Utah State University

DigitalCommons@USU

1975

Use of the Normal Generating Distribution for Estimating Population Survival

H. C. Romesburg

Recommended Citation

Utah State University
MERRILL-CAZIER LIBRARY

USE OF THE NORMAL GENERATING DISTRIBUTION FOR ESTIMATING POPULATION SURVIVAL

H. C. Romesburg
Utah State University

ABSTRACT

Counting the individuals in a population before and after an annual period of environmental stress allows the proportion s of the initial population surviving the period to be computed. A series of such observations over n annual periods gives a sequence $s_1, s_2, \ldots, s_n$. A statistical model is formulated from axioms describing the survival process, and it is concluded that these observed values may usefully be regarded as realizations of a random variable that arises from the normal generated distribution ($n.g.d.$). Equations for estimating the $n.g.d.$ parameters ζ and τ^2 from observed survival proportions by the method of moments and maximum likelihood are given. The distributions of parameter estimates $\hat{\zeta}$ and $\hat{\tau}^2$ are obtained and discussed in the context of testing hypotheses comparing survival among different populations. Finally, the dependence of the $n.g.d.$ upon parameters ζ and τ^2 is examined in terms of altering survival, either by population self-regulation mechanisms or man-induced controls. The intent is to provide insight into the relationship between the $n.g.d.$ and its supporting axioms and, more generally, basic knowledge of population processes.

INTRODUCTION

Consider a population of animals or plants over a period of time within the year, in which natural mortality is the only means of changing the number of individuals. The proportion of the population surviving the period is $s = N_2/N_1$, where N_1 is the number alive at the beginning of the period, N_2 are alive at the end of the period. Observation over n years gives a sequence $s_1, s_2, \ldots, s_n$ which, apart from sampling error, may differ. A natural question is whether the observed values can be usefully regarded as realizations of a random variable s that arises from a probability density function $q(s)$. This paper develops a model, based upon the normal distribution, which leads to the conclusion that such observations can often be adequately described by the normal generated distribution ($n.g.d.$) $q(s)$:

$$q(s) = \varphi\left[(1/\tau)(\Phi^{-1}(s) - \zeta)\right]/\tau\,\varphi\,(\Phi^{-1}(s))$$

$$0 < s < 1, -\infty < \zeta < \infty, 0 < \tau^2 < \infty \quad (1)$$

where ζ and τ^2 are parameters, $\varphi\,(\cdot)$ is probability density function of the standard normal distribution and $\Phi\,(\cdot)$ the standard normal cumulative distribution function. The $n.g.d.$ was first derived by Chiu (1974).

The development is as follows. A set of axioms are given leading to Equation 1 for $q(s)$. The axioms are statements of the factors that characterize the population and its environment and which are assumed to determine survival. The discussion of Equation 1 for $q(s)$ centers on three topics. First, equations for estimating the parameters ζ and τ^2 using observed survival proportions both by the method of moments and by maximum likelihood are given. Second, testing hypotheses concerning the true parameters ζ and τ^2 are discussed. Third, the dependence of $q(s)$ upon parameters ζ and τ^2 is examined in terms of altering survival, either by population self-regulation mechanisms or man-induced controls. One practical use of these results is that fall-to-spring survival, which is often quite uncertain, can be estimated for a number of different years, these observations used to estimate the parameters and, finally, the $n.g.d.$ $q(s)$ used to make probability statements concerning future survivals.

DERIVATION OF NORMAL GENERATED DISTRIBUTION $q(s)$

Consider a period of time from t_1 to t_2 (such as fall to spring) where mortality is the only means of change in population size, and assume:

Axiom 1: An individual is subject to a stress u by its environment. The environment is defined as everything exogenous to an individual and includes predators, competing individuals and weather.

Axiom 2: An individual is characterized by a strength v for resisting an environmental stress u.

Axiom 3: If $u > v$ the individual dies; otherwise survives.

