

AD-A052 468

MASSACHUSETTS INST OF TECH CAMBRIDGE DEPT OF MATHEMATICS F/G 12/1  
ESTIMATING THE INFORMATION PARAMETER OF A TWO-WAY TABLE WITH AP--ETC(U)  
JAN 78 W H DUMOUCHEL, N ODEN N00014-75-C-0555  
TR-10 NL

UNCLASSIFIED

| OF |

AD  
A052468



END

DATE

FILMED

5 -78

DDC

AD A 052468

*See back  
page for 1473*

12 *J*

ESTIMATING THE INFORMATION PARAMETER OF A TWO-WAY TABLE  
WITH APPLICATIONS TO ANIMAL COMMUNICATION

BY

WILLIAM H. DuMOUCHEL  
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

NEAL ODEN  
UNIVERSITY OF MICHIGAN

TECHNICAL REPORT NO. 10  
JANUARY 30, 1978

AD NO. —  
DDC FILE COPY

PREPARED UNDER CONTRACT  
N00014-75-C-0555 (NR-042-331)  
FOR THE OFFICE OF NAVAL RESEARCH

DDC  
RECEIVED  
APR 10 1978

*J*  
B

DEPARTMENT OF MATHEMATICS  
MASSACHUSETTS INSTITUTE OF TECHNOLOGY  
CAMBRIDGE, MASSACHUSETTS

**DISTRIBUTION STATEMENT A**  
Approved for public release;  
Distribution Unlimited

ESTIMATING THE INFORMATION PARAMETER OF A TWO-WAY TABLE  
WITH APPLICATIONS TO ANIMAL COMMUNICATION

by

William H. DuMouchel

Neal Oden (University of Michigan)

Technical Report No. 10

January 30, 1978

Prepared Under Contract

N00014-75-C-0555 (NR-042-331)

For the Office of Naval Research

Reproduction in Whole or in Part is Permitted  
for any purpose of the United States Government

Approved for public release; distribution unlimited

DEPARTMENT OF MATHEMATICS  
MASSACHUSETTS INSTITUTE OF TECHNOLOGY  
CAMBRIDGE, MASSACHUSETTS

1. Introduction

Recently a number of studies of animal behavior (Hazlett and Bossert, 1965, 1966; Dingle, 1969, 1972; Altmann, 1965) have been concerned with estimating the dependence between the successive acts of two communicating animals. If one is given  $p = \{p_{jk}\}$ , the probability that act type  $k$  follows act type  $j$ ,  $j = 1, \dots, J$ ;  $k = 1, \dots, K$  where the first act has  $J$  possibilities and the second act has  $K$  possible types, (often  $J = K$ ), a measure of dependence between successive acts is:

$$I(p) = \sum_j \sum_k p_{jk} \log_2(p_{jk}/p_{j.} p_{.k}) ,$$

where  $p_{j.}$  indicates summation with respect to  $k$ , and similarly for  $p_{.k}$ .

If communication is equivalent to dependence, this quantity may be (and has been) construed to measure the amount of "information" transmitted from one animal to another. Other workers have used  $I$  or related measures to assess the information transmitted by the waggle dance of honey bees (Haldane and Spurway, 1954) and the odor trail of fire ants (Wilson, 1962), and to measure dependence between successive phrases in the call of the wood pewee (Chatfield and Lemon, 1970). A good review of the pitfalls involved in the measure of information transfer is given by Cronbach (1955).

In applying this information measure, workers in animal behavior

ON for		
White Section	<input checked="" type="checkbox"/>	
Buff Section	<input type="checkbox"/>	
INDEXED	<input type="checkbox"/>	
ATION		
BY _____		
DISTRIBUTION/AVAILABILITY CODES		
Dist.	AVAIL. and/or	SPECIAL
A		

usually replace  $p_{jk}$  by

$$\hat{p}_{jk} = n_{jk}/n ,$$

where  $n_{jk}$  is the total number of times act type  $k$  is observed to follow act type  $j$  , and  $n = n_{..}$  is the total number of observed pairs of adjacent acts.

The usefulness of  $I(\hat{p})$  has been impaired by lack of knowledge of its distribution. Although a few workers (Chatfield and Lemon, 1970; Oden, 1977) have noted that, under the null hypothesis of no dependence,  $I(\hat{p})$  is closely related to  $G^2$  , the likelihood ratio test of independence in a two-way table (and is therefore roughly related to Pearson's  $X^2$ ) by the equation:

$$\begin{aligned} G^2 &= 2 \sum_j \sum_k n_{jk} \log_e (n_{jk} n_{..} / n_{j.} n_{.k}) , \\ &= 2n \log_e 2 I(\hat{p}) \\ &\doteq 1.39n I(\hat{p}) , \end{aligned}$$

tests of alternate hypotheses have not been performed. Since nonzero values of  $I$  are of interest, for example, in comparing inter-act dependence in various species, or in the same species under various treatments, a method of estimating  $I$  should be of use to behavioral scientists.

