varies.

Figure 2 shows the resulting graph for the case $(\theta, m, n) = (1.5, 5, 20)$ when approximate 95% confidence limits are sought. In this graph, as in others, we have plotted error rate, i.e. one minus coverage frequency, versus mean A value for each group of samples. The graph illustrates the general tendency for P_c and LR to yield accurate confidence limits for most values of A, whereas the method based on P_u has coverage deviating considerably from the average 95% when I deviates from $ni_{\hat{\theta}}$. Figure 3 shows similar effects for the case $(\theta, m, n) = (2.33, 5, 20)$.

To an approximate degree the performance of P_u is predictable, if the standard normal approximations for P_c, P_u and A are accurate, since $|P_u| \geq 1.96$ is nearly the same as

$$|P_c| \geq 1.96\sqrt{(1 \pm A\gamma_\theta/\sqrt{n})} \quad ,$$

as may be seen from (3.4), (3.6) and (3.8). Using the normal approximations we have computed what the conditional error rates would be for the 95% confidence limit method based on P_u at the "outer limits" $A = \pm 2$, for various values of $\gamma_\theta/\sqrt{n}$. These are given in Table 2.

Figures 2,3 here

Table 2. Approximate values of the conditional error rates of the 95% method based on P_u when $A = \pm 2$.

$\gamma_\theta/\sqrt{n}$ = c.v.(I)	0	0.1	0.2	0.3	0.4
error rate at A = -2	5	3	2	1	<1
error rate at A = +2	5	8	13	22	38

These figures should only be taken as rough guides, of course, but they do seem to give accurate indications of what is observed in Figures 3 and 4, where $\gamma_\theta/\sqrt{n}$ is about 0.2.

A word of caution is necessary at this point. The results shown so far have not been seriously affected by the discreteness of the values of $\hat{\theta}$ and I, but in some cases the normal and chi-square approximations for P_c and LR can be very inaccurate. This will typically happen when the vast majority of experiments end with $N_x = 0$, i.e. Y = 0, as is the case if $\theta \leq 1$ and m is large; $pr(Y=0) \geq m^{-1}$ when $\theta \leq 1$. Figures 4 and 5 illustrate this effect for nominal 95% confidence limits by contrasting the cases $(\theta, m, n) =$ (1, 10, 20) and (3, 10, 20); only 1,000 samples were used here. The performances of P_c and P_u are quite similar in Figure 5 because $\gamma_\theta/\sqrt{n}$ is considerably lower than for the cases in Figures 2 and 3.

The distinction between conditional and unconditional methods is highlighted in Figure 4, where error rates vary from near 0% to about 20% depending on A. The unconditional error rates are about 8%. Clearly in a case such as this there would need to be a rather detailed evaluation of the conditional distribution of $\hat{\theta}$ in order that erroneous conclusions be avoided.

Figures 4,5 here

As we mentioned at the end of Section 3, the standard normal approximation for P_c can sometimes be improved by suitable choice of working parameter. Both suggestions that we made there for transforming θ are precluded here by mathematical complications, but inspection of ℓ_θ suggests that for $\psi = \theta^{1/3}$, the "skewness" $E(-\overset{\cdots}{\ell}_\psi)$ will be nearly zero. The pivot corresponding to P_c is

$$P_c^*(\hat{\psi},\psi) = 3\hat{\theta}^{2/3} P_c(\hat{\theta}^{1/3},\theta^{1/3}) \quad .$$

Figure 6 compares the normal plots for P_c and P_c^* from 49 samples of the case $(\theta, m, n) = (1, 3, 20)$. This plot suggests superiority of P_c^*. Note the evident bumpiness of the distributions. For the difficult cases illustrated in Figure 4, results for P_c^* are very close to those for LR. This should be true in general, since

$$LR(\hat{\theta},\theta) - \{P_c^*(\hat{\psi},\psi)\}^2 \doteq 1/3(\hat{\psi}-\psi)^3 \overset{\cdots}{\ell}_{\hat{\psi}}$$

[Figure 6 here]

In all of the cases illustrated above the observed distribution of A was very close to standard normal.