Stress u and strength v are quantitative variables, but may not be directly measurable. We can only measure aspects of component factors such as, for example, air chill and snow pack conditions, which Verme (1968) found positively correlated with winter deer mortality in northern Michigan. What is being postulated by introducing the concepts of stress and strength is an underlying dimension of variation where positions on the dimension can be associated with an interval scale of measurement. Used in this sense, there is a continuum of values for the variable "stress u" which summarizes all of the environmental information relating to mortality. An organism could then, in theory, be subject to increasing values of stress u, and the value above which it could not survive corresponds to its strength v.

The next assumption makes allowance for differing strengths v and stresses u within the population and environment.

Axiom 4: u and v are distributed in a bivariate normal distribution with means μ_u and μ_v, variances σ_u^2 and σ_v^2 and correlation coefficient ρ.

Both u and v are taken as the sum of a great many fundamental but stochastic effects and therefore, under certain conditions, are reasonably asserted as being normally distributed. The conditions for this holding can, to a degree,

be forced by properly defining the population. For example, consider a group of young individuals and a second group of old individuals. If the strength distribution in each group is normally distributed but not identical, a "population" formed from the merger of both groups could not have its overall strength distribution normally distributed. A population should be defined so that no known factors exist which affect an individual's ability to survive. That is, a collection of organisms should be stratified on the basis of age-, density- and location-related factors which can affect survival. Any remaining differences in individual survival within a strata can only be accounted for by stochastic effects, and each strata can then be treated as a separate population. For example, an appropriate population might be juvenile deer living in an area where habitat and climate are not predictably location-dependent and where total herd size is not large enough to affect survival strongly.

Letting x be the difference of v and u, the distribution of x is univariate normal with mean:

$$\mu_x = \mu_v - \mu_u \qquad (2)$$

and variance:

$$\sigma_x^2 = \sigma_v^2 + \sigma_u^2 - 2\rho\,\sigma_v\,\sigma_u \qquad (3)$$

where ρ is the correlation between u and v. The random variable x is termed the "extensity" of the survival process.

An individual survives the period when $x > 0$, so that the fraction of the population surviving is obtained from the integral of the normal probability density $N(x;\mu_x,\sigma_x^2)$:

$$s = \int_0^\infty N(x;\mu_x,\sigma_x^2)\,dx \qquad (4)$$

Transforming x to standard form by letting $z = (x - \mu_x)/\sigma_x$ gives the final form for the single-period model:

$$s = \int_{-\infty}^{\mu_x/\sigma_x} \varphi(z)\,dz = \Phi(\mu_x/\sigma_x) \qquad (5)$$

This model appears in reliability and psychometric theory. In reliability, v represents the strength of a mechanical or electronic part and u the stress acting on the part; s is the probability a part will fail in service (Shooman 1968, pp. 441-452). The model is also identical to the well-known Thurstone (1959, pp. 19-38) judgmental model where a stimulus u interacts with a subject's ability to discriminate the stimulus, or v. The variable s is the probability a randomly chosen subject will be able to differentiate a randomly selected stimulus. In both cases the models are derived on the basis of the normal distribution.

The single-period model (Equation 5) holds for any one season with extensity parameters μ_x and σ_x^2 constant, but there is no reason to suppose these parameters are constant among seasons. For simplicity, only variation in μ_x is assumed. The parameter μ_x is assumed to follow a normal

distribution, and the argument is essentially a repeat of that presented previously in support of Axiom 4; an appeal to the central limit theorem.

Axiom 5: The difference between mean strength μ_v and mean stress μ_u, or μ_x, is normally and independently distributed with mean μ_1 and variance σ_1^2, designated by $g(\mu_x)$.

The final form of the multiperiod model results from transforming $g(\mu_x)$ using Equation 5 in the general form $s = h(\mu_x)$. For fixed σ_x^2, the marginal density for survival s among years, $q(s)$, is obtained from the transformation relation:

$$q(s) = g[h^{-1}(s)]dh^{-1}(s)/ds \qquad (6)$$

where $h^{-1}(s)$ is the inverse form of Equation 5, giving μ_x as a function of s, and is found to be

$$q(s) = (\sigma_x/\sigma_1)\exp(\mu_x^2/2\sigma_x^2 - (\mu_x - \mu_1)^2/2\sigma_1^2) \qquad (7)$$

Equation 7 can be algebraically manipulated to arrive at the n.g.d. Equation 1 if new parameters $\zeta = (\mu_1/\sigma_x)$ and $\tau^2 = (\sigma_1/\sigma_x)^2$ are used.

