Weighted Distributions and Size-Biased Sampling with Applications to Wildlife Populations and Human Families
Author(s): G. P. Patil and C. R. Rao
Source: *Biometrics*, Vol. 34, No. 2 (Jun., 1978), pp. 179-189
Published by: International Biometric Society
Stable URL: http://www.jstor.org/stable/2530008
Accessed: 11/11/2014 10:00

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

International Biometric Society is collaborating with JSTOR to digitize, preserve and extend access to *Biometrics.*

http://www.jstor.org

This content downloaded from 50.78.147.85 on Tue, 11 Nov 2014 10:00:25 AM
All use subject to JSTOR Terms and Conditions

BIOMETRICS 34, 179–189
June, 1978

Weighted Distributions and Size-Biased Sampling with Applications to Wildlife Populations and Human Families

G. P. PATIL

The Pennsylvania State University, University Park, Pennsylvania 16802, U.S.A.

C. R. RAO

Indian Statistical Institute, New Delhi 110029, India

Summary

When an investigator records an observation by nature according to a certain stochastic model, the recorded observation will not have the original distribution unless every observation is given an equal chance of being recorded. A number of papers have appeared during the last ten years implicitly using the concepts of weighted and size-biased sampling distributions.

In this paper, we examine some general models leading to weighted distributions with weight functions not necessarily bounded by unity. The examples include: probability sampling in sample surveys, additive damage models, visibility bias dependent on the nature of data collection and two-stage sampling. Several important distributions and their size-biased forms are recorded. A few theorems are given on the inequalities between the mean values of two weighted distributions. The results are applied to the analysis of data relating to human populations and wildlife management.

For human populations, the following is raised and discussed: Let us ascertain from each male student in a class the number of brothers, including himself, and sisters he has and denote by k the number of students and by B and S the total numbers of brothers and sisters. What would be the approximate values of $B/(B + S)$, the ratio of brothers to the total number of children, and $(B + S)/k$, the average number of children per family? It is shown that $B/(B + S)$ will be an overestimate of the proportion of boys among the children per family in the general population which is about half, and similarly $(B + S)/k$ is biased upwards as an estimate of the average number of children per family in the general population. Some suggestions are offered for the estimation of these population parameters. Lastly, for the purpose of estimating wildlife population density, certain results are formulated within the framework of quadrat sampling involving visibility bias.

1. *Introduction*

When an investigator records an observation by nature according to a certain stochastic model, the recorded observation will not have the original distribution unless *every* observation is given an equal chance of being recorded. For example, suppose that the original observation X has $f(x)$ as the *pdf* (which may be probability when X is discrete and probability density when X is continuous), and that the probability of recording the observation x is $0 \leq w(x) \leq 1$, then the *pdf* of X^w, the recorded observation is

$$f^w(x) = \frac{w(x) f(x)}{\omega} \tag{1.1}$$

Key Words: Weighted distribution; Probability sampling; Damage model; Quadrat sampling; Visibility bias.

This content downloaded from 50.78.147.85 on Tue, 11 Nov 2014 10:00:25 AM
All use subject to JSTOR Terms and Conditions

where ω is the normalizing factor obtained to make the total probability equal to unity. Thus ω may be referred to as the *visibility factor*. Note that $f^w = f$ if and only if $w(x)$ is a constant.

Rao (1965) introduced distributions of the type (1.1) with an arbitrary non-negative weight function $w(x)$ which may exceed unity and gave practical examples where $w(x) = x$ or x^α are appropriate. He called distributions with arbitrary $w(x)$ weighted distributions. In this paper we use this definition of a weighted distribution with arbitrary $w(x)$, of which (1.1) is a special case. The weighted distribution with $w(x) = x$ is also called a sized biased distribution. We shall show in Section 2 of this paper how the weight $w(x) = x$ occurs in a natural way in many sampling problems.

A number of papers have appeared during the last ten years implicitly using the concepts of weighted and size-biased sampling distributions, the results of which are briefly surveyed (with relevant references) in a recent paper by the authors (Patil and Rao 1977). A study of size-biased sampling and related form-invariant weighted distributions was made by Patil and Ord (1975).

In this paper, we examine some general models leading to weighted distributions with weight functions not necessarily bounded by unity. The results are applied to the analysis of data relating to human populations and wildlife management.

2. *General Models Leading to Weighted Distributions*

2.1 *Probability Sampling in Sample Surveys*

A well known example is what is called pps (probability proportional to size) sampling in sample survey methodology where the original pdf of a variable is changed according to a given design of selection of samples to improve the efficiency of estimators of unknown parameters.