## 2. The Distribution of $I(\hat{p})$ when Acts are Dependent.

For convenience, we will discuss the distribution of  $G^2$  rather than  $I(\hat{p})$  , and define the analogous parameter

$$\begin{aligned}\lambda &= \lambda(p) = 2n \log_e 2 I(p) \\ &= 1.39n I(p) .\end{aligned}$$

When  $n$  is large and  $\lambda = I = 0$  (i.e., acts are independent), it is well-known (Wilks, 1938) that  $G^2$  has approximately a chi-squared distribution with  $(J-1)(K-1)$  degrees of freedom. The large sample distribution of  $G^2$  (and thus of  $I(\hat{p})$ ) is more complicated if  $\lambda$  (and thus  $I(p)$ ) is positive. In general, the distribution of  $G^2$ , even for  $n$  large, depends in a complicated way upon the  $p_{jk}$ , which are usually unknown. See Bishop, Fienberg, and Holland (1975, pp. 518-519) for a discussion and Broffitt and Randles (1977) for a related discussion. However, if  $I(p)$  is near 0, but not equal to 0, then it was proved by Wald (1943, Theorem IX), and under simpler assumptions by Feder (1968) and by Davidson and Lever (1970), that when  $n$  is large the distribution of  $G^2$  is approximately that of a non-central chi-squared variable, with  $(J-1)(K-1)$  degrees of freedom, and with noncentrality parameter equal to  $\lambda$ . (The references cited prove this for likelihood ratio tests in general. In the appendix it is shown that their definition of the non-centrality parameter reduces to our definition of  $\lambda$  in the present case. See Johnson and Kotz (1970) for a discussion of the non-central chi-squared distribution, which has mean  $df + \lambda$  and variance  $2df + 4\lambda$ , if  $df$  is the degrees of freedom.) The meaning of the phrase "I(p) near 0 and  $n$  large" is that one envisions a sequence of alternative hypotheses  $p_{jk} \rightarrow p_j \cdot p_k$ , and sample sizes  $n \rightarrow \infty$ , such that the product  $nI(p)$ , or equivalently  $\lambda$ , approaches a fixed number. In any application, of course, there is only one value of  $n$  and  $I(p)$ , and the question then arises of how closely the noncentral chi-squared distribution approximates that of  $G^2$ .

The adequacy of this approximation depends on two factors:

- 1) the size of  $np_{jk}$ , the expected cell frequencies, and
- 2)  $|p_{jk} - p_{j.}p_{.k}|$ , the cell deviations from independence.

The former quantities should be large; the latter, small. How large, and how small, depends on the degree of approximation desired, and to a lesser extent, on the other characteristics of the  $p_{jk}$ 's. Many rules of thumb have been suggested for how large the expected cell sizes should be for chi-squared tests of contingency tables to be valid. Most authorities would probably agree that if every cell in the observed table has at least 2 observations, the approximations are practically adequate. Also, if fewer than half the cells have at least two observations, it is certainly wrong to use this theory. In between these rather broad limits, the results should be interpreted with caution. When many of the cell sizes are too small, the low-frequency types of acts could be excluded, or combined if they group themselves into natural categories. As for the size of  $|p_{jk} - p_{j.}p_{.k}|$ , the simulation study presented in Table I suggests that if the cell sizes are not too small and if every  $|p_{jk} - p_{j.}p_{.k}| \leq .1$ , the noncentral chi-squared approximation to the distribution of  $G^2$  is fairly good for the larger percentage points, though not so good for the lower tail.

Table I presents the results of a computer simulation of the distribution of  $G^2$  for 5 tables (sets  $\{p_{jk}\}$ ). Four of the

True Table ( $p_{jk}$ )	I(p) (df)	Max $ p_{jk} - p_{j \cdot} p_{\cdot k} $ j, k	Sample Size n	Min np <sub>jk</sub> j, k	Noncentrality $\lambda$	Count of times $G^2 \leq \chi^2_{df, \alpha}(\lambda)$ in 1000 simulations				
						Expected $\pm 2$ (Standard Error)				
						25 $\pm$ 10	100 $\pm$ 20	500 $\pm$ 32	900 $\pm$ 20	975 $\pm$ 10
T <sub>1</sub>	.3 .2	.05 (1)	20 100	4 20	.804 4.02	41* 32	104 108	545* 497	903 909	973 976
T <sub>2</sub>	.4 .1	.15 (1)	20 100	2 10	7.71 38.5	22 18	76* 75*	437* 467*	897 913	988* 980
T <sub>3</sub>	.05 .05	.04 (1)	20 100	1 5	2.50 12.5	217* 60*	289* 136*	476 509	849* 878*	961* 967
T <sub>4</sub>	.08 .02	.07 (1)	20 100	.4 2	7.15 35.8	178* 53*	183* 139*	402* 493	892 874*	989* 975
T <sub>5</sub>	.25T <sub>1</sub> .25T <sub>4</sub>	.10 (9)	20 100 500 2500	.1 .5 2.5 12.5	12.7 63.4 317 1586	21 15 16 21	85 77* 93 95	539* 482 481 505	963* 954* 898 911	997* 988* 978 984

TABLE I. A Computer simulation of the distribution of  $G^2$  for 5 sets  $\{p_{jk}\}$ .

\*Differs significantly (2 standard errors) from that expected by the noncentral chi-squared distribution.

tables,  $T_1, \dots, T_4$  have  $J = K = 2$ , and the  $p_{jk}$  are given in the column labeled "True Table" of Table I. Notice that  $T_1$  and  $T_2$  have marginal frequencies of 50-50 for the two types of act, while  $T_3$  and  $T_4$  have a more unbalanced 90-10 marginal distribution. Table  $T_5$  is a larger,  $4 \times 4$ , table formed by dividing the cell probabilities of  $T_1$  through  $T_4$  by 4, and arranging them as indicated. The next columns of Table I contain the information measure  $I(p)$ , the largest value of  $|p_{jk} - p_{j \cdot} p_{\cdot k}|$ , the simulated sample size  $n$ , the smallest expected cell size, and the noncentrality parameter,  $\lambda = 2n(\log_e 2)I(p)$ , for each simulated experiment. For each combination of  $n$  and  $\{p_{jk}\}$ , 1000 random samples from a simple multinomial (neither margin fixed) were taken, and the value of  $G^2$  computed. Computer generated uniformly-distributed pseudo-random numbers were used to generate the samples, computed using a multiplicative congruential generator. (See Ericson and Fox, 1976, pp. 66-69.) The last five columns of Table I give the empirical cumulative distribution of  $G^2$ , namely the number of times (out of 1000) that the observed  $G^2$  was less than or equal to the 2.5, 10, 50, 90, and 97.5 percentiles of the noncentral chi-squared distribution with the corresponding  $df$  and  $\lambda$ . The table entries marked with an asterisk are those values of the observed cdf of  $G^2$  which differed significantly (2 standard errors) from that of the