A straightforward integration of Equation 1 gives the cumulative distribution function $Q(s)$:

$$Q(s) = \Phi\left(\tfrac{1}{\tau}(\Phi^{-1}(s) - \zeta)\right) \qquad (8)$$

For particular values of s, $q(s)$ and $Q(s)$ are easily obtained using tables or appropriate computer functions: $\wp(s)$ can be evaluated using the error function and $\Phi(s)$ using the standard normal cumulative distribution function.

PARAMETER ESTIMATION

Chiu (1974) has derived the mean $E(s)$ and variance $Var(s)$ of $q(s)$. Defining $c = \zeta/\sqrt{(1+\tau^2)}$ and $w = \tau^2/(1+\tau^2)$, the expressions are

$$E(s) = \Phi(c) \qquad (9)$$

$$Var(s) = Pr\{y \leqslant c, z \leqslant c; w\} - E^2(s), \qquad (10)$$

where y and z have a joint bivariate normal distribution with zero means, unit variances and correlation of w. The procedure for obtaining moment estimates $\hat{\zeta}$ and $\hat{\tau}^2$ of parameters ζ and τ^2 is to compute the first two sample moments using the observations $s_1, s_2, \ldots, s_n$ and equate these to $\Phi(c)$ and $Pr\{y \leqslant c, z \leqslant c; w\}$, the first two theoretical moments. Values of $\hat{\zeta}$ and $\hat{\tau}^2$ are then searched for according to an iterative scheme until a sufficiently accurate solution is obtained. Tables given by Owen (1962) can be used for evaluating the bivariate normal distribution.

Maximum likelihood estimates of ζ and τ^2 can be obtained by forming the likelihood function $q^n(s)$ from Equation 1. The likelihood method yields estimates:

$$\hat{\zeta} = \sum_{i=1}^{n} \Phi^{-1}(s_i)/n \qquad (11)$$

$$\hat{\tau}^2 = \sum_{i=1}^{n} (\Phi^{-1}(s_i) - \hat{\zeta})^2/n \qquad (12)$$

A check for bias in these estimators shows that $\hat{\zeta}$ is unbiased while $\hat{\tau}^2$ can be made unbiased by multiplying Equation 12 by the quantity $n/(n-1)$.

HYPOTHESIS TESTING CONCERNING n.g.d. q(s)

In certain instances it may be required to test hypotheses concerning $q(s)$. For example, to test H_0: $q(s) = q'(s)$, where $q'(s)$ is a specified form of $q(s)$, against an appropriate alternative, amounts to testing H_0: $\zeta = \zeta'$ and H_0: $\tau^2 = \tau^{2'}$ against the chosen alternatives. Such tests can be performed by observing that $\hat{\zeta}$ given by Equation 11 is normally distributed while $(n-1)\hat{\tau}^2/\tau^2$ is chi-square distributed with $n-1$ d.f. The estimate $\hat{\tau}^2$ is calculated from Equation 12 and corrected for bias. The observations $s_1, s_2, \ldots, s_n$ must, of course, be independent for these distributions to hold.

The distributions of $\hat{\zeta}$ and $\hat{\tau}^2$ follow from Equation 5 which can be written:

$$\mu_x = \sigma_x \Phi^{-1}(s) \qquad (13)$$

Since μ_x is by definition normally distributed, Equation 13 implies the same for $\Phi^{-1}(s)$. Therefore, from Equations 11 and 12, $\hat{\zeta}$ and the unbiased form of $\hat{\tau}^2$ are normally and chi-square distributed, respectively.

MODEL PARAMETER DEPENDENCE

Measures of mean survival and the probability of survival less than a specified level s^* are of interest to population control. The dependence of each upon the model parameters μ_1, σ_1 and σ_x is examined in the following.