Let Y be a *rv* concomitant with the main *rv* X under study and $p_1(y)$, $p_2(x \mid y)$, $p_3(x)$ and $p_4(y \mid x)$ be the marginal *pdf* of Y, the conditional *pdf* of X given Y, the marginal pdf of X and the conditional *pdf* of Y given X respectively. Let us first observe a value y of a *rv* Y^w with *pdf*

$$g(y)/\, p_1(y) \Big/ \int g(y)\, p_1(y) dy \tag{2.1}$$

where $g(y)$ is a chosen non-negative function. Then an observation is made on X using the conditional distribution $p_2(x \mid y)$. The *pdf* of the resulting *rv* X^w is

$$\frac{\int p_2(x \mid y)\, g(y)\, p_1(y) dy}{\int g(y)\, p_1(y) dy} = p_3(x)\, w(x) \Big/ \int p_3(x)\, w(x) dx \tag{2.2}$$

where

$$w(x) = \int p_4(y \mid x)\, g(y)\, dy \tag{2.3}$$

which is a weighted distribution of X with the weight $w(x)$ not necessarily bounded by unity. Further $w(x)$, in such a situation, may involve the unknown parameters of $p_3(x)$. Thus, for the observed value X^w the appropriate *pdf* is the weighted *pdf* (2.2). In practice $p_1(y)$ is completely known and $g(y)$ is chosen to maximize the efficiency of inference on the unknown parameters of interest in $p_3(x)$.

This content downloaded from 50.78.147.85 on Tue, 11 Nov 2014 10:00:25 AM
All use subject to JSTOR Terms and Conditions

2.2 Damage Model of Rao (1965)

Suppose that we are sampling from a *pdf* $f(x)$, but while realizing an observation x it goes through a 'damage process' with the result that we finally have an observation z from the conditional distribution with *pdf* $c(z \mid X = x)$ which we may denote simply by $c(z \mid x)$. The marginal *pdf* of the observed value z is

$$\int c(z \mid x) f(x)\, dx \tag{2.4}$$

which has the form of a mixture of distributions.

Let us suppose that an observation is recorded only when the original value is unchanged through the damage process. The *pdf* of such an observation X^w is

$$c(x \mid x) f(x)/E(c(X \mid X)), \tag{2.5}$$

which is a weighted distribution with $w(x) = c(x \mid x)$, where $c(x \mid x) = c(Z = x \mid X = x)$, the conditional probability that an observation x remains unchanged. Examples of such distributions are considered by Rao (1965) and Rao and Rubin (1964). A truncated distribution is a special case of (2.5), where $c(x \mid x)$ takes the value zero in a certain region of the sample space of X and unity in the complement.

2.3 Visibility bias

Let us consider a discrete random variable X with *pdf* $f(x)$. For instance, X may be the number of individuals in a group or a colony in which case $f(x)$ is the probability that a group consists of x individuals. Let us suppose that a group gets recorded only when at least one of the individuals in the group is sighted and each individual has an independent chance β of being sighted. Then the probability that an observed group has x individuals is

$$f^w(x) = w(x) f(x)/E(w(X)) \tag{2.6}$$

where $w(x) = [1 - (1 - \beta)^x]$. It may be noted that $E(w(X)$ is the probability of observing a group.

The limit of $f^w(x)$ as $\beta \to 0$ is easily seen to be

$$x f(x)/E(X) \tag{2.7}$$

which is thus an approximation to (2.6) when β is very small. The weight $w(x) = x$ corresponds to size biased sampling and the distribution (2.7) provides an appropriate model in many practical situations. See, for example, the analysis of data on sex ratio in Rao (1965) and Section 4 of this paper.

The significance of the weight function $w(x) = 1 - (1 - \beta)^x$ and the limiting value x would be apparent from the following examples.

Example 1 (Haldane 1938, Fisher 1934, Rao 1965, Neel and Schull 1966, etc.). If we wish to study the distribution of the number X of albino children (or children with a rare anomaly) in families with proneness to produce such children, a convenient sampling method is first to discover an albino child and through it obtain the albino count X^w of the family to which it belongs. If the probability of detecting an albino is β, then the probability that a family with x albinos gets recorded is $w(x) = 1 - (1 - \beta)^x$, assuming the usual independence of Bernoulli trials. In such a case X^w has a weighted binomial distribution with the weight function as defined above.