6. Concluding Remarks

For setting confidence limits, the use of the conditional pivotals $P_c(\hat{\theta},\theta)$ and $LR(\hat{\theta},\theta)$ appears to give accurate results irrespective of the actual information in the performed experiment. The same cannot be said of the unconditional pivot $P_u(\hat{\theta},\theta)$. It is noteworthy that the conditional methods agree with approximate Bayesian methods, and indeed share the property that different sampling experiments with the same likelihood yield

the same inferences. If confidence limits are to be based on $\hat{\theta}$ and an estimated standard error, then it appears to be beneficial to choose the scale of θ so that the third derivative of log likelihood is small; this choice is unnecessary when LR is used.

It is interesting to note that the approximate normality of the pivotal $P_c(\hat{\theta},\theta)$ may hold even when $P_u(\hat{\theta},\theta)$ is never approximately normal. This happens for long-term observation of the linear birth (Yule) process, as shown by Feigin and Reiser (1979).

The approximate calculations of variability for observed information I (Section 4) may be of value in designing the experiment. For example, if in the cancer experiment all experiments with nm fixed were equally convenient (which is unlikely), then the choice $m = 2$ could essentially guarantee the amount of information, since I has a very small standard error ($\gamma = 0$).

When discrete distributions are involved, as here, it is obviously important to learn as much as possible about the effect of the discreteness on the large-sample approximations, no matter what method is used to obtain confidence limits.

Appendix

We outline here the derivations of various properties of the likelihood ℓ_θ (3.1). The main point to note is that successive derivatives $\dot{\ell}_\theta(X,Y),\ldots$ are linear combinations of X and Y, which are defined in (2.2). Thus in computing i_θ and γ_θ we need only to evaluate the joint first and second moments of (X,Y), which are well-known in the study of the simple random walk with absorbing barriers. The required moments can be obtained easily from the results

$$E[e^{-\lambda N_X}\{f^*(\lambda)\}^X] = 1$$

and

$$E\{z^X | Y = 1\} = \frac{F_{m-1}(z)}{F_{m-1}(1)}$$

where $f^*(\lambda) = \pi e^{\lambda} + (1-\pi)e^{-\lambda}$, $F_a(z) = (\lambda_1-\lambda_2)/(\lambda_1^{a+1}-\lambda_2^{a+1})$ and $\lambda_1(z)$, $\lambda_2(z)$ are the solutions of $zf^*(\lambda) = 1$; see Cox and Miller (1965, Sections 2.2(ii) and 2.3(v)). For the special case $\theta = 1$, i.e. $\pi = 1/2$, we find that

$$E(X) = m-1,\ E(Y) = \frac{1}{m},\ \mathrm{Var}(X) = \frac{m(m-1)(m-2)}{3},\ \mathrm{Cov}(X,Y) = \frac{(m-1)(m-2)}{3m},$$

$$\mathrm{Var}(Y) = \frac{m-1}{m^2}.$$

The general moments are complicated and will not be given here.