Using $\zeta = (\mu_1/\sigma_x)$ and $\tau^2 = (\sigma_1/\sigma_x)^2$, the mean survival $E(s)$ given by Equation 9 can be reformulated as

$$E(s) = \Phi[\mu_1/\sqrt{(\sigma_x^2 + \sigma_1^2)}] \qquad (14)$$

It follows from the sigmoid relation $\Phi(\cdot)$ that, for a given σ_x^2 and σ_1^2, $E(s)$ increases with mean extensity μ_1, the greatest rate of increase occurring where $\mu_1 = 0.0$. Conversely, holding μ_1 constant and varying σ_x^2 or σ_1^2 changes $E(s)$ according to the value assumed by μ_1. If $\mu_1 < 0$, $E(s)$ increases with increasing σ_x^2 or σ_1^2; if $\mu_1 = 0$, $E(s)$ does not change with σ_x^2 or σ_1^2; if $\mu_1 > 0$, $E(s)$ decreases with increasing σ_x^2 or σ_1^2.

A second measure is the probability a realized value for survival less than a specified level s^* occurs. The concern is that a low value of s in a given year reduces the population size to a point where its resource value vanishes or, at the extreme, becomes nonsustaining. The condition imposed is

$$Q(s^*) = \Phi[(1/\tau)(\Phi^{-1}(s^*) - \zeta)] < p \qquad (15)$$

where $Q(s^*)$, as given by Equation 8, is restricted to being less than a specified value p. The objective is to determine the parameter space satisfying Equation 15. With some algebra this translates to:

$$\Phi^{-1}(s^*)\sigma_x - \Phi^{-1}(p)\sigma_1 \leqslant \mu_1 \qquad (16)$$

Of general interest to extinction are small values of s^* and p such that $s^*, p < 0.5$. This implies $\Phi^{-1}(s^*)$, $\Phi^{-1}(p) < 0$. With this understanding, if Equation 16 is satisfied for any set of values σ_x, σ_1, μ_1, it will remain satisfied if σ_x and/or μ_1 are made arbitrarily large or σ_1 is reduced to zero.

Habitat manipulation which raises the mean strength μ_v by making food more abundant, or decreases the mean stress μ_u by creating an improved shelter complex, leads to increased mean extensity μ_1. From the above it follows that increasing μ_1 increases the mean survival $E(s)$ and reduces the probability of extinction $Q(s^*)$. Habitat control, however, would likely alter σ_x^2, and it is conceivable, although unconfirmed, that manipulation might raise μ_1 but alter σ_x^2 so that survival s actually decreases. Ideally, the strategy for a maximal increase in $E(s)$, in addition to increasing μ_1, is to increase σ_x^2 if $\mu_1 < 0$ and decrease σ_x^2 if $\mu_1 > 0$. Whether μ_1 is positive or negative can be decided by estimating ζ from data using Equation 11 and testing H_0: $\zeta = 0$ against the appropriate alternative.

It does not follow, however, that increased mean survival $E(s)$ implies a reduced probability of extinction. If $\mu_1 > 0$ and σ_x^2 is decreased, then $E(s)$ will increase but $Q(s^*)$ will also increase. This can be explained by reference to Figure 1, curves 7 and 9. If $\mu_1 > 0$, then $\zeta > 0$, and if $\tau^2 < 1$, then decreasing σ_x^2 will at some point cause $\tau^2 > 1$. This process results in an increase not only in $E(s)$ but in $Var(s)$ as well, and an increase in the probability contained in the lower tail of the $q(s)$ curve.

Equation 16 makes explicit the relationship between climatic variation among years and variation in extensity exhibited by the organism in its environment, σ_1^2 and σ_x^2 respectively. The two are opposed; when great variation in climatic stress exists among years (σ_1^2 large), the probability of extinction is minimized by making σ_x^2 as large as possible by habitat manipulation. Also, the organisms' behavior may be relevant in changing σ_x^2 through the dependence of σ_x^2 upon the strength-stress correlation ρ (Equation 3). If strong animals occupy low-stress microhabitats and weak animals occupy a high-stress area, $\rho < 0$ and σ_x^2 is increased compared to that for random habitat occupancy. It follows that the optimal behavioral strategy to minimize extinction over the stress period is for nature to prescribe $\rho < 0$.

