Faculty Publications   **Forestry**

1977

Sex-Ratio Estimation, Sequential Sampling, and the Programmable Packet Calculator

Paul C. Johnson
Arthur Temple College of Forestry and Agriculture, Stephen F. Austin State University

Sex-Ratio Estimation, Sequential Sampling, and the Programmable Pocket Calculator[1]

By P. C. JOHNSON

School of Forestry, Stephen F. Austin State University, Nacogdoches, Texas 75962

Determination of sex-ratios associated with large numbers of insects (trap catch, quadrat samples, etc.) is a recurrent problem in entomological research, frequently involving tedious microscopic examination of individuals. Considerable savings in time and money may result from use of an appropriately designed sequential sampling procedure for estimating sex-ratio. Sequential sampling has enjoyed a deserved popularity in recent years (e.g., see González 1970), as an efficient means of estimating infestation levels of pest populations with regard to their location above or below a critical value (usually the economic threshold). The basic techniques for this type of sequential sampling (reviewed by Waters 1955) are, however, inappropriate for sex-ratio estimation.

Kuno (1969) has presented a somewhat different approach to sequential sampling. Utilizing the regression of mean crowding on mean density to quantify the variance/mean relationship (Iwao 1968, Iwao and Kuno 1968), Kuno (1969) develops a sequential sampling routine which determines mean population density with a predetermined precision in terms of the standard error/mean ratio. Moreover, he has extended this technique (Kuno 1969) to include estimates of proportions from binomially distributed populations.

The routine developed here is similar to that presented by Kuno (1969) for sampling from a binomial distribution of limited size without replacement of sampled individuals (i.e., sampling from the hypergeometric distribution), with the following modifications:

1. An unbiased estimate of the standard error $SE_{\hat{p}}$ of the estimated proportion $\hat{p}$ has been employed, as recommended by Steel and Torrie (1960).

2. Preset precision levels expressed as confidence limits are substituted for precision levels in terms of the standard error/mean ratio as used by Kuno (1969).

3. Replacement of graphic techniques with the programmable pocket calculator has increased the ease of application, one of the major disadvantages of the graphic approach (Kuno 1969).

Although developed for estimation of sex-ratios associated with southern pine beetles caught on sticky traps during a study on flight dispersal, the routine is applicable to sequential sampling of any binomially distributed character where: (1) samples are drawn at random without replacement and (2) the population size is known or may be estimated. Use of a programmable calculator greatly facilitates application of the routine, but the calculations are easily handled by any standard calculator.

Description of the Statistic

Let N represent the total number of individuals in the population being sampled (e.g., number of insects in a single trap). [An estimate does not seriously alter the results when values of N are large.] Let $\hat{p}$ represent the

estimated proportion of individuals in one class over all sampled individuals, and $\hat{q} = 1-\hat{p}$, or the proportion of individuals in the other class. If n is the total number of individuals sampled, then the standard error of the estimate $(SE_{\hat{p}})$ for sampling from a hypergeometric distribution is given as (Steel and Torrie 1960):

$$SE_{\hat{p}} = \sqrt{\frac{\hat{p}\hat{q}}{n-1}\left(\frac{X-n}{X}\right)};\qquad (1)$$

where the term $\left(\dfrac{X-n}{X}\right)$ is the proportion of the total population left unsampled.

If we let B be the bounds on the error associated with $\hat{p}$, and let C be a constant determining the confidence level associated with these bounds, then:

$$B = C\sqrt{\frac{\hat{p}\hat{q}}{n-1}\left(\frac{X-n}{X}\right)}\qquad (2)$$

Since repeated estimates of p may be expected to approach a normal distribution about the true value of p, setting C=1.64 will insure that our estimate of p will fall within $\hat{p}\pm B$ 90% of the time (i.e., 90% confidence level). Similarly, setting C=1.96 corresponds to a 95% confidence level, and setting C=2.58 corresponds to a 99% confidence level.

Thus, for any desired confidence level and any number of individuals sampled, we may use equation (2) to calculate the bounds on the error of our estimate, $\hat{p}$. Sequentially drawing random samples of size k_i from the population and comparing B (based on the total number of individuals sampled, $\Sigma k_i = n$), to a preset level of precision (B_E) allows us to terminate sampling as soon as a sufficiently precise estimate of p has been reached.

It should be noted that individual sample size, k_i, at each sampling step (i) may be variable since $\hat{p}$ and B are based on the total sample, n. This allows large sample size during the early stages of the procedure, and progressively smaller sample sizes as B approaches B_E, thus allowing the user to maximize efficiency of the operation. Moreover B_E and C may be varied independently to provide any degree of precision and confidence level desired.