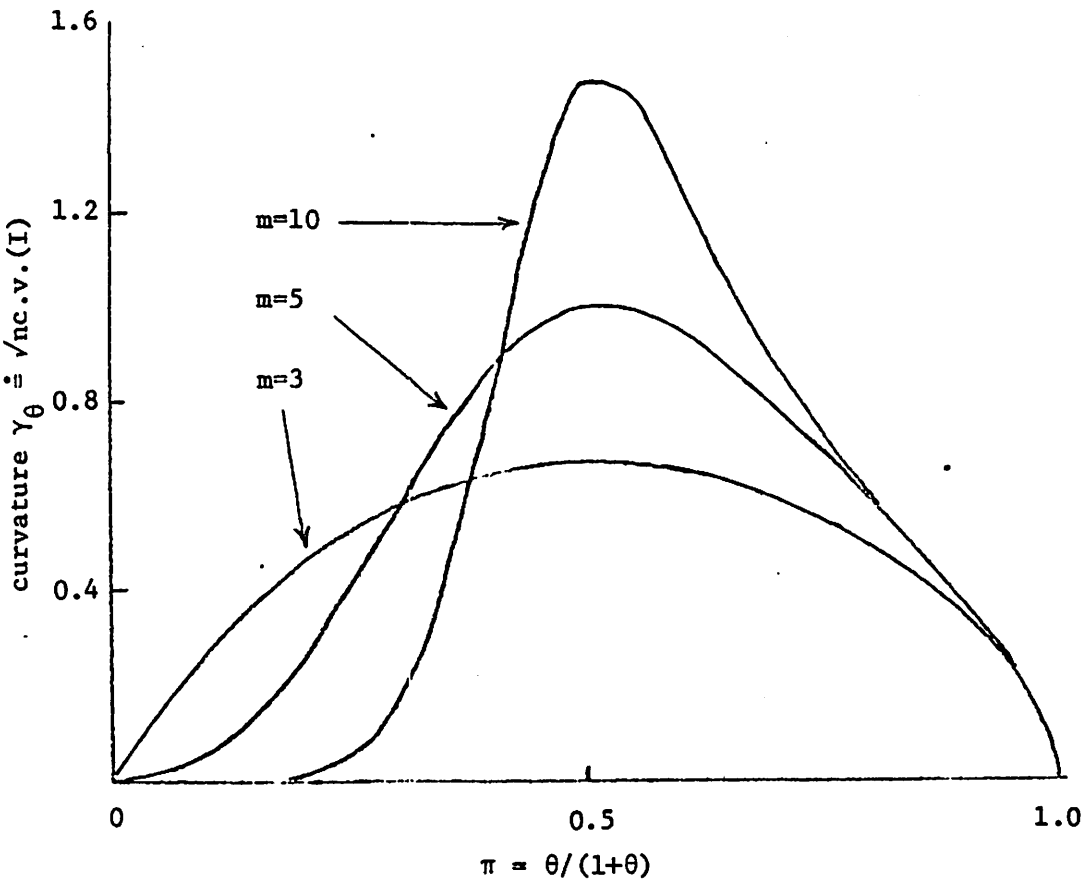


Fig. 1. Curvature $\gamma_\theta \doteq \sqrt{nc.v.(I)}$ for m=3,5,10.