noncentral chi-squared distribution. As was mentioned above, the upper tails of the observed distributions of  $G^2$  agree well with approximation whenever there are 2 or more expected in each cell, and every  $|p_{jk} - p_{j.} p_{.k}| \leq .1$ . This suggests that approximate lower confidence limits for  $\lambda$ , and thus  $I(p)$ , can profitably be found, using percentiles of the noncentral chi-squared distribution. Since Table I shows that the lower tails of  $G^2$  are not so well approximated for these sample sizes, it seems that rather larger expected cell frequencies, and/or smaller values of  $|p_{jk} - p_{j.} p_{.k}|$  are needed in order to trust upper confidence limits for  $\lambda$ . Fortunately, in most applications one is more interested in establishing lower limits for  $I(p)$ .

### 3. Obtaining Approximate Confidence Limits for $I(p)$ .

If the noncentral chi-squared approximation to the distribution of  $G^2$  is accurate, and if we let  $\chi_{df,\alpha}^2(\lambda)$  be the  $\alpha^{\text{th}}$  percentile of the distribution with degrees of freedom  $df = (J - 1)(K - 1)$  and noncentrality  $\lambda$ , then if the value  $G^2$  is observed, the solution to

$$\chi_{df,\alpha}^2(\lambda) = G^2$$

is an approximate one-sided confidence limit for the true  $\lambda$ ,

with confidence coefficient  $\alpha$  for an upper bound. This is converted to a bound on the information  $I(p)$  by the relation

$$I(p) = \lambda/1.39n.$$

If either  $df$  or  $\lambda$  is quite large, then Johnson (1959) shows that  $(\chi^2 - df - \lambda + 1)/(2\chi^2 + 2\lambda)^{1/2}$  is approximately a standard normal deviate. This leads easily to the following approximate confidence limits  $\lambda_\alpha$ :

$$\lambda_{.5} = G^2 - df + 1$$

$$\lambda_\alpha = \lambda_{.5} + z_\alpha^2 + z_\alpha [2(G^2 + \lambda_{.5}) + z_\alpha^2]^{1/2}$$

where  $z_\alpha$  is the  $\alpha^{\text{th}}$  percentile of the standard normal distribution. For  $df$  and  $\lambda$  only moderately large it is necessary to solve the equation  $\chi_{df,\alpha}^2(\lambda) = G^2$  directly, and Figures A and B are designed to do this easily, for  $\alpha = 2.5\%$  and  $97.5\%$  respectively. One merely finds the value of  $G^2$  on the abscissa, goes up vertically to the curve labeled with appropriate degrees of freedom, interpolating with respect to  $df$  if necessary, and then reads the value of  $\lambda_\alpha$  on the ordinate scale. Figure A is used for an upper limit and