The Sampling Routine

The sequential sampling routine proceeds as follows:

1. Select a random sample of k_i individuals, in which I_i represents the number of individuals in Class I;

   Calculate $\hat{p} = \dfrac{\Sigma I_i}{n}$ and $\hat{q} = 1-\hat{p}$.

2. Calculate B from equation (2) for the desired confidence level.

3. If B is greater than a selected value, B_E, continue sampling at step 1).

Direction of flow ··· mandatory

Direction of flow ··· optional

(START)   **Start or end of sampling program**

X, Y, Z, T   **Keyboard input into operational stack**

X   **Display of output**

Relation   **Conditional branching**

Process   **Processing procedure**

X →R2 Y →R6   **Storage in addressable registers (R1–R9)**

(A)   **Labeled subroutine**

(RS)   **Run / Stop ... pause for display and / or input**
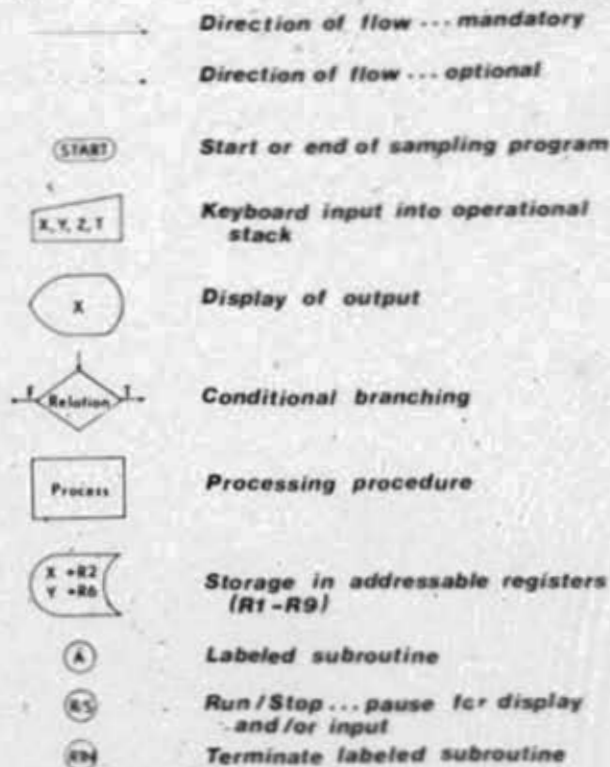
(RTN)   **Terminate labeled subroutine**

Fig. 1.—Symbols used in flowcharting the sequential sampling procedure.

4. If B is less than or equal to B_a, stop sampling and accept $\hat{p} \pm B$ as your estimate at the specified confidence level.

This procedure was programmed for the Hewlett-Packard HP-65* programmable calculator as outlined in the flowchart in Fig. 2. Flowchart symbols are provided in Fig. 1.

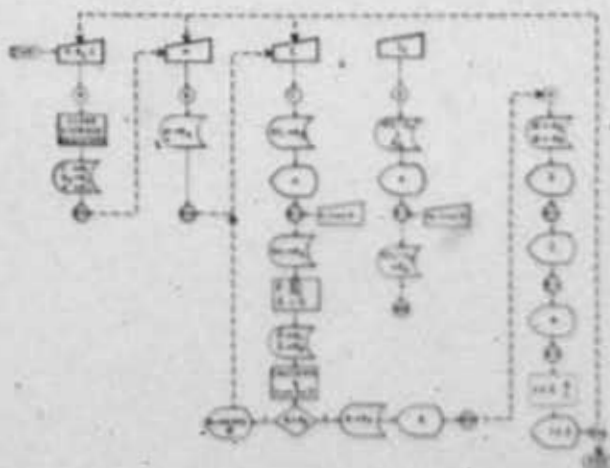Fig. 2 outlines the most general case for the sequential sampling being discussed. Various modifications may simplify the programming as dictated by the sampling situation. For example, the user will frequently select a confidence level and precision which he wishes to apply throughout the sampling program. In this case, an initialization routine storing the selected values for C, B_a and K without data input may be substituted for subroutine A. Similarly, if a standard sample size, K, is to be drawn throughout, the pause for input of k_i may be deleted.

The decision box ending subroutine C, the blinking zero display (indicating need for further sampling), etc. are "window-dressing." A simple visual display of B would suffice. The inclusion of an error correction subroutine such as subroutine E in Fig. 2 is recommended, but not essential (I_i and k_i correspond to the erroneous I_i and k_i).