Example 2 (Cook and Martin 1974). In aerial census data collected for estimating wildlife population density, visibility bias is generally present because of the failure to observe some

This content downloaded from 50.78.147.85 on Tue, 11 Nov 2014 10:00:25 AM
All use subject to JSTOR Terms and Conditions

animals. Suppose that the animals are found in groups and group count X had *pdf* $f(x)$ and the probability of sighting an animal is β. Conditional on observing at least one animal in a group, a complete count is made of the group and the number of animals is recorded. If the sampling process is such that each animal has an independent chance β of being sighted, then the selection probability is $w(x) = 1 - (1 - \beta)^x$. The observed group count X^w has the *pdf* $w(x)f(x)/E(w(X))$.

In both these examples, the parameter β may be small, in which case the weight function $w(x) = x$ would be appropriate. The exact treatment of quadrat sampling for animal populations for given β is given in Section 5 of this paper.

2.4 Two-Stage Sampling (A Limit Theorem)

First let us consider a discrete r.v. X such that

$$P(X = i) = \pi_i, i = 1, \cdots . \tag{2.8}$$

Suppose that nature has produced a large sample of size N from the distribution (2.8) and in the sample the observation i occurs n_i times, so that $n_1 + n_2 + \cdots = N$. Further let us suppose that we take a subsample of size n from the finite set of N observations by drawing one observation at a time with replacement and giving a chance proportional to $n_i\, w(i)$ to the observation i. Then the probability that the subsample consists of r_1 ones, r_2 twos, $\cdots$ is proportional to

$$\frac{[n_1 w(1)]^{r_1}[n_2 w(2)]^{r_2}\cdots}{[n_1 w(1) + n_2 w(2) + \cdots]^n} = \frac{[(n_1/N)w(1)]^{r_1}[(n_2/N)w(2)]^{r_2}\cdots}{[(n_1/N)w(1) + (n_2/N)w(2) + \cdots]^n}. \tag{2.9}$$

As $N \to \infty$, the expression (2.9) tends in probability to

$$\frac{[\pi_1 w(1)]^{r_1}[\pi_2 w(2)]^{r_2}\cdots}{[\pi_1 w(1) + \pi_2 w(2) + \cdots]^n} . \tag{2.10}$$

In the limit $r_1, r_2, \cdots$ constitute a sample of size n from the weighted distribution

$$P(X^w = i) = w(i)\pi_i/\Sigma\, w(i)\pi_i. \tag{2.11}$$

It is seen that in (2.11), the weight $w(i)$ can be arbitrarily subject to the condition that a chance mechanism exists for drawing a sample from the finite set $(x_1, \cdots, x_N)$ giving a chance proportional to $w(x_i)$ for x_i. Consider, for instance, the problem of estimating the probability that a child inherits a certain defect. For this purpose we may obtain the list of children referred to a clinic and record the number of defective and non-defective children in each of the distinct families to which they belong. In such a method of sampling, the chance that a family with r defective children is brought to record is proportional to r, if the children referred to the clinic can be considered as a random sample of all defective children in the population under study.

A similar result holds in the case of a continuous *rv* X with *pdf* $f(x)$ when we subsample from an original large sample (say of size N), giving a chance proportional to $w(x)$ to observation x in the original sample. In such a case as $N \to \infty$, the subsample of fixed size n may be considered as a random sample on a *rv* X^w with the *pdf*

$$w(x)\, f(x)/E(w(X)). \tag{2.12}$$

Examples of such sampling arise if we want to determine particle size distribution by choosing a sample of particles hit by random points selected in the space enclosing the particles.

This content downloaded from 50.78.147.85 on Tue, 11 Nov 2014 10:00:25 AM
All use subject to JSTOR Terms and Conditions

2.5 Examples

It may be worthwhile now to record some of the important distributions and their modified forms under size-bias with $w(x) = x$. It may be noted that in the case of discrete distributions X^w must necessarily have $P(X^w = 0) = 0$.

3. Some Properties of Mean Values of Weighted Distributions

In this section, we prove a number of theorems on the inequalities between the mean values of two weighted distributions.

Theorem 1: Let a non-negative *rv* X have *pdf* $f(x)$ with $E(X) < \infty$. Let X^w have *pdf* $f_w(x) = xf(x)/E(X)$. Then $E(X^w) - E(X) = V(X)/E(X)$, where V stands for variance, and therefore $E(X^w) > E(X)$ for non-degenerate X.

Proof: Straightforward.

Remark: The inequality $E(X^w) > E(X)$ substantiates the intuitive result that the size-biased X^w records larger values of X more often than their natural frequency, and its smaller values less often.