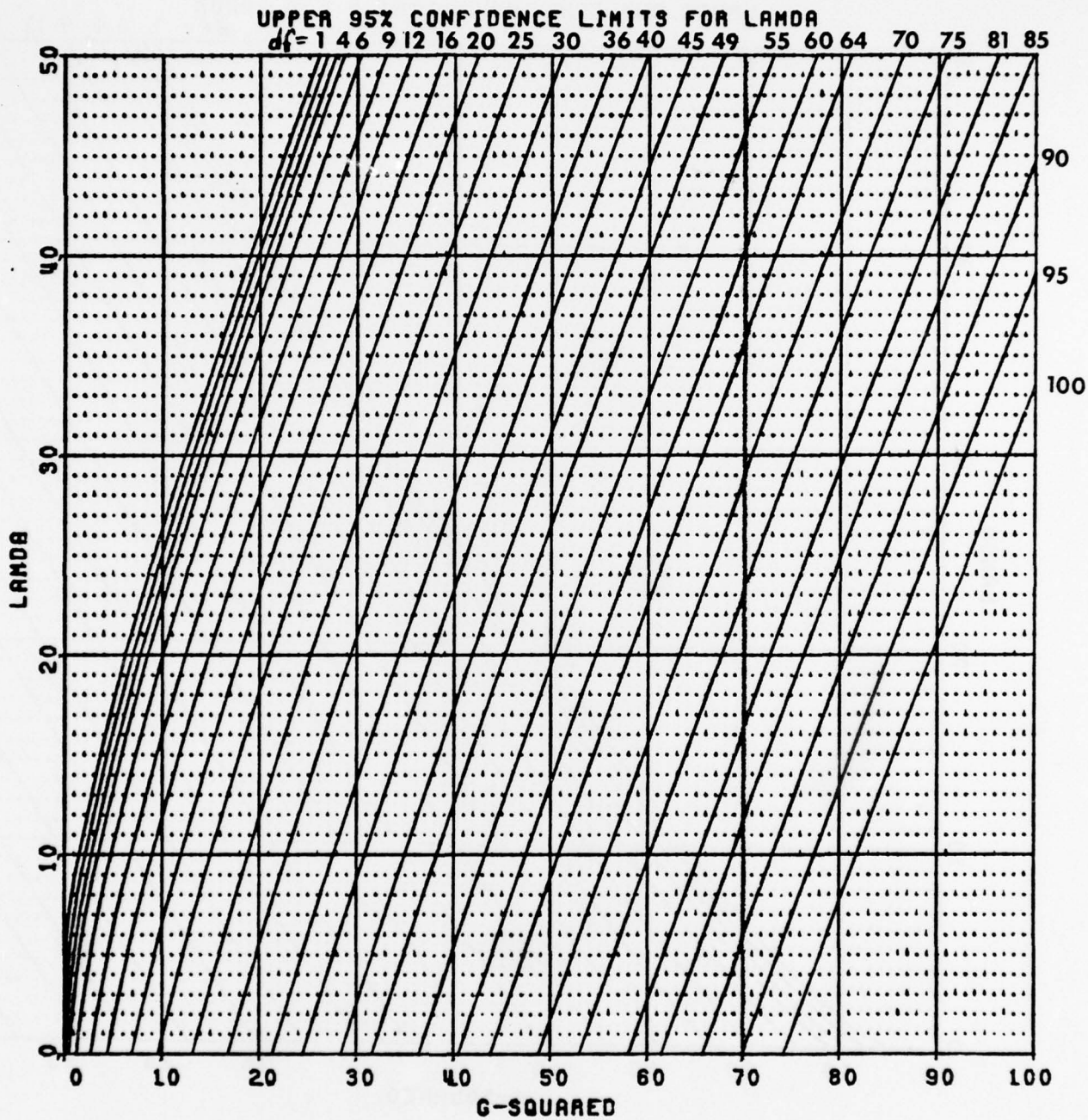


Figure A. 95% confidence intervals for the noncentrality  $\lambda$ , if  $G^2$  is an observation from a noncentral chi-squared  $\chi_{df}^2(\lambda)$  distribution: upper endpoints.

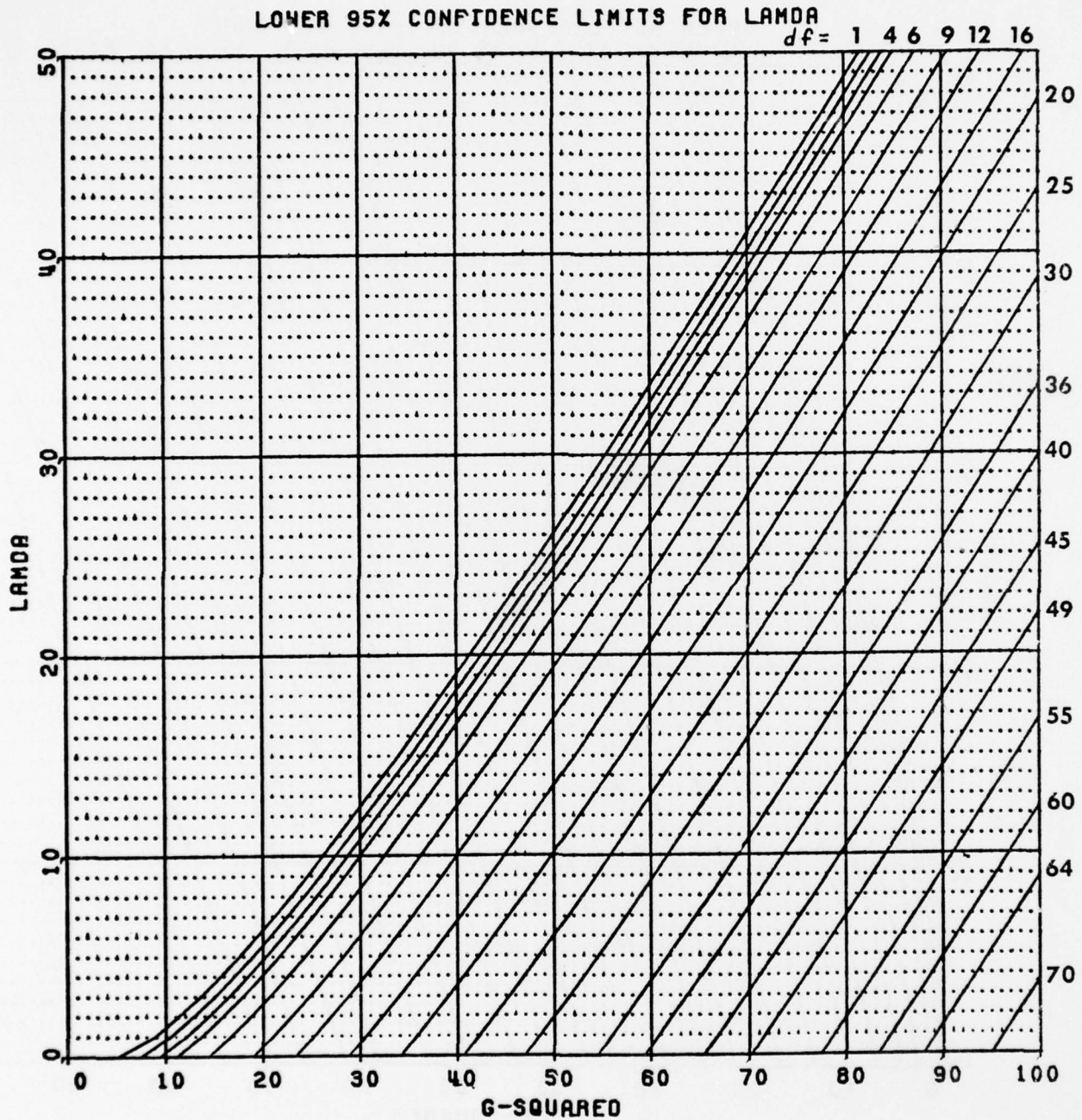


Figure B. 95% confidence intervals for the noncentrality  $\lambda$ , if  $G^2$  is an observation from a noncentral chi-squared  $\chi_{df}^2(\lambda)$  distribution: lower endpoints.