The format of the output subroutine D, is another area in which the particular sampling situation may suggest modification of the program. For example, numbers of individuals in each class ($\hat{p}N$ and $\hat{q}N$) may be the appropriate output units, $SE\frac{3}{2}$ may be deleted, etc.

Copies of the individual steps for the HP-65 program outlined in Fig. 2, including a default initialization routine if C, B_a and K are not set, are available on request.

Estimating Total Sampling Size

The equation,

$$n = \frac{pq + \dfrac{B_a^2}{C}}{\dfrac{pq}{n} + \dfrac{B_a^2}{C}}, \qquad (3)$$

may be used to estimate the avg number of individuals which must be sampled to obtain $B \le B_a$ for a population of size N at the confidence level determined by C. Equation (3) was used to generate sample size curves for populations of 100, 1,000 and 10,000 individuals at confidence levels of 90%, 95% and 99% (Fig. 3-5). The shaded bands represent the range of probable sample size (n expressed as %N) for specified degrees of precision

Fig. 2.—Flowchart for sequential sampling routine as programmed for the HP-65 programmable calculator (Hewlett-Packard).
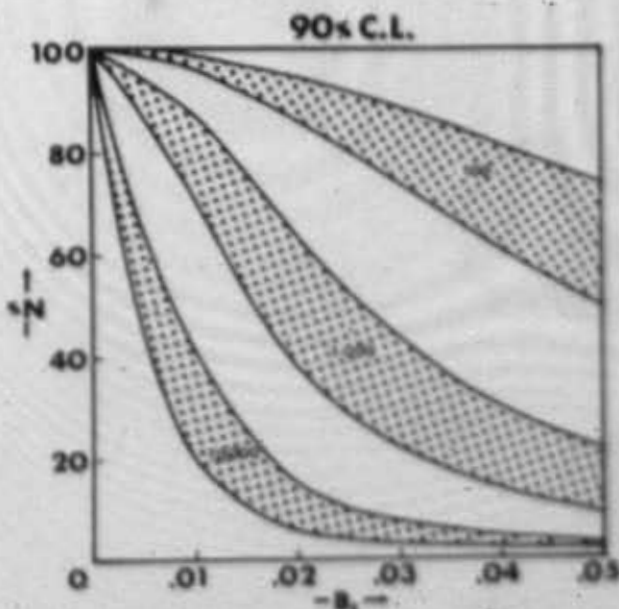

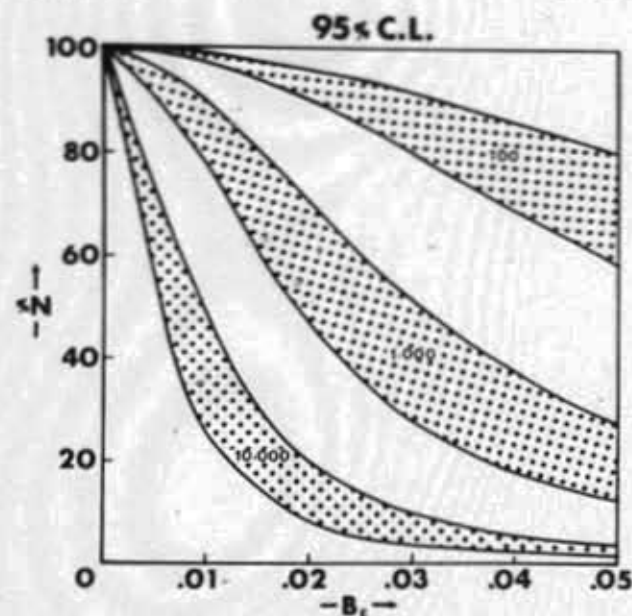Fig. 3.—Sample Size Curves for populations of 100, 1000 or 10,000 individuals at the 90% confidence level.

Fig. 4.—Sample Size Curves for populations of 100, 1000 or 10,000 individuals at the 95% confidence level.

($B_E = 0.0$ to $B_E = 0.05$). The upper limit is set by $p=0.5$, while the lower limit represents both $p=0.1$ and $p=0.9$. Therefore, the greater the discrepancy between p and q, the lower the total sample size required.

The sample size curves may be used in planning an appropriate B_E and confidence level based on projected population levels as well as for determining appropriate sample sizes (k_i) for use in the sequential sampling procedure. Consideration of population levels involved in a particular study and the required sample size (n) to obtain a specific degree of precision allows the setting of reasonable bounds on the precision of the estimate $\hat{p}$.