Theorem 2: Let *rv*X have *pdf* $f(x)$. Further let the weight function $w(x) > 0$ have $E(w(X)) < \infty$. Let X^w be the w-weighted *rv* of X with *pdf* $f^w(x) = w(x)f(x)/E(w(X))$. Then $E(X^w) > E(X)$ if $\text{cov}[X,w(X)] > 0$ and $E(X^w) < E(X)$ if $\text{cov}\ [X,w(X)] < 0$.

TABLE 1
Certain Basic Distributions and their Size-Biased Forms

Random variable (rv)	pf(pdf)	Size-biased rv
Binomial, $B(n,p)$	$\binom{n}{x}p^x(1-p)^{n-x}$	$1 + B(n-1,p)$
Negative Binomial, $NB(k,p)$	$\binom{k+x-1}{x}q^xp^k$	$1 + NB(k+1,p)$
Poisson, $Po(\lambda)$	$e^{-\lambda}\lambda^{\xi}\lvert x \quad \$$	$1 + Po(\lambda)$
Logarithmic series, $L(\alpha)$	$\{-\log(1-\alpha)\}^{-1}\alpha^x/x$	$1 + NB(1,\alpha)$
Hypergeometric, $H(n,M,N)$	$\binom{n}{x}M^{(x)}(N-M)^{(n-x)}/N^{(n)}$	$1 + H(n-1,M-1,N-1)$
Binomial beta, $BB(n,\alpha,\gamma)$	$\binom{n}{x}\beta(\alpha+x,\gamma+n-x)/\beta(\alpha,\gamma)$	$1 + BB(n-1,\alpha,\gamma)$
Negative binomial beta, $NBB(k,\alpha,\gamma)$	$\binom{k+x-1}{x}\beta(\alpha+x,\gamma+k)/\beta(\alpha,\gamma)$	$1 + NBB(k+1,\alpha,\gamma)$
Gamma, $G(\alpha,k)$	$\alpha^kx^{k-1}e^{-\alpha x}/\Gamma(k)$	$G(\alpha,k+1)$
Beta first kind, $B_1(\delta,\gamma)$	$x^{\delta-1}(1-x^{\gamma-1}/\beta(\delta,\gamma)$	$B_1(\delta+1,\gamma)$
Beta second kind, $B_2(\delta,\gamma)$	$x^{\delta-1}(1+x)^{-\gamma}/\beta(\delta,\gamma-\delta)$	$B_2(\delta+1,\gamma-\delta-1)$
Pearson type V, $Pe(k)$	$x^{-k-1}\exp(-x^{-1})/\Gamma(k)$	$Pe(k-1)$
Pareto, $Pa(\alpha,\gamma)$	$\gamma\alpha^\gamma x^{-(\gamma+1)},\ x \geq \alpha$	$Pa(\alpha,\gamma-1)$
Lognormal, $LN(\mu,\sigma^2)$	$(2\pi\sigma_2)^{-1/2}\exp -\frac{1}{2}\left(\frac{\log x-\mu}{\sigma}\right)^2 \cdot\frac{1}{x}\cdot$	$LN(\mu+\sigma^2,\sigma^2)$

This content downloaded from 50.78.147.85 on Tue, 11 Nov 2014 10:00:25 AM
All use subject to JSTOR Terms and Conditions

Proof: Straightforward.

Theorem 3: Let X and X^w be defined as in Theorem 2. Further let X be non-negative. Then $E(X^w) > E(X)$ if $w(x) \uparrow$ in x and $E(X^w) < E(X)$ if $w(x) \downarrow$ in x, where $\uparrow$ means increasing and $\downarrow$ means decreasing.

Proof: Follows from Theorem 2 since $\text{cov}[X, w(X)] > 0$ if $w(x) \uparrow x$ and the reverse is true if $w(x) \uparrow x$, which can b easily demonstrated.

Theorem 4: Let non-negative *rv* X have *pdf* $f(x)$. Let the weight functions $w_i(x) > 0$ have $E(w_i(X)) < \infty$ for $i = 1, 2$, defining the corresponding w_i-weighted *rv*'s of X denoted by X^{w_i}. Then $E(X^{w_2}) > E(X^{w_1})$ if $r(x) = w_2(x)/w_1(x) \uparrow$ in x and $E(X^{w_2}) < E(X^{w_1})$ if $r(x) \downarrow$ in x.

Proof: Follows when one notes that X^{w_2} is r-weighted *rv* of X^{w_1}.