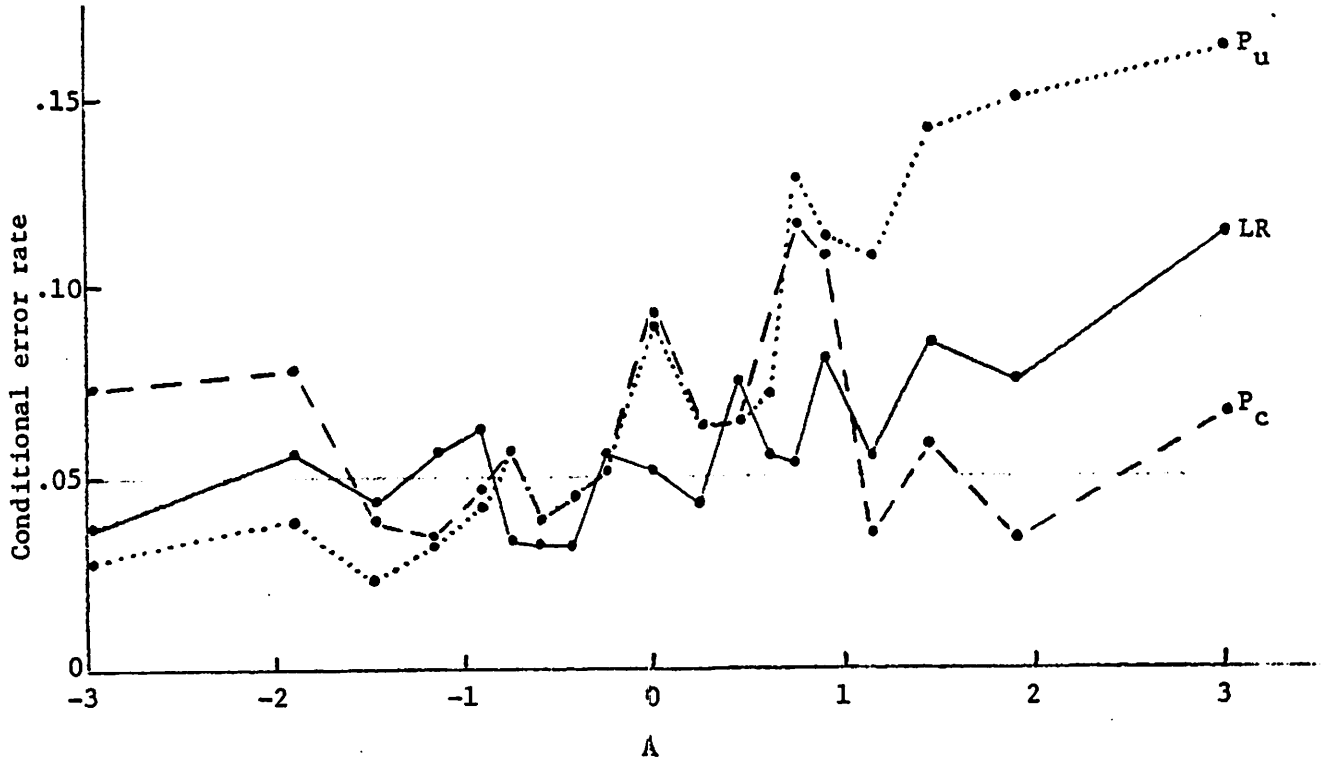


Fig. 2. Error rates by interval values of A for nominal 95% confidence limits on θ . Case $(\theta,m,n) = (1.5,5,20)$.