Figure B for a lower limit. The two limits together form an approximate 95% confidence interval for  $\lambda$  and are converted to an interval for  $I(p)$  upon division by  $1.39n$ . The noncentral chi-squared percentiles displayed in the figures were computed using Pearson's (1959) approximation for most values of  $\lambda$  and  $df$ , and a series expansion for the distribution function (see Johnson and Kotz, 1970, vol. 2, p.132) in the few cases where Pearson's approximation was not accurate enough for the figures.

If a point estimate of  $I(p)$  is desired, one may take  $\hat{I} = \lambda_{.5}/1.39n$ , which is approximately unbiased for large  $n$ , and has approximate standard error  $[2df + 4\lambda]^{1/2}/1.39n$ . If  $n$  is large, the difference between  $\hat{I}$  and  $I(\hat{p})$  will not be important. However, if  $n$  is not large compared to  $df$ , then the fact that  $\hat{I}$  "corrects" for  $df$  can be important, at least in terms of mean square error (MSE) of the estimator. For example, in the simulations, for the  $4 \times 4$  Table  $T_5$  (with  $df = 9$ ,  $n = 20$ , and  $I(p) = .458$ ), the 1000 simulated  $G^2$ 's had a sample mean of 20.5 and sample standard deviation of 6.4. Simple calculations then show that  $\sqrt{MSE}$  was .36 for  $I(\hat{p})$  but only .23 for  $\hat{I}$ . The  $\sqrt{MSE}$  would be even smaller for the estimate defined to be  $\max(0, \hat{I})$ , which is of course the estimate one would report in practice. Often it is desired to compare values of  $I(p)$  for different populations. A simple intuitive test is to notice whether confidence intervals for

the separate population parameters overlap. Although this criterion will be more conservative than a proper likelihood ratio test for  $I(p_1) = I(p_2)$ , it is much simpler. If one desires to estimate the common value of  $I(p)$  for two or more populations, one can use the present methods after adding the corresponding values of  $G^2$ ,  $df$ , and  $n$  from the various samples. This is practically equivalent to averaging the various values of  $\hat{I}$ , weighted by their sample size. However this may not be as efficient as the much more complicated maximum likelihood procedure (see Johnson and Kotz, 1970, p. 136).

#### 4. Applications to the Literature

Some examples of the application of the distribution theory discussed above to experimental data are presented in Table II and will now be discussed. It should be noted that the use of information statistics or a chi-squared analysis is not strictly appropriate for most of these data, since there are often too many cells with low expected values. However, published tables with sufficiently high expected values that depict information transfer in animals are rare, and so we have decided to perform the analyses on these data. Our results must certainly be more accurate than the heretofore uniform practice of ignoring the variability of  $G^2$  and merely assuming  $I(p) = I(\hat{p})$ . (Actually, only the Altmann (1964) data, discussed next, flagrantly violate the rules of thumb established in section 2.)

Line	$\hat{I}(p)$	$\hat{I}$	Approximate 95% Confidence Limits	Species	Source	Remarks	n	df	Line
1	2.01	.08	.03-.13	Rhesus	Altmann (1965 and personal communication)		4571	12210	1
2	1.18	1.14	1.08-1.19	"	Altmann's data grouped (see text)		4571	289	2
3	.36	.18	.13-.24	Pyloporurus operculatus	Hazlett & Bossert (1965)	between indiv.	1030	256	3
4	.35	.29	.25-.33	Clibanarius tricolor	"	"	2000	169	4
5	.39	.27	.22-.34	Calcinus tibicen	"	"	1040	169	5
6	.52	.36	.30-.43	Pagurus marshi	"	"	1060	225	6
7	.63	.53	.43-.65	Gonodactylus bredini	Dingle (1969)	1st 10 min-- between indiv.	523	72	7
8	1.03	.88	.72-1.07	"	"	2nd 10 min.	355	72	8
9	.78	.65	.51-.80	"	"	middle 20 min.	393	72	9
10	.7	.53	.38-.70	"	"	last 20 min.	272	64	10
11	.58	.51	.41-.64	"	"	1st 10 min-- within indiv.	447	42	11
12	.79	.75	.63-.90	"	"	last 50 min.	690	42	12
13	.40	.27	.18-.45	Gonodactylus spinulosus	Dingle (1972)	1st 10 min-- within indiv.	235	42	13
14	.64	.52	.04-.65	"	"	1st min-- between indiv.	427	72	14
15	.5	.38	.29-.50	G. spinulosus vs. G. bredini	"	1st min-- between indiv.	446	72	15

Table II. Application of the method for confidence limits to examples in the literature.