Remark: The result in Theorem 4 is interesting in that the ratio of the weight functions is a decisive criterion, and not any direct inequality between the weight functions, as one may think. For example, let $w_1(x) = x(x - 1)$, and $w_2(x) = x^2$, from which $r(x) = w_2(x)/w_1(x) = x/(x - 1) = 1/[1 - (1/x)] \downarrow x$, implying that $E(X^{w_2}) < E(X^{w_1})$, and not the reverse because one observes $w_2(x) > w_1(x)$.

4. *A Natural Example of Weighted Binomial Distribution*

4.1 Test of the Model

Let us ascertain from each male student in a class the number of brothers, including himself, and sisters he has and denote by k the number of students and by B and S the total numbers of brothers and sisters. What would be the approximate values of $B/(B + S)$, the ratio of brothers to the total number of children and $(B + S)/k$, the average number of children (a.n.c.) per family? It is easily seen that $B/(B + S)$ will be an overestimate of the proportion of boys among the children in the general population which is about half, and similarly $(B + S)/k$ is biased upwards as an estiamte of the a.n.c. per family in the general population. Surprisingly, when the number of boys in the class is not very small, you can make a fairly accurate prediction of the relative magnitudes of B and S, and the ratio $B/(B + S)$. This was stated in the form of an empirical theorem in Rao (1977). It was observed that $(B - k)/(B + S - k)$ is closer to half than $B/(B + S)$ (see Table 2). This was explained on the hypothesis that the distribution of b, the number of brothers given $b + s$, the total number of brothers and sisters, is a weighted binomial with weight proportional to the size of the observation, i.e., writing C for the binomial coefficient,

$$\begin{aligned} p(b \mid b + s) &= b\, C(b + s, b)(.5)^{b+s}/E(b) \\ &= C(b + s - 1, b - 1)(.5)^{b+s-1}. \end{aligned} \tag{4.1}$$

Let (b_i, s_i), $i = 1, \cdots, k$ be the observation of k boys, and $B = \Sigma b_i$, $S = \Sigma S_i$. Then under the model (4.1)

$$p(B - k = x \mid B + S - k = T) = C(T, x)(.5)^T. \tag{4.2}$$

To test the hypothesis that b has a weighted distribution of the type (4.1), we have computed

$$\chi^2 = (2x - T)^2/T \text{ on 1 d.f.} \tag{4.3}$$

from the data for each city (see Table 2). The chi-square values are small, indicating the plausibility of the hypothesis. A more detailed test of the hypothesis, i.e. the validity of the model (4.1), was carried out by computing chi-square values for each family size within each city. The chi-square values were again small.

This content downloaded from 50.78.147.85 on Tue, 11 Nov 2014 10:00:25 AM
All use subject to JSTOR Terms and Conditions

TABLE 2
Estimates of Sex Ratio and Family Size from Data on Male Respondents

Place and year		k	B	S	$\frac{B}{B+S}$	$\frac{B-k}{B+S-k}$	χ^2	$\frac{B+S}{k}$	$\hat{\delta}$
Delhi	:75	29	92	66	.58	.488	.07	5.45	4.25
Calcutta	:63	104	414	312	.57	.498	.04	6.96	5.30
Waltair	:69	39	123	88	.58	.488	1.09	5.41	4.36
Hyderabad	:74	25	72	53	.58	.470	.36	5.00	3.46
Tirupati (students)	:75	592	1902	1274	.60	.507	.50	5.36	4.24
Tirupati (staff)	:76	50	172	130	.57	.484	.25	6.04	4.20
Poona	:75	47	125	65	.66	.545	1.18	4.04	3.15
Tehran	:75	21	65	40	.62	.500	.19	5.00	3.18
Isphahan	:75	11	45	32	.58	.515	.06	7.00	5.70
Tokyo	:75	50	90	34	.73	.540	.49	2.48	2.25
Columbus	:75	29	65	52	.62	.523	2.91	4.00	2.79
State College	:75	28	80	37	.68	.584	2.53	4.18	3.21
College Station	:76	63	152	90	.63	.497	.01	3.84	3.04
London & Bradford	:76	43	80	39	.67	.487	.02	2.77	2.15

* It is interesting to note that Tokyo has the smallest family size (number of children). Among the Indian cities Poona has a smaller value, and it would be of interest to investigate this phenomenon.

Note: It has not been possible to ascertain the actual family sizes in the populations of different cities quoted in Table 2 except in the cases of Indian cities. In these cases the figures were close to δ, as predicted.