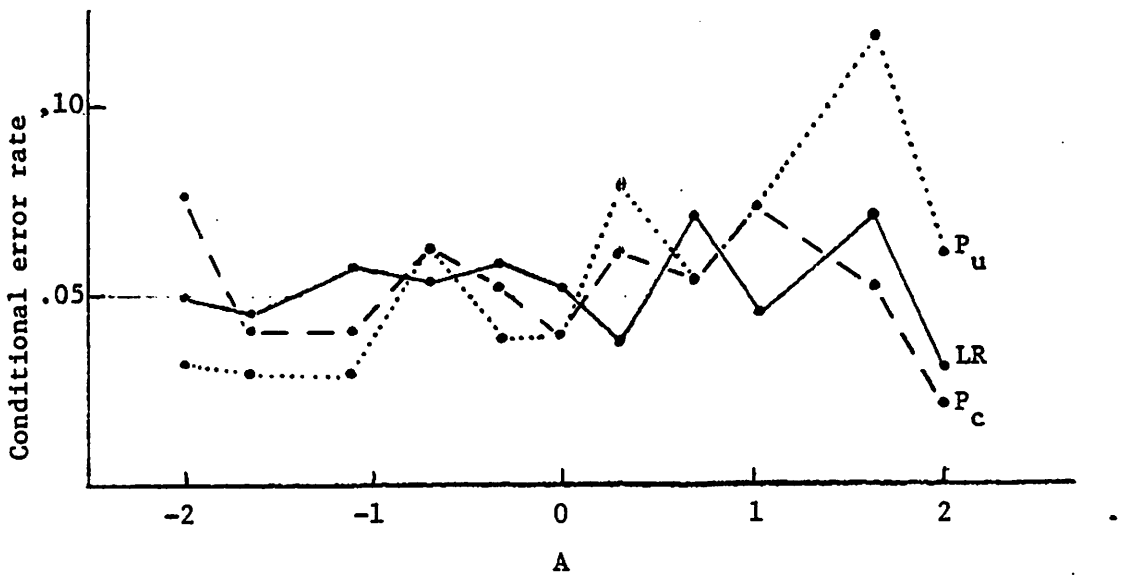


Figure 3. As for Figure 2. Case $(\theta,m,n) = (2.33,5,20)$.

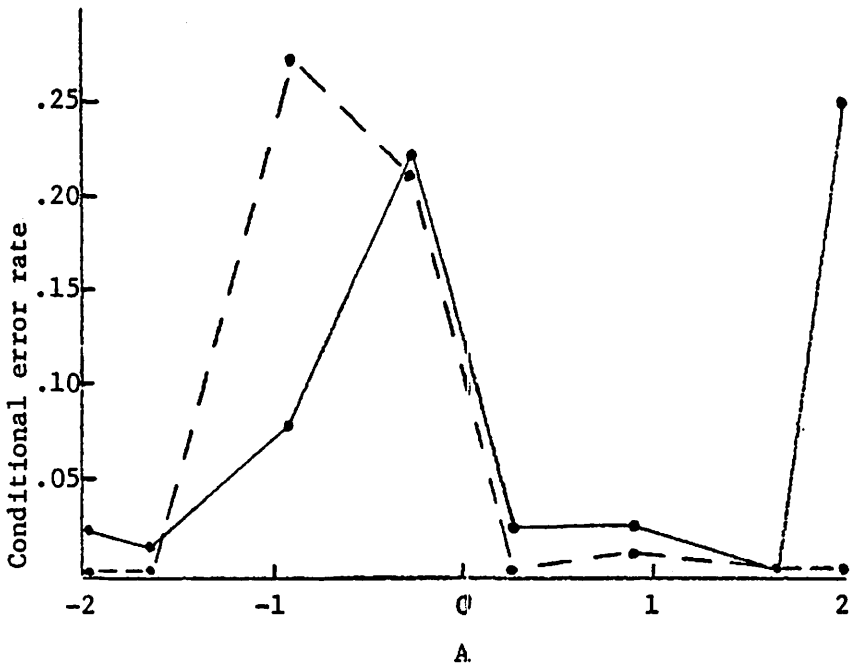


Fig. 4. As for Fig. 2. Case $(\theta,m,n) = (1,10,20)$.