CONSEQUENCES OF NONNORMALITY

The assumption of normality appears in Axiom 5 where extensity μ_x is taken as normally distributed as $g(\mu_x)$. The transforming function, of general form $s = h(\mu_x)$, is the cumulative normal distribution given by Equation 5. The

sigmoid curve $s = h(\mu_x)$, in Figure 2 is purposely skewed to represent a nonnormally distributed extensity, and $g(\mu_x)$ is also skewed. The transformation (Equation 6) can be evaluated graphically by drawing $q(s)$ so that any bounded area under the $q(s)$ curve equals the corresponding area under the $g(\mu_x)$ curve, as indicated by the shading in Figure 2. Under conditions where μ_1 is varied and where $s = h(\mu_x)$ remains sigmoid and $g(\mu_x)$ bell-shaped, although both are quite skewed, the transformed curve can be adequately fit by $q(s)$ (Equation 1); the resulting shapes are well characterized by those shown in Figure 1. Thus, exact normality is not important. However, severe nonnormality which departs from being bell-shaped, such as multimodal distributions, yields striking cases which are ill fit by the n.g.d. $q(s)$.

EXAMPLES OF ESTIMATED $q(s)$ CURVES

Figure 1 gives typical $q(s)$ curves for a partition of the ζ and τ^2 parameter space. In practice, $\hat{\zeta}$ and $\hat{\tau}^2$ can be calculated from the observations $s_1, s_2, \ldots, s_n$ using Equations 11 and 12 and Figure 1, used to identify the general form of the survival density. The specific form can be obtained from Equation 1 and probability statements concerning survival from Equation 8.

Nine sets of population survival data taken from the literature were used to generate $q(s)$ curves: 1) grouse brood survival, Connecticut Hill study area (Darrow 1947a, p. 315); 2) grouse brood survival, Adirondack study area (Darrow 1947a, p. 315); 3) adult grouse over-winter survival, Connecticut Hill study area (Darrow 1947b, p. 531); 4) adult grouse over-winter survival, Adirondack study area (Darrow 1947b, p. 531; 5) juvenile partridge over-winter survival (Severtzoff 1934, p. 419); 6) great tit summer-to-spring survival (Lack 1968, pp. 60-61); 7) bobwhite over-winter survival (Errington 1945, p. 13); 8) bobwhite over-winter survival (Kozicky and Hendrickson 1952, p. 484); 9) yearly survival of young roe deer (Severtzoff 1934, p. 422). Page references give location of tabular data as shown in Figure 3 and Table 1; none is bimodal as in curves 3, 6 and 9 in Figure 1. Bimodal curves result when the variance of μ_x among years exceeds the variance of the extensity, that is $(\sigma_1/\sigma_x)^2 = \tau^2 > 1$. Referring to the graphical transformation (Figure 2) for given μ_1 and σ_x^2, the larger σ_1^2 becomes, the greater is the spread of realizations of μ_x along the abscissa and hence the

probability of s being near 0 or 1 increases. The effect is most conspicuous when the mean of $g(\mu_x)$, or μ_1, is zero, so that $q(s)$ is symmetric. If $\tau^2 = 1$, $q(s)$ is the uniform distribution; if $\tau^2 < 1$, $q(s)$ is unimodal with mode at $s = 0.5$; if $\tau^2 > 1$, there are modes at $s = 0$ and 1 and an antimode at $s = 0.5$ (Figure 1, cases 5, 4, 6).

Thus, as σ_1^2 becomes larger than σ_x^2, $q(s)$ becomes bimodal and there is an increase in the probability of low (and high) survival. Hence bimodal forms may be rare or nonexistent in nature. This behavior can also explain the end of range for a population of animals or plants. For example, the terminus of a forest occurs where σ_1^2 begins to exceed σ_x^2 even though μ_1 remains constant as the boundary is approached.

