Reprint Recent Advances in Primatology, Vol. 1 (pp. 279-282), 1969

Estimating Paired-Interaction Probabilities

S. S. Wagner

Printed in Switzerland

S. Karger Basel (Switzerland) New York

Proc. 2nd int. Congr. Primat., Atlanta, GA 1968
vol. 1, pp. 279–282 (Karger, Basel/New York 1969)

ESTIMATING PAIRED-INTERACTION PROBABILITIES

S. S. WAGNER

Yerkes Regional Primate Research Center, Emory University, Atlanta, GA

Suppose that the various members of some group of K individuals, called senders, are directing acts of some kind toward the various members of that or another group of K individuals, called receivers, and that, during some period of observation, they made a total of N such acts. There are then K^2 different sender-receiver combinations, and hence K^2 possible outcomes at each trial, or occurrence of a single act. The problem is to account for the observed distribution of the N acts among the K^2 possible outcomes. The individuals may be countries, people, monkeys, or whatever, the acts may be trade shipments, threats, groomings or whatever, and the receivers may be the same as or different from the senders; for the sake of concreteness, however, we shall discuss the problem in terms of monkeys and groomings, with the group of senders the same as the group of receivers.

Letting n_i and m_i denote the number of times monkey i groomed and was groomed respectively, and n_{ij} the number of times monkey i groomed monkey j, we have $n_i = n_{i1} + \dots + n_{iK}$, $m_i = n_{1i} + \dots + n_{Ki}$, and $n_1 + \dots + n_K = m_1 + \dots + m_K = N$. Such data are sometimes displayed in the form of a 'contingency table'

n_{11}	n_{12}	. . .	n_{1K}	n_1
n_{21}	n_{22}	. . .	n_{2K}	n_2
.	.	.	.	.
.	.	.	.	.
n_{K1}	n_{K2}	. . .	n_{KK}	n_N
m_1	m_2	. . .	m_K	N

whence the n_i's and m_i's may be called the row and column totals (or marginals) respectively.

If the groomings occur independently of one another they will have a multinomial distribution. That is, if at each trial there remains an unchanging probability p_{ij} that it will be monkey i grooming monkey j, then the likelihood l of obtaining, in N trials, exactly the distribution observed will be given by

$$(1) \qquad l = c \cdot p_{11}^{n_{11}} \cdot p_{12}^{n_{12}} \cdot \ldots \cdot p_{1K}^{n_{1K}} \cdot \ldots \cdot p_{K1}^{n_{K1}} \cdot \ldots \cdot p_{KK}^{n_{KK}}$$

where $c = \dfrac{N!}{n_{11}!\, n_{12}! \ldots n_{KK}!}$.

We include the cases where some $p_{ij} = 0$ by agreeing that 0^0 is 1. The p_{ij}'s can be arranged in a table in the same way as the n_{ij}'s.

In the case of an ordinary table (one with no p_{ij}'s known *a priori* to be zero), one may then test the null hypothesis that there is independence between sender and receiver, i.e., that the various ratios obtainable between p_{ij}'s in a given row are the same for all rows (and analogously for columns), which is the same as the hypothesis that $p_{ij} = p_i \cdot q_j$ for all i and j – where the row total $p_i = p_{i1} + \ldots + p_{iK}$ is the probability that monkey i will groom, and the column total $q_i = p_{i1} + \ldots + p_{iK}$ is the probability that monkey i will be groomed. This is called the independence-model, or hypothesis of independence. Under it, (1) becomes

$$(2) \qquad l = c \cdot p_1^{n_1} \cdots p_K^{n_K} \cdot q_1^{m_1} \cdots q_K^{m_K}$$

which, except for c, no longer depends on the individual entries n_{ij} but only upon the marginals n_i and m_i. The foregoing is well-known and discussed in many textbooks of probability and statistics.