Altmann, in his 1965 study of social communication in rhesus monkeys on Cayo Santiago Island, reports on the amount of information held in common between adjacent acts (line 1). This number,  $I(\hat{p}) = 2.01 \text{ bits}^*$ , is larger than that reported for any other species except man. Altmann notes that his study is beset by a number of methodological difficulties, including the fact that he did not distinguish between intra- and inter-individual sequences and the fact that he did not distinguish between various kinds of behavior, such as maternal, courtship, aggressive, etc. From the point of view of assessing the significance of Altmann's findings, another difficulty is the fact that he recognized 120 possible types of acts. The results is a two-way table with 14 400 cells but only 4 571 observations! A formal application of our procedure yields  $\hat{I} = .08$  with a 95% confidence interval for  $I(p)$  of from .03 to .13. The degrees of freedom used in the computation is  $df = 12\ 210$  and reflects the fact that some rows and columns of the observed table are entirely empty (and thus omitted) because some rare acts occurred only at the beginnings or ends of sequences. However, before this data could be used to support or reject Altmann's contention that rhesus monkeys do communicate, one must either develop a theory for the distribution of  $I(\hat{p})$  when most cells are empty, or combine similar types of acts to reduce the degrees of freedom and increase the cell sizes.

Using as a guide the groups of acts suggested by Altmann (1965, Table 2) and using the data kindly provided by Professor

\* Actually Altmann (1965) reports 1.96 bits per act, but a recomputation of his figure (using data kindly supplied to us by Professor Altmann) with a larger computer yields the figure 2.01. The difference between the two values is negligible for our purposes.

Altmann, we lumped the 120 acts into 18 groups. Groups 1-13 each contained a single act, namely all those which occurred 98 times or more and originally numbered 2-5, 9, 19, 25, 41, 43, 45, 50, 51, and 99. The remaining acts were placed into one of the 5 groups 14-18, with rough meanings as follows:

- group 14: fear, submission and compound acts containing fearful and submissive elements
- group 15: attack, threat, and their compounds
- group 16: friendly acts and compounds
- group 17: vocalizations
- group 18: miscellaneous -- other acts hard to group logically.

This grouping has no doubt lumped some dissimilar acts, especially in groups 17 and 18, but our purpose here is not to propose a particular set of group definitions for this data, but merely to demonstrate the effectiveness of grouping in general when estimating  $I(p)$ . The resulting  $18 \times 18$  table has 324 cells, of which 95 have  $n_{jk} \geq 10$ , 100 cells have  $2 \leq n_{jk} \leq 9$ , and 129 have  $n_{jk} = 0$  or 1. Every value of  $|\hat{p}_{jk} - \hat{p}_j \cdot \hat{p}_{.k}|$  is less than .04. The rules of thumb given in section 2 suggest using our approximate theory with a cautious interpretation. The results are given in line 2 of Table II:  $I(\hat{p}) = 1.18$ ,  $\hat{I} = 1.14$ , and  $1.08 < I(p) < 1.19$  with 95% confidence. Thus we find evidence to support the value 1.1 bits per act for communication among rhesus monkeys, but not the value 2.0 reported by Altmann (1965).

Hazlett and Bossert (1965) present data on aggressive communication in a number of hermit crabs, some of which is presented in lines 3-6. They assert that, while Clibanarius tricolor and Calcinus tibicen have roughly the same value for the information parameter, that for Pagurus marshi is higher. Hazlett and Bossert then provide plausible biological reasons why information transfer is more essential to Pa. marshi than Cl. tricolor or Ca. tibicen. Although our analysis does not contradict these explanations, inspection of the confidence intervals shows that there is no strong evidence differentiating these three species, although  $I(p)$  for Py. operculatus is quite surely below that of Pa. marshi. Hazlett (personal communication) has been unable to suggest an interpretation of this result.

In Dingle's (1969) experiments of aggressive communication in the mantid shrimp, Gonodactylus bredini, two adults of the same sex were placed in a finger bowl and allowed to interact for one hour. Data obtained from twenty such pairings were lumped together. Each 60 minutes of data were broken into two 10-minute and two 20-minute periods. A separate transition matrix was formed for each period. To test stationarity of the process, column marginals of temporally adjacent tables were compared using a chi-squared test, revealing significant differences except between the two 20-minute tables. Dingle asserted that information transferred during the second 10-minute period is significantly higher than any of the other periods. Furthermore, this dependence of the information statistic on time was also found in Gonodactylus spinulosus and in matches

between G. brendini and G. spinulosis (not shown). Dingle also found a decline in the frequency of aggressive acts in G. bredini, and a difference in the types of transitions as time went on. He attributes these effects to the establishment of a dominant-subordinate relation in the species mentioned during the second 10 minutes.

Our analysis of the four time periods, which is presented in lines 7-10 of table II, shows that information transferred during the second period is indeed significantly higher than during either the first or the last period, but not significantly different from the third.

Dingle also compiled tables depicting transitions between each individual's act, and the next act of the same animal. **It should be noted that this type of intra-individual analysis confounds two-step dependence with two steps of one-step dependence (Oden, 1977).** However, we have calculated confidence intervals for these data and present them in lines 11,12 and, for G. spinulosis,  
13

In calculating the information statistic for G. spinulosis and matches between G. spinulosis and G. bredini, Dingle (1972) followed essentially the same protocol as for G. bredini. The results are reported in lines 13-15.

Dingle observed that, generally, values for mantid shrimp exceed those for hermit crabs. He suggests that this might reflect the greater seriousness of physical combat in the mantid shrimp.

Although the two independent values of  $\hat{I}$  between individual mantid shrimps (lines 7 and 14) are each higher than all 4  $\hat{I}$ 's from experiments with hermit crabs (lines 3 - 6), this in itself is not much support for Dingle's interpretation, since a Mann-Whitney U-test on the ranks of the  $\hat{I}$  or  $I(\hat{p})$  statistics for mantid shrimp and hermit crabs fails to reveal a significant difference between the two classes ( $\alpha = .13$ , two-tailed) . However, the fact that there is only one overlap in the 95% confidence intervals between the two groups (G. spinulosis, line 14, and P. marshi, line 6), while there is much overlap within the hermit crabs, tends to confirm his assertion.