SUMMARY

Given a set of independent survival data $s_1, s_2, \ldots, s_n$ for a given population obtained over n years, the normal generated distribution (n.g.d.; Equation 1) developed by Chiu (1974) can be used to obtain the distribution of survival proportions. Maximum likelihood estimates of the parameters ζ and τ^2 can be found from Equations 11 and 12 and the cumulative distribution function $Q(s)$ (Equation 8) used to give probability estimates of survival. Hypothesis tests concerning the distribution of survival can be performed based upon the distributions of the parameter estimates $\hat{\zeta}$ and $\hat{\tau}^2$.

The axioms leading to the n.g.d. depend upon assumptions of normality regarding the abstract concept of extensity. These cannot be subjected to empirical study. However, the model is robust for departures from normality that retain the bell-shaped feature. In any application, the researcher has knowledge of the survival process beyond the observations $s_1, s_2, \ldots, s_n$. Comparison of this knowledge with probability statements derived from the model is the most practical way to judge model validity.

The relationships among the parameters μ_1, σ_1^2 and σ_x^2 concisely describe how man- or self-regulating controls affect survival. These relationships hold in an idealized setting given by the model axioms. Still, they raise useful questions regarding the concepts of habitat management and population self-regulation strategies.
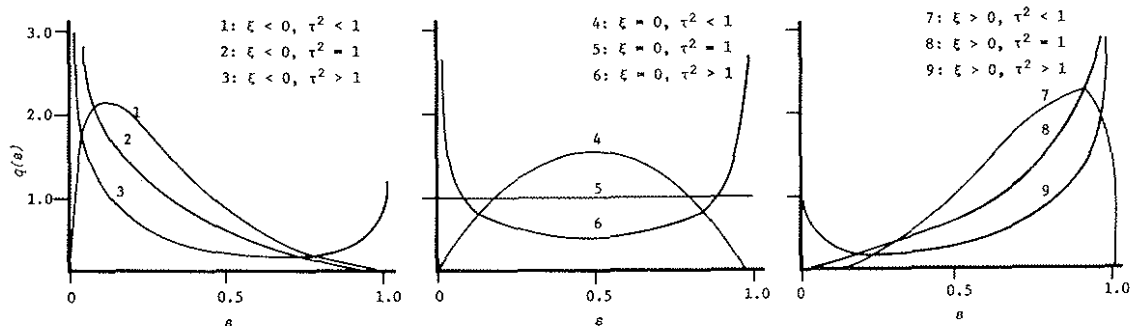