4.2 Estimation of Family Size

Let $f(b, s)$ be the relative frequency of families with b brothers and s sisters in the general population. Then under the assumption made in Section 4.1 the probability of the observation (b, s) or (b, t) where $t = b + s$ coming into our record is

$$f^w(b, s) = b\,f(b, s)/E(B) = b\,p(t)\,p(b \mid t)/E(B) \tag{4.4}$$

where

$$p(b \mid t) = C(t, b)\,\pi^b(1-\pi)^{t-b}. \tag{4.5}$$

Then

$$p^w(t) = t\,p(t)/\delta,\ \delta = E(T) \tag{4.6}$$

where δ is the parameter of interest. Hence

$$E_w(1/T) = 1/\delta \tag{4.7}$$

which shows that the harmonic mean of the observed t_i is an estimator of δ.

If the exact form of $p(t)$ is not known then we may estimate δ by

$$\bar{\delta} = k/\Sigma(1/t_i). \tag{4.8}$$

The estimate $\bar{\delta}$ of δ is computed for each city and given in Table 2. It is seen that the observed average $(B + S)/k$ is larger than the estimated value in each case.

Since $p(b \mid t)$ given in (4.5) is independent of δ, we observe that under the assumed model, t is sufficient for δ. If the form of $p(t)$ is known, we may use the likelihood function

$$\prod_{i=1}^{k} p^w(t_i) = \delta^{-k} \prod_{i=1}^{k} t_i p(t_i) \tag{4.9}$$

to estimate δ. For instance, if $p(t)$ is geometric, then the maximum likelihood estimate of δ is

This content downloaded from 50.78.147.85 on Tue, 11 Nov 2014 10:00:25 AM
All use subject to JSTOR Terms and Conditions

$$\bar{\delta}_g = \frac{T}{2K} + \frac{1}{2} \tag{4.10}$$

where $T = t_1 + \cdots + t_k$. If $p(t)$ is logarithmic, then

$$\bar{\delta}_l = |-(1 - \hat{\alpha}) \log (1 - \hat{\alpha})|^{-1}\hat{\alpha} \tag{4.11}$$

where $\hat{\alpha} = 1 - (k/T)$. The formulae (4.10) and (4.11) are not appropriate for the present data as $p(t)$ does not seem to be either geometric or logarithmic, as claimed in some studies (see Feller, p. 141).

5. *Quadrat Sampling with Visibility Bias*

For the purpose of estimating wildlife population density, quadrat sampling has been found generally preferable. Quadrat sampling is carried out by first selecting at random a number of quadrats of fixed size from the region under study and ascertaining the number of animals in each. Following Cook and Martin (1974) we make the assumptions as given below:

A_1: Animals occur in groups within each quadrat and the number of groups within a quadrat has a specified distribution.

A_2: The number of animals in a group has a specified distribution.

A_3: The number of groups within a quadrat and the numbers of animals within the groups are all independently distributed.

A_4: The method of sampling is such that the probability of sighting (recording) a group of x animals is $w(x)$.

Let X and X^w be the *rv*'s representing the number of animals in a group in the population and as ascertained. Similarly, let N and N^w be the *rv*'s for the number of groups within a quadrat. It is clear that since the method of ascertainment does not give equal chance of selection to groups of all sizes (unless $w(x)$ is constant), the *rv*'s X and X^w do not have the same distribution, and so is the case with N and N^w. The following theorem provides the basic results in quadrat sampling theory.

Theorem 5: Under the assumptions $A_1 - A_4$ we have the following results.

(*i*)
$$P(N^w = m \mid N = n) = \binom{n}{m}\omega^m(1 - \omega)^{n-m}$$

where

$$\omega = \sum_1^\infty w(x)\, P(X = x)$$

is the visibility factor (the probability of recording a group).

(ii)
$$P(N^w = m) = \sum_{n=m}^\infty \binom{n}{m} \omega^m(1 - \omega)^{n-m}\, P(N = n),$$

i.e. the visibility bias induces an additive damage model on the true quadrat frequency with binomial survival distribution (see Rao 1965). (iii) The probability that m observed groups in a quadrat have $x_1, \cdots, x_m$ animals is

$$P(X_1{}^w = x_1, \cdots, X_m{}^w = x_m \mid N^w = m) = \prod_{i=1}^m P(X^{wi} = x_i)$$