For example (Fig. 5), in sampling from a population of 10,000 individuals at the 99% confidence level with $B_E = 0.01$, we might expect a total sample size (n) of between 3700 (37% of 10,000) and 6200 (62% of 10,000). If this would be prohibitively large, we may sacrifice precision by selecting $B = 0.02$ and use a 95% confidence level (Fig. 4). The estimated sample size then ranges from 800 (8% of 10,000) to 1900 (19% of 10,000).

Discussion

Sequential sampling can be an efficient technique for estimating proportions in a binomial distribution. Limitations associated with the technique presented here include:

1. random selection of samples
and 2. estimation of total population size.

Limitation number (1) applies to virtually all sampling programs and must be considered during the design of any sampling procedure. Limitation number (2) is peculiar to sampling from a finite population of limited size (such as trap catch). Under the conditions for which the routine was developed (i.e., sex-ratio estimation for trapped southern pine beetles), this was not a problem since the beetles were to be counted (or estimated) anyway. In some sampling situations, however, this may be a severe drawback to the routine, and the consequences of inaccurate estimation of population size must be considered.

The accuracy of the population estimate becomes critical only when the total sample size (n) approaches the population size (N), and it influences primarily the bounds on the error of the estimate $\hat{p}$ (B). Indirectly, it influences the estimate $\hat{p}$ through modification of the sample size required. Where the accuracy of the population estimate is in doubt, an overestimate will insure over sampling and increase confidence in $\hat{p}$, although the resultant confidence limits will be misleading (i.e., the true B will be less than that reported).

Use of a programmable pocket calculator increases the ease of application of the sequential sampling routine when compared to standard graphic techniques. The Kuno (1969) approach to sequentially sampling the binomial distribution involved preparation of a 'stop line' graph for each population size (awkward under field conditions and inordinately time consuming where a wide range of population sizes are encountered). With the pocket calculator, the 'stop line' graph is, in effect, stored internally and redesigned for each new population size with a single keystroke entry. Successive entry of the number of individuals of one type for each sample triggers the comparison to the 'stop line' and results are immediately obtainable following the decision to stop sampling. It is felt that programmable pocket calculators have broad application to sequential sampling in general and represent a major improvement over the graphic approach.
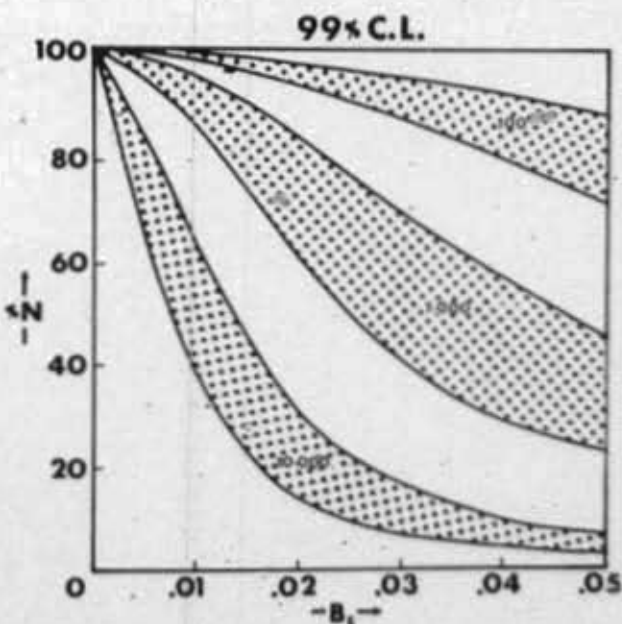
Fig. 5.—Sample Size Curves for populations of 100, 1000 or 10,000 individuals at the 99% confidence levels.

REFERENCES CITED

González, D. 1970. Sampling as a basis for pest management strategies. Tall Timbers Conference on Ecological Animal Control by Habitat Management 2: 83–101.

Iwao, S. 1968. A new regression method for analyzing the aggregation pattern of animal populations. Res. Popul. Ecol. 10: 1–20.

Iwao, S., and E. Kuno. 1968. Use of the regression of mean crowding on mean density for estimating sample size and the transformation of data for the analysis of variance. Ibid. 10: 210–14.

Kuno, E. 1969. A new method of sequential sampling to obtain the population estimates with a fixed level of precision. Ibid. 11: 127–36.

Steel, R. G. D., and J. H. Torrie. 1960. Principles and Procedures of Statistics. McGraw Hill, New York.

Waters, W. E. 1955. Sequential sampling in forest insect surveys. Forest Sci. 1: 68–79.