##### 5. Acknowledgements

We would like to thank Brian Hazlett and Lien-Ju Chao for comments on an earlier version of this work, Stuart Altmann for the use of his data, Herman Chernoff for comments on the proof in the appendix, and Mary Coffey for some help with the computer programming.

## REFERENCES

- Altmann, S. A. (1965). Sociobiology of Rhesus Monkeys II: Stochastics of Social Communication. J. Th. Biol. 8:490-522.
- Bishop, Feinberg and Holland (1975). Discrete Multivariate Analysis, M.I.T. Press, Cambridge, Mass.
- Broffit, J. D. and Randles, R.H. (1977). A Power Approximation for the Chi-Square Goodness of Fit Test: Simple hypothesis Case. JASA 72:604-607.
- Chatfield, C. and R. E. Lemon (1970) Analysing Sequences of Behavioral Events. J. Th. Biol. 29:427-455.
- Cronbach, L. J. (1955) On the non-rational application of information measures in psychology. In H. Quastler (ed) Information Theory in Psychology: 14-26. Free Press, Glecncoe, Ill.
- Davidson R. R. and Leve, W. E. (1970) The limiting distribution of the likelihood ratio statistic under a class of local alternatives. Sankhya Ser. A 32:209-224.
- Dingle, H. A. (1969). Statistical and Information Analysis of Aggressive Communication in the mantis shrimp Gonodactylus bredini Manning. Anim. Behav. 17:561-579.
- Dingle, H.A. (1972). Aggressive Behavior in Stomatopods and the Use of Information Theory in the Analysis of Animal Communication. In H. E. Winn and B. L. Olla (eds), Behavior of Marine Animals: Current Perspectives in Research, vol. 1, Invertebrates: 126-156. Plenum Press, N. Y.
- Ericson, W. A., and Fox, Daniel J. (1976). Simulation with MIDAS. Statistical Research Laboratory, University of Michigan, Ann Arbor, Michigan 48109
- Feder, P. I. (1968). On the distribution of the log likelihood ratio test statistic when the true parameter is "near" the boundaries of the hypothesis regions. Annals Mathematics Statistics 39: 2044-2055.
- Haldane, J. B. S. and Spurway, H. (1954). A statistical analysis of communication in Apis mellifera and a comparison with communication in other animals. Insects Sociaux, 1(3): 247-283.

- Hazlett, B. A. and W. H. Bossert. (1965). A statistical analysis of the aggressive communication systems of some hermit crabs. Anim. Behav., 13:357-373.
- Hazlett, B. A. and W. H. Bossert (1966). Additional observations on the communication systems of hermit crabs. Anim. Behav. 14:546-549.
- Johnson, N. L. (1959). On an extension of the connection between Poisson and  $\chi^2$ -distributions. Biometrika 46:352-363.
- Johnson, N. L. and Kotz, S. (1970). Continuous Univariate Distributions, Vol. 2. Houghton-Mifflin Company, Boston.
- Kullback, S. (1959). Information Theory and Statistics, New York: John Wiley and Sons.
- Oden, N. L. ( 1977 ). Partitioning Dependence in Nonstationary Behavioral Sequences. In B. A. Hazlett (ed.) Quantitative Methods in the Study of Animal Behavior. 203-220. Academic Press.
- Pearson, E. S. (1959). Note on an approximation to the distribution on non-central  $\chi^2$  . Biometrika 46:364.
- Wald, A. (1943). Test of statistical hypotheses concerning several parameters when the number of observations is large. Trans. Amer. Math. Soc. 54:426-482.
- Wilks, S. S. (1938). The large sample distribution of the likelihood ratio for testing composite hypotheses. Ann. Math. Stat. 9:60-62.
- Wilson, E. O. (1962). Chemical communication among workers of the fire ant Solenopsis saevissima (Fr. Smith): 2, an information analysis of the odor trail. Anim. Behav. 10(1,2):134-164.

## APPENDIX

The noncentral chi-squared approximation to the distribution of the likelihood ratio statistic.

This appendix will show that the definition of the non-centrality parameter  $\lambda$  in Wald's (1943) proof of the asymptotic distribution of the likelihood ratio statistic is the same, for  $n$  large and  $I(p)$  small, as  $\lambda(p) = 2n \log_e I(p)$ , where

$$I(p) = \sum_{j,k} p_{jk} \log_2 p_{jk}/p_{j \cdot} p_{\cdot k} .$$

Let  $X$  be a discrete-valued random vector with probabilities  $p(x;\theta)$ . The parameter  $\theta$  belongs to a subset  $\Theta$  of Euclidian space  $E^q$  and is unknown. Let  $g(\cdot)$  be a function over the parameter space which takes values in  $E^r$  ( $r \leq q$ ) and has continuous first partial derivatives. To test the hypothesis  $g(\theta) = 0$  versus  $g(\theta) \neq 0$  using a sample of  $n$  independent observations of  $X$ , consider the likelihood ratio statistic

$$G^2 = 2 \log \left[ \frac{\sup_{\theta \in \Theta} \prod_{i=1}^n p(x_i; \theta)}{\sup_{g(\theta)=0} \prod_{i=1}^n p(x_i; \theta)} \right]$$