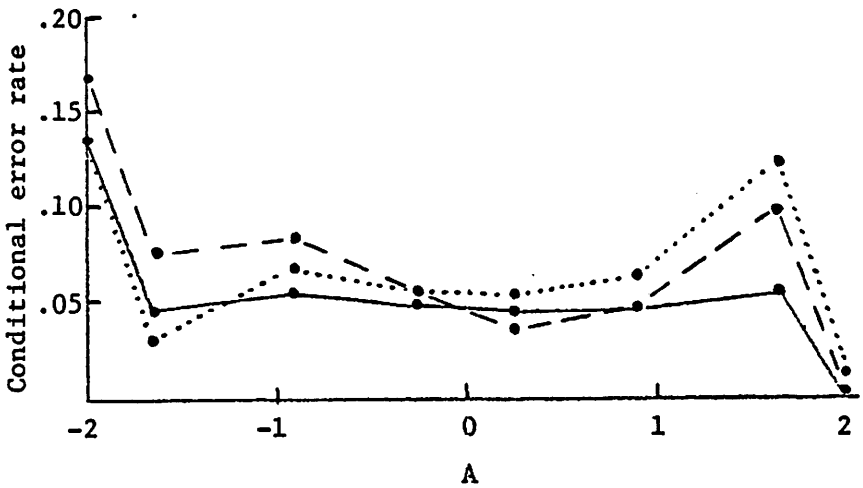
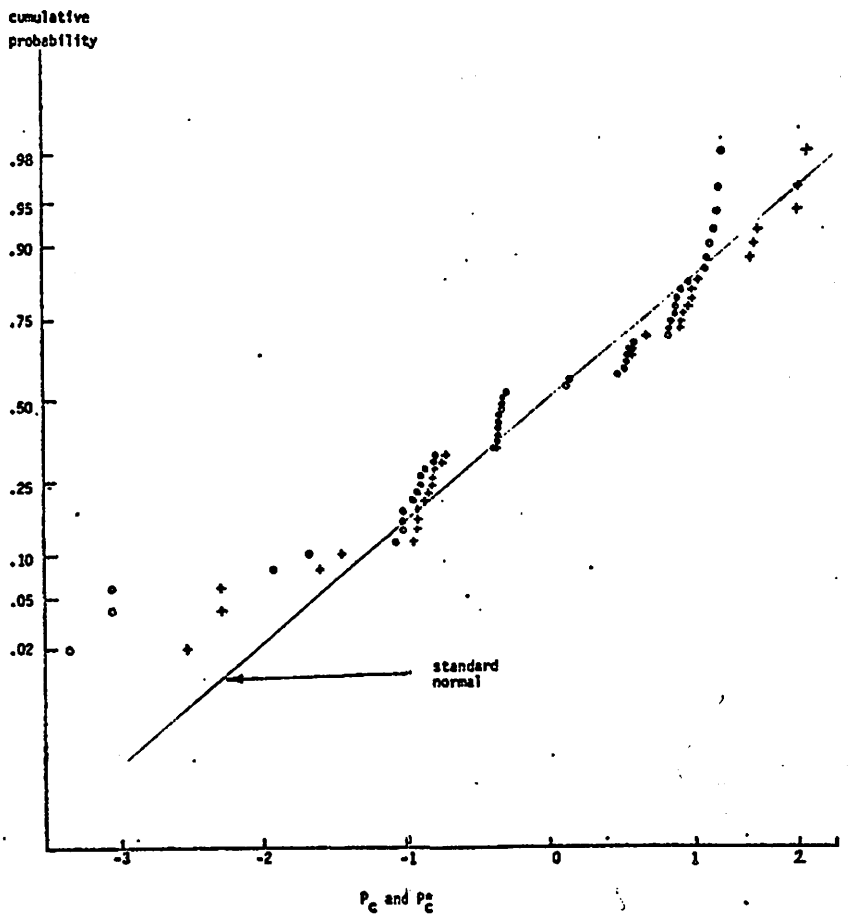


Fig. 5. As for Fig. 2. Case $(\theta,m,n) = (3,10,20)$.

Fig. 6 Normal plots of P_c (o) and P_c^e (+) from 49 samples of the case $(\ell,m,n)=(1,3,20)$



References

Bliss, C.I. & James, A.T. (1966) Fitting the rectangular hyperbola. Biometrics, 22, 573-602.

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

Cox, D.R. & Miller, H.D. (1965) The Theory of Stochastic Processes. London: Methuen

Downham, D.Y. & Green, D.H. (1976) Inference for a two-dimensional stochastic growth model. Biometrika, 63, 551-554.

Efron, B. & Hinkley, D.V. (1978) Assessing the accuracy of the maximum likelihood estimator: Observed versus expected information. Biometrika, 65,

Feigin, P.D.& Reiser, B.(1979) Another look at statistical inference for stochastic processes. Biometrika, 66 (to appear)

Fisher, R.A. (1925) Theory of statistical estimation. Proc. Camb. Phil. Soc., 22, 700-725.

Sprott, D.A. (1975) Application of maximum likelihood methods to finite samples. Sankhȳa, 37B, 259-270.