This content downloaded from 50.78.147.85 on Tue, 11 Nov 2014 10:00:25 AM
All use subject to JSTOR Terms and Conditions

where it may be noted,

$$P(X^w = x) = w(x)\, P(X = x)/\omega.$$

(iv) Let $S^w = X_1{}^w + \cdots + X_m{}^w$. Then

$$P(S^w = y) = \sum_{m=1}^{\infty} P(N^w = m)\, P(S^w = y \mid m)$$

$$P(S^w = y \mid m) = \sum_{\Sigma x_i = y} \frac{w(x_1)}{\omega} \cdots \frac{w(x_m)}{\omega} P(X_1 = x_1) \cdots P(X_m = x_m).$$

Proof: Under the assumptions and notations used we have the basic probability equality

$$P(N = n, N^w = m, X_1{}^w = x_1, \cdots, X_m{}^w = x_m, X_{m+1} = x_{m+1}, \cdots, X_n = x_n)$$

$$= P(N = n) \binom{n}{m} \prod_{j=1}^{m} P(X_j = x_j) w(x_j) \prod_{j=m+1}^{n} [1 - w(x_j)]\, P(X_j = x_j). \quad (5.1)$$

From (5.1) summing out $X_{m+1}, \cdots, X_n$ we have

$$P(N = n, N^w = m, X_1{}^w = x_1, \cdots, X_m{}^w = x_m)$$

$$= P(N = n) \binom{n}{m} \omega^m (1 - \omega)^{n-m} \prod_{j=1}^{m} P(X_j{}^w = x_j). \quad (5.2)$$

Then the results (i), (ii), and (iii) of the theorem follow from (5.2). Summing (5.2) over n from m to ∞, we have

$$P(N^w = m, X_1{}^w = x_1, \cdots, X_m{}^w = x_m) = P(N^w = m) \prod_{j=1}^{m} P(X_j{}^w = x_j), \quad (5.3)$$

from which the result (iv) follows.

Note 1: The expression (5.3) enables us to write down the joint likelihood of the numbers of groups observed in different quadrats and the numbers of animals observed in all the groups sighted. Thus, if $m_1, \cdots, m_k$ are the numbers of groups in k quadrats and x_{ij} is the number of animals in the *jth* group of the *ith* quadrat, the joint likelihood is the product of

$$\prod_{i=1}^{k} P(N^w = m_i) \quad (5.4)$$

and

$$\prod_{i=1}^{k} \prod_{j=1}^{m_i} P(X_{ij}{}^w = x_{ij}). \quad (5.5)$$

Results (ii) and (iii) of the theorem give the methods of computing the individual terms in (5.4) and (5.5) from the population distributions of N and X and the weight function $w(x)$. In general, the unknown parameters are those occurring in the specified distributions of N and X and the additional visibility factor ω (or p the probability of sighting an animal). All these could be estimated using the product of (5.4) and (5.5) as the likelihood function.

Note 2: Cook and Martin (1974) consider the special case where

$$N \sim P_0(\lambda), \text{ Poisson with parameter } \lambda, \quad (5.6)$$

$$X \sim a_x\, \theta^x / g(\theta), \text{ power series distribution}, \quad (5.7)$$

$$w(x) = 1 - (1 - \beta)^x. \quad (5.8)$$

This content downloaded from 50.78.147.85 on Tue, 11 Nov 2014 10:00:25 AM
All use subject to JSTOR Terms and Conditions

It may be noted that whatever $w(x)$ may be, $N^w \sim P_0(\delta)$, $\delta \doteq \lambda\omega$ where

$$\omega = \Sigma\, s_x\, w(x)\, \theta^x/g(\theta) \quad \text{and} \quad X^w \sim a_x\, w(x)\, \theta^x/\omega g(\theta).$$

Thus, there are three parameters δ, ω and θ. Then the parameter δ is estimated from the likelihood (5.4) and ω, θ from (5.5). Cook and Martin (1974) provided the necessary computations in such a case, choosing $w(x)$ as in (5.8).

If N is not a Poisson variable then the distribution of N^w involves ω as an additional parameter (see Rao 1965 and Sprott 1965), in which case the product of (5.4) and (5.5) provide the joint likelihood for the estimation of all the unknown parameters.

In the special case where N and X are as distributed in (5.6) and (5.7) respectively and $w(x) = \beta^x$ (i.e., when a group is observed if and only if all the animals are sighted),

$$N^w \sim P_0(\delta), \delta = \lambda\omega \quad \text{and} \quad X^w \sim a_x\, \phi^x/g(\phi), \phi = \theta\beta$$

so that the parameters λ, θ and β are confounded and are not individually estimable. In a different context, Kemp (1973, 1975) observes the confounding of θ and β in the special case of N being degenerate at one.