Suppose  $\{\theta_n\}$  is a sequence of alternatives converging at rate  $n^{-1/2}$  to a point  $\theta_0$  such that  $g(\theta_0) = 0$  and the matrix  $\partial g(\theta_0)/\partial \theta$  is of full rank, i.e.,  $\theta_n = \theta_0 + C n^{-1/2}$

for some fixed vector  $C$ . Then Wald's (1943) theorem IX states that the limiting distribution of  $G^2$  is the noncentral chi-squared distribution with  $r$  degrees of freedom and non-centrality

$$(1) \quad \lambda = \lim_{n \rightarrow \infty} n [g(\theta_n)]' \Sigma^{-1}(\theta_n) [g(\theta_n)] ,$$

where  $\Sigma(\theta)$  is the asymptotic covariance matrix of the quantity  $n^{1/2} [g(\hat{\theta}) - g(\theta)]$  as  $n \rightarrow \infty$  with  $\theta$  fixed and  $\hat{\theta}$  denoting the maximum likelihood estimator of  $\theta$  based on sample size  $n$ .

Now the quantity

$$(2) \quad \delta(\theta_n) = [g(\theta_n)]' \Sigma^{-1}(\theta_n) [g(\theta_n)]$$

may be interpreted as the squared distance, in  $E^q$ , between the point  $\theta_n$  and the nearest point on the surface  $\{\theta: g(\theta) = 0\}$ , where the metric used to define distance is that given by the Fisher information matrix  $H(\theta_n)$ , defined by

$$H(\theta) = E[(\partial \log p(x; \theta) / \partial \theta)(\partial \log p(x; \theta) / \partial \theta)']$$

$H^{-1}(\theta)$  is the asymptotic covariance matrix of  $n^{1/2} \hat{\theta}$ , and  $\Sigma(\theta) = \dot{g}' H^{-1}(\theta) \dot{g}$ , where  $\{\dot{g}_{\ell m}\} = \{\partial g_{\ell}(\theta) / \partial \theta_m\}$   $\ell = 1, \dots, r$  and  $m = 1, \dots, q$ . It is well known (see e.g., Kullback (1959, p. 26-28) that Fisher information is a limiting form of the Kullback-Leibler information measure: namely for two points  $\theta_0$ ,  $\theta_1$  close together,  $(\theta_1 - \theta_0)' H(\theta_1) (\theta_1 - \theta_0) = 2 \int p(x; \theta_1) \log_e p(x; \theta_1) / p(x; \theta_0) + o(|\theta_1 - \theta_0|^3)$ . Thus, if  $g(\theta_0) = 0$ , the minimum distance  $\delta(\theta_n)$  is

$$(3) \quad \delta(\theta_n) = \min_{\{\theta: g(\theta)=0\}} 2 \sum_x p(x; \theta_n) \log_e p(x; \theta_n) / p(x; \theta) + o(|\theta_n - \theta_0|^3).$$

In our case, where  $x = (j, k)$ ,  $p(x; \theta) = p_{jk}$ , and  $\{\theta: g(\theta) = 0\} = \{p_{jk} : p_{jk} - p_{j.} p_{.k} = 0\}$  it can be checked that the sum on the right of (3) is minimized by  $p_{jk} = p_{j.}^{(n)} p_{.k}^{(n)}$

so that

$$\delta(\theta_n) = \delta(\{p_{jk}^{(n)}\}) = 2 \sum_{jk} p_{jk}^{(n)} \log_e p_{jk}^{(n)} / p_{j.}^{(n)} p_{.k}^{(n)} + o(n^{-3/2}),$$

assuming that  $p_{jk}^{(n)} = p_{jk}^{(0)} + c_{jk} n^{-1/2}$ , and that  $p_{jk}^{(0)} - p_{j.}^{(0)} p_{.k}^{(0)} = 0$  for all  $j, k$ . Thus, combining (1) - (4), we have  $\lambda = \lim_{n \rightarrow \infty} 2 n (\log 2) I(p^{(n)})$ , as desired.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

12 27p.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 10 14 TR-10	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER 11 30 Jan 78
4. TITLE (and Subtitle) Estimating the Information Parameter of a Two-Way Table with Applications to Animal Communication.	5. TYPE OF REPORT & PERIOD COVERED 9 Technical Report.	
7. AUTHOR(s) 10 William H. DuMouchel Neal/Oden (University of Michigan)	6. CONTRACT OR GRANT NUMBER(s) 15 N00014-75-C-0555	
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Mathematics M.I.T., Cambridge, Mass. 02139	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (NR-042-331)	
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 436 Arlington, Virginia 22217	12. REPORT DATE January 30, 1978	
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	13. NUMBER OF PAGES 23	
	15. SECURITY CLASS. (of this report) Unclassified	
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE		
16. DISTRIBUTION STATEMENT (of this Report) Approved for Public Release; Distribution Unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Information, Shannon measure, animal communication, non-central chi-squared distribution, two-way table, confidence limits		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) (see reverse side)		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE  
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

220 021 - Hill

A much-used measure of dependence in two-way tables is

$I(p) = \sum_j \sum_k p_{jk} \log_2(p_{jk}/p_{j.}p_{.k})$ . It is shown that, for large  $n$  and small  $I(p)$ , estimation of  $I(p)$  is equivalent to estimating  $\lambda$ , the noncentrality parameter of the noncentral chi-squared distribution. Methods for calculating estimates and confidence limits for  $I(p)$  are presented. Monte Carlo simulation suggests rule-of-thumb meanings for 'large  $n$ ' and 'small  $I(p)$ ', and shows that the lower confidence limits are more accurate than the upper ones. Applications to several previously published studies of animal behavior are presented.