In the case of a non-ordinary table (one or more p_{ij}'s known *a priori* to be zero), one may test the null hypothesis of *quasi-independence* between sender and receiver, i.e., that the various ratios obtainable between non-*a-priori*-zero p_{ij}'s in a given row are the same for all rows wherever applicable, which is equivalent to the hypothesis that there exist non-negative parameters P_i and Q_i with $P_1 + \ldots + P_K = Q_1 + \ldots + Q_K = 1$ such that $p_{ij} = \dfrac{P_i \cdot Q_j}{t}$ for every p_{ij} not *a priori* zero – where $t = 1 - \sum_{(i,j)\in\Gamma} P_i \cdot Q_j = \sum_{(i,j)\notin\Gamma} P_i \cdot Q_j$, with the index set Γ defined by: $(i, j) \varepsilon \Gamma$ only if p_{ij} *a priori* zero. (Division by the total, t, is necessary to normalize the p_{ij}, that is, to obtain $\Sigma p_{ij} = 1$.) Thus, for example, if grooming is never self-directed (the zero-diagonal case), t will equal $1 - P_1Q_1 - \ldots - P_KQ_K = \sum_{i \neq j} P_iQ_j$, while if grooming cannot proceed down the order-hierarchy (the upper-right-triangle-zero case), t will equal $1 - \sum_{i<j} P_iQ_j = \sum_{i \geqslant j} P_iQ_j$. This is called the quasi-independence-model, or hypothesis of quasi-independence. The parameters P_i and Q_i may be called the 'theoretical tendencies' of monkey i to groom and be groomed respectively; they are not in general equal to the true tendencies (probabilities) $p_i = \sum_j p_{ij}$ and $q_i = \sum_j p_{ji}$ respectively. Under this model, (1) becomes

$$L = c \cdot \frac{P_1^{n_1} \cdot \ldots \cdot P_K^{n_K} \cdot Q_1^{m_1} \cdot \ldots \cdot Q_K^{m_K}}{t^N} \tag{3}$$

which, as before, no longer depends upon the individual entries n_{ij}.

In order to test either the independence-model or the quasi-independence model, one must have at hand some estimate of the parameters involved. The estimates ordinarily sought are the 'maximum likelihood' estimates, that is, in the first case, the p's and q's making l as large as possible, and, in the second case, the P's and Q's making L as large as possible. Now it is well-known that the unique solution to the first problem is given by

$$p_i = \frac{n_i}{N} \text{ and } q_i = \frac{m_i}{N}$$

for all i; for the second, however, no general solution is known.

In [1] we give all solutions for the zero-diagonal case, showing that for the usual table of n_{ij}'s, the solution is unique and may be obtained by finding the value of α for which $K - 2 - R_1 - \ldots - R_K$ is 0, where

$$R_i = \sqrt{\left(1 - \frac{n_i + m_i}{\alpha}\right)^2 - \frac{4n_im_i}{\alpha^2}}, \text{ and setting}$$

$$P_i = \tfrac{1}{2}\left(1 + \frac{n_i - m_i}{\alpha} - R_i\right) \text{ and } Q_i = \tfrac{1}{2}\left(1 + \frac{m_i - n_i}{\alpha} - R_i\right)$$

for all i; in [2] we give all solutions for the upper-right-triangle-zero case, showing that, for the usual table, the solution is unique and may be obtained by setting

$$P_1 = \frac{n_1}{m_1} \qquad Q_K = \frac{m_K}{n_K}$$

$$P_2 = \frac{n_2(1-P_1)}{m_1+m_2-n_1} \qquad Q_{K-1} = \frac{m_{K-1}(1-Q_K)}{n_{K-1}+n_K-m_K}$$

$$P_3 = \frac{n_3(1-P_1-P_2)}{m_1+m_2+m_3-n_1-n_2} \qquad Q_{K-2} = \frac{m_{K-2}(1-Q_{K-1}-Q_K)}{n_{K-2}+n_{K-1}+n_K-m_{K-1}-m_K}$$

etc.

A general procedure, which would include these two as well as other cases, is obviously to be desired.

Having the P's and Q's, one can find t, and hence the p_{ij}'s. The predicted or expected values for the table are then given by $E_{ij} = p_{ij} \cdot N$ and may be compared with the observed values by means of a χ^2-test (df $= K^2 - 2K + 1 -$ number of cells in Γ) to see whether the disagreement is significant enough to cause rejection of the model.

SUMMARY

One of the assumptions underlying a standard model for the distribution of hits, threats, etc. among a group of monkeys is that each monkey i has 'theoretical tendencies', $P(i)$ and $Q(i)$, to hit and be hit respectively. Various iterative procedures have been suggested for finding the maximum likelihood values of the parameters $P(i)$ and $Q(i)$, but none of them are known to be convergent, nor to yield a solution if convergent. This paper discusses solutions to two cases of that problem.

REFERENCES

1. WAGNER, S. S.: The maximum-likelihood estimate for contingency tables with zero diagonal (to appear).
2. WAGNER, S. S.: The maximum-likelihood estimate for contingency tables with upper-right-triangle zero (to appear).

Author's address: STEPHEN S. WAGNER, Yerkes Primate Research Center, Emory University, *Atlanta, GA 30322* (USA).