Figure 1. Typical probability densities of survival proportions, $q(s)$ by the normal generated distribution (Equation 1) for parameter space ranges: $\zeta <, =, > 0$; $\tau^2 <, =, > 1$.
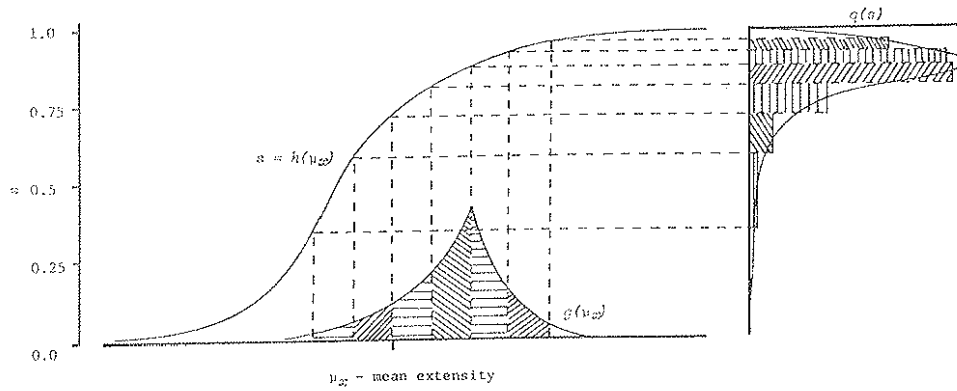
Figure 2. Graphical method for transforming $g(\mu_x)$ using $s = h(\mu_x)$ to obtain $q(s)$. The transformation preserves corresponding shaded areas under the $g(\mu_x)$ and $q(s)$ curves. $g(\mu_x)$ and $s = h(\mu_x)$ are not normal, but $q(s)$ is still approximately $n.g.d.$ distributed.
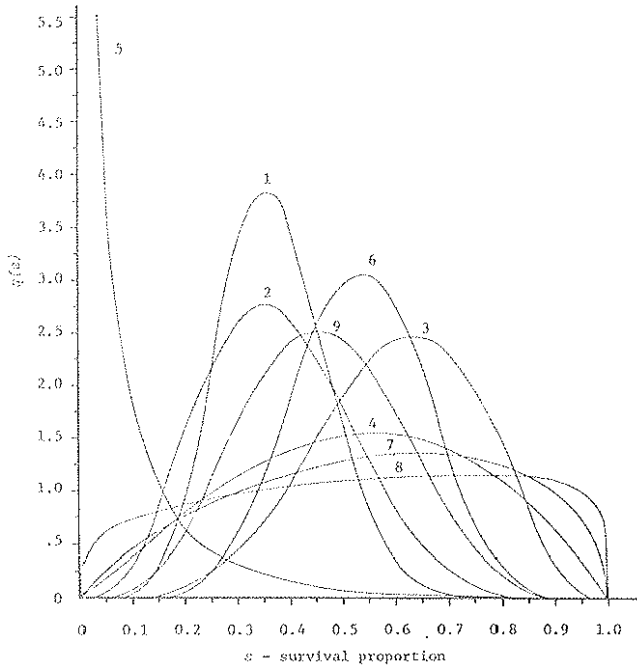

Figure 3. Normal generated distribution $q(s)$ fitted to survival realization $s_1, s_2, \ldots, s_n$ taken from the literature. Case numbers identify curves; references to data used are given in text.

Table 1. Estimated parameters $\hat{\zeta}$ and $\hat{\tau}^2$, mean $E(s)$ and standard deviation S.D. (s) and number of observations n for $q(s)$ curves given in Figure 3

Case	$\hat{\zeta}$	$\hat{\tau}^2$	$E(s)$	S.D. (s)	n
1	− .351	.0772	.368	.101	13
2	− .335	.150	.377	.138	11
3	.277	.177	.601	.151	12
4	.0858	.417	.529	.218	7
5	−2.03	.805	.0656	.107	23
6	.081	.106	.531	.123	11
7	.159	.565	.551	.241	15
8	.131	.799	.539	.270	11
9	− .103	.158	.462	.147	12

LITERATURE CITED

Chiu, W. K. 1974. A new prior distribution for attributes sampling. Technometrics 16:93-102.

Darrow, R. W. 1947a. Predation. Pages 307-352 in G. Bump, R. W. Darrow, F. C. Edminster, and W. F. Crissey, contributors. The ruffed grouse. New York State Conservation Dept.

Darrow, R. W. 1947b. Productivity of grouse population. Pages 511-554 in G. Bump, R. W. Darrow, F. C. Edminster, and W. F. Crissey, contributors. The ruffed grouse. New York State Conservation Dept.

Errington, P. L. 1945. Some contributions of a fifteen-year local study of the northern bobwhite to a knowledge of population phenomena. Ecol. Monogr. 15:1-34.

Kozicky, E. L., and Hendrickson, G. O. 1952. Fluctuations in bob-white populations. Iowa St. Coll. J. Sci. 26:483-489.

Lack, D. 1968. Population studies of birds. Oxford Univ. Press, London.

Owen, D. B. 1962. Handbook of statistical tables Addison-Wesley, London.

Severtzoff, S. A. 1934. On the dynamics of populations of vertebrates. Quart. Rev. Biol. 9(4):409-437.

Shooman, M. L. 1968. Probabilistic reliability: an engineering approach. McGraw-Hill, New York.

Thurstone, L. L. 1959. The measurement of values. Univ. Chicago Press, Chicago.

Verme, L. J. 1968. An index of winter weather severity for northern deer. J. Wildl. Manage. 32(3):566-574.