

CSIRO
Division of Fisheries and Oceanography

REPORT 108

**Tables of Probabilities
Associated with the Fission
of Replicate Samples
in Classifications**

R. L. Sandland and P. C. Young

1979

**COMMONWEALTH SCIENTIFIC AND INDUSTRIAL RESEARCH ORGANIZATION
DIVISION OF FISHERIES AND OCEANOGRAPHY
P.O. BOX 21, CRONULLA, NSW 2230**

National Library of Australia Cataloguing-in-Publication Entry

Sandland, R.L.

Tables of probabilities associated with the fission of replicate samples in classifications.

(Division of Fisheries and Oceanography report; 108)

Bibliography

ISBN 0 643 02507 3

1. Fission (Biology) – Tables. 2. Probabilities. I. Young, P.C., joint author. II. Title. (Series: Commonwealth Scientific and Industrial Research Organization. Division of Fisheries and Oceanography. Report; 108)

574.16

© CSIRO, 1979

Printed by CSIRO, Melbourne

TABLES OF PROBABILITIES ASSOCIATED WITH THE FISSION OF REPLICATE SAMPLES IN CLASSIFICATIONS

R.L. Sandland¹ and P.C. Young²

¹CSIRO Division of Mathematics and Statistics

²CSIRO Division of Fisheries and Oceanography
P.O. Box 21, Cronulla, NSW 2230

CSIRO Aust. Div. Fish. Oceanogr. Rep. 108 (1979)

INTRODUCTION

Sandland and Young (In Press) described how, by the use of replicated sampling and combinatorial analysis, it is possible to develop a test of random noise and an objective stopping rule for use in hierarchical classification.

The problem is as follows. We have a number (n) of sites which we wish to classify into an *a priori* unknown number of groups on the basis of a number of variates measured at each of r replicates at each site. It is required that the number of replicates at each site be the same. The questions to be answered are (i) does the hierarchy shown by the classification represent real non-random patterns of intersample similarities? (ii) at which level of the classification does this no longer apply?

The statistic of interest in determining our test and stopping rule is $N(n,r;k)$ which is the number of sites which have all their r replicates in any one group at the kth stage (or on classification into k groups). We are interested in whether the event $N(n,r;k) \geq \ell$ the observed number of sites which have all their replicates in one group, is likely to have arisen by chance when all the data are random noise,

i.e. where all the data are replicates of the same situation.

The usual order of the hierarchy is reversed and we consider the fission of one group into two. We start from the point where each individual belongs to the same group and work towards the point where each individual constitutes one group.

The probability is calculated separately for each fission. As each is a split of one group into two, we shall primarily be interested in the probability of the event $N(n_k,r;2) \geq \ell_{k+1}$ where n_k is the number of sites with all their replicates in any one group in the group from the kth stage which is split at the (k+1)th stage and ℓ_{k+1} is the observed number of sites with all their replicates in one of the two new groups formed at the (k+1)th stage. While the paper (Sandland and Young, *ibid.*) gives the formula for the general $P(N(n,r;k) \geq \ell)$, for practical purposes we are only interested in $P(N(n,r;2) \geq \ell)$ and these are the probabilities tabulated here. Each is a measure of the probability of the observed configuration occurring by chance and hence may be used as a test of significance for each fission in the hierarchy.

Sandland and Young (*ibid.*) give the formula for $P(E_k)$, where E_k denotes the event $N(n,r;k) \geq \ell$, as follows:

$$P(E_k) = NF(E_k) / \Delta^k 0^{nr} \quad (1)$$

$$\begin{aligned} \text{where } NF(E_k) = & \left\{ \binom{n}{\ell} \Delta^k 0^{(n-\ell)r+\ell} - \binom{n}{\ell+1} \Delta^k 0^{(n-\ell-1)r+\ell+1} \right. \\ & + \binom{\ell+1}{2} \binom{n}{\ell+2} \Delta^k 0^{(n-\ell-2)r+\ell+2} + \dots \\ & \left. + (-1)^{n-\ell} \binom{n-1}{n-\ell} \binom{n}{n} \Delta^k 0^n \right\} \end{aligned}$$

$$\text{and } \Delta^k 0^{nr} = \sum_{j=0}^k \binom{k}{j} (-1)^j (k-j)^{nr}$$

This report tabulates these probabilities for up to 50 sites and up to 5 replicates per site.

To use the tables, at the k^{th} stage find the group which is split into two to give the $(k+1)^{\text{th}}$ stage and record the number of sites (N_1) with all their replicates in this group. Then record the number of sites (N_2) with all their replicates in either of the two new groups formed by the fission. $P((N_1, r; 2) \geq N_2)$ is then read off from the table and is used to decide whether the data are random noise (at the first stage) or whether the additional classification stage is warranted (at later stages). The full details of how to use the probabilities are found in Sandland and Young (*ibid.*).

The tables given here record the probabilities correct to five decimal places, which is adequate for the purpose in hand. If the ratio

criterion (Sandland and Young, *ibid.*) is to be used, the probabilities will have to be computed directly from equation (1).

On any page of the table, the probabilities are given only up to the N_2 value for which all probabilities associated with it are zero. For higher values of N_2 and all the N_1 values on that page, the probabilities are zero.

REFERENCE

- Sandland, R.L., and Young, P.C.
Probabilistic tests and stopping rules associated with hierarchical classification techniques. *Aust. J. Ecol.* (In Press).

NUMBER OF REPLICATES PER SAMPLE 5

-N1- NUMBER OF SITES WITH ALL REPLICATES IN THE ORIGINAL GROUP
 -N2- NUMBER OF SITES WITH ALL REPLICATES IN ONLY ONE OF THE TWO SUBGROUPS FORMED ON FISSION OF THE ORIGINAL

-N1-

	11	12	13	14	15	16	17	18	19	20
1	.50832	.53905	.56786	.59487	.62019	.64393	.66618	.68704	.70660	.72494
2	.14775	.17029	.19333	.21674	.24038	.26411	.28785	.31150	.33497	.35820
3	.02756	.03507	.04353	.05289	.06313	.07421	.08608	.09869	.11199	.12592
4	.00353	.00503	.00691	.00919	.01193	.01513	.01882	.02302	.02775	.03302
5	.00032	.00052	.00080	.00118	.00168	.00232	.00312	.00411	.00529	.00669
6	.00002	.00004	.00007	.00012	.00018	.00028	.00040	.00057	.00079	.00108
7	0	0	0	.00001	.00002	.00003	.00004	.00006	.00010	.00014
8	0	0	0	0	0	0	0	.00001	.00001	.00001
9	0	0	0	0	0	0	0	0	0	0

-N2-

CSIRO
Division of Fisheries and Oceanography

HEADQUARTERS

202 Nicholson Parade, Cronulla, NSW

P.O. Box 21, Cronulla, NSW 2230

NORTHEASTERN REGIONAL LABORATORY

233 Middle Street, Cleveland, Qld

P.O. Box 120, Cleveland, Qld 4163

WESTERN REGIONAL LABORATORY

Leach Street, Marmion, WA 6020

P.O. Box 20, North Beach, WA 6020