PLANT PROTECTION CONFERENCE
March 22–25, 1978

The Plant Protection Conference, jointly organized by the Malaysian Plant Protection Society and the Rubber Research Institute of Malaysia, will be held at the International Hall, Hotel Merlin, in Kuala Lumpur from Wednesday, March 22 to Saturday, March 25, 1978. Emphasis will be given to studies and methods of control of pests, diseases and weeds on crops of economic importance in the Southeast Asian and Pacific region. All correspondence should be addressed to:

> *Secretary*
> Plant Protection Conference
> Rubber Research Institute of Malaysia
> P.O. Box 150
> Kuala Lumpur 01-02
> MALAYSIA

AAAS PROJECT ON THE HANDICAPPED IN SCIENCE SEEKING NAMES FOR DIRECTORY

The Project on the Handicapped in Science of the American Association for the Advancement of Science (AAAS) is developing a directory of scientists with handicapping conditions which will be distributed to public agencies, handicapped scientists, and consumer organizations to facilitate communication among them. Research for, and publication of the directory is supported by the National Science Foundation (Grant No: SP177-19913).

Each directory entry will contain basic personal and professional biodata, nature and age of onset of handicapping condition, and a summary of educational and professional experience including problems faced and overcome. The scientists will be asked if they would be willing to act as advisors, consultants and/or evaluators on many levels concerning science and the handicapped. This information will also be included in the directory. It is anticipated that the directory will identify handicapped scientists of specific scientific specialization who are willing to share experiences and information on coping strategies; educate the scientific community concerning their existence and their problems; organize action when needed; and provide the basis for problem analysis and professional planning for scientific societies, educational institutions, consumer organizations, and federal agencies.

The project defines science broadly to include all the natural and social sciences, mathematics, engineering and medicine—all fields for which a rigorous scientific education is required. It will seek to include those with degrees in these areas or equivalent training who are working in scientific, medical or technological fields. The directory will also include individuals trained in science but working in other areas or presently unemployed.

At this time the AAAS has names of over 500 disabled scientists who have been serving as a resource group to the Association. The Association is interested in increasing this number so that the directory will include as many handicapped scientists as possible. All handicapped scientists and graduate students of science not currently included in the resource group are urged to participate in the project by identifying themselves to Martha Ross Redden, Project Director, or Janette Alsford Owens, Project Assistant, Project on the Handicapped in Science, Office of Opportunities in Science, AAAS, 1776 Massachusetts Avenue NW, Washington DC, 20036, telephone number (202) 467-4497 (TTY or voice).

INTERNATIONAL COMMISSION OF ZOOLOGICAL NOMENCLATURE
A.N.(S) 103
ANNOUNCEMENT

The required six months' notice is given of the possible use of plenary powers by the International Commission on Zoological Nomenclature in connection with the following names listed by case number: (see *Bull. Zool. Nom.* 34, part 2, 31st August, 1977).

400   *Trombidium akamushi* Brumpt, 1910 (Acarina): proposed validation.

2115   *Glyphipterix* Hübner, [1825] (Lepidoptera, GLYPHIPTERYGIDAE): proposed designation of a type-series.

2130   *Stethaspis* Hope, 1837 (Coleoptera): proposed designation of a type-species.

2186   PIERIDAE Duponchel, [1835]: proposal to give precedence over COLIADINAE Swainson, 1827 (Insecta, Lepidoptera).

2193   *Campylosteira* Fieber, 1844 (Hemiptera): designation of type-species.

2194   *Baeocera* Erichson, 1845 (Coleoptera): designation of type-species.

2204   *bjerkandrella*, *Tinea*, Thunberg, 1784, and *cardui*, *Phalaena* (*Noctua*) Hübner, 1790 (Insecta, Lepidoptera): proposed conservation.

2201   MORPHIDAE Boisduval, 1836 (Insecta, Lepidoptera): request for revision of the Official List.

2206   HENICOPIDAE Pocock, 1901: proposal to give precedence over CERMATOBIIDAE Haase, 1885 (Myriapoda: Chilopoda).

2207   *Athelges* Gerstaecker, 1862 (Crustacea, Isopoda): proposed conservation.

Comments should be sent in duplicate (if possible within six months of the date of publication of this notice), citing case number to:

> R. V. Melville, The Secretary
> International Commission Zoological Nomenclature
> c/o British Museum (Natural History)
> Cromwell Road
> LONDON, SW7 5BD
> England

Those received early enough will be published in the Bulletin of Zoological Nomenclature.