Résumé

Quand un chercheur recueille une observation qui suit de par sa nature un modèle stochastique, l'observation recueillie n'obéira à la distribution d'origine, que si chaque élément observable reçoit la même chance d'être recueilli. Un certain nombre d'articles sont apparus ces dix dernières années, qui utilisent implicitement les concepts de distributions pondérées avec échantillonnage d'effectif biaisé.

Dans cet article, nous examinons quelques modèles généraux conduisant à des distributions pondérées non nénessairement limitées par l'unité. Les exemples comprennent : l'échantillonnage probabiliste dans des enquêtes sur échantillons, les modèles de dommages additifs, le biais de visibilité lié à la nature du recueil des données et l'échantillonnage à deux niveaux. On rapporte plusieurs distributions importantes et leurs déformations provoquées par un biais d'effectif. On donne quelques théorèmes sur les inégalités entre valeurs moyennes de deux distributions pondérées. Les résultats sont appliqués à l'analyse de données se rapportant à des populations humaines et au contrôle des populations sauvages.

Pour des populations humaines, la question suivante est posée et discutée : demandons à chaque étudiant d'une classe, de sexe mâle, le nombre de frères (lui même y compris) et de soeurs de sa fratie. Soit h le nombre d'étudiants, B et S le nombre total de frères et de soeurs. Que seront les valeurs approchées de $B/(B + S)$, rapport du nombre de frères au nombre total denfants, et de $(B + S)/h$, nombre moyen d'enfants par famille ? On montre que $B/(B + S)$, est une surestimation de la proportion de garçons parmi les enfants de la population générale, proportion églale à 1/2 environ.Et que de même, $(B + S)/h$ est biaisé vers les valeurs élevées en tant qu'estimation du nombre d'enfants par famille dans la population générale. On propose quelques suggestions pour l'estimation de ces paramètres de population.

Enfin, en vue d'estimer la densité des populations sauvages on formule quelques résultats dans le cadre de l'échantillonnage quadratique avec biais de visibilité.

References

Cook, R. C. and Martin, F. B. (1974). A model for quadrat sampling with visibility bias. *Journal of the American Statistical Association 69*, 345–349.

Feller, W. (1968). *An Introduction to Probability Theory and its Applications.* Volume 1, John Wiley and Sons, Inc., New York.

Fisher, R. A. (1934). The effects of methods of ascertainment upon the estimation of frequencies. *Annals of Eugenics 6*, 13–25.

This content downloaded from 50.78.147.85 on Tue, 11 Nov 2014 10:00:25 AM
All use subject to JSTOR Terms and Conditions

Haldane, J. B. S. (1938). The estimation of the frequency of recessive conditions in man. *Annals of Eugenics (London) 7,* 255–262.

Kemp, C. D. (1973). An elementary ambiguity in accident theory. *Accident Analysis and Prevention,* 371–373.

Kemp, C. D. (1975). Models for visibility bias in quadrat sampling. *Proceedings of the 40th Session of the International Statistical Institute,* Warsaw.

Neel and Schull (1966). *Human Heredity.* University of Chicago Press, Chicago, 211–227.

Patil, G. P. and Ord, J. K. (1975). On size-biased sampling and related form-invariant weighted distributions. *Sankhyā, Series B 38,* 48–61.

Patil, G. P. and Rao, C. R. (1977). Weighted distributions: a survey of their applications. In *Applications of Statistics.* P. R. Krishnaiah (ed.), North Holland Publishing Company, 383–405.

Rao, C. R. (1977). A natural example of a weighted binomial distribution. *American Statistician 31,* 24–26.

Rao, C. R. (1965). On discrete distributions arising out of methods of ascertainment. In *Classical and Contagious Discrete Distributions.* G. P. Patil (ed.), Pergamon Press and Statistical Publishing Society, Calcutta, 320–332.

Rao, C. R. and Rubin, H. (1964). On a characterization of the Poisson distribution. *Sankhyā, Series A 25,* 295–298.

Sprott, D. A. (1965). Some comments on the question of identifiability of parameters raised by Rao. In *Classical and Contagious Discrete Distributions.* G. P. Patil (ed.), Statistical Publishing Society, Calcutta and Pergamon Press, 333–336.

Received December 1976; *Revised August* 1977

This content downloaded from 50.78.147.85 on Tue, 11 Nov 2014 10:00:25 AM
All use subject to JSTOR Terms and Conditions