INSTITUTE OF ECOLOGY-POLISH ACADEMY OF SCIENCES

EKOLOGIA POLSKA

Vol. XVIII Warszawa 1970 No. 40

Institute of Ecology, Interdepartmental Laboratory of Statistics, Warszawa

Head: Dr. Teresa Wierzbowska

Teresa WIERZBOWSKA

TRUNCATED DISTRIBUTIONS AND THEIR APPLICATION IN ECOLOGY

(Ekol. Pol. 18: 837–848). The relations between the parameters of "total" distribution (defined for all values of the random variable) and moments of truncated distribution (this is the distribution of the analysed random variable defined in a certain range of values of this variable) for the exponential and geometrical distributions are described in this study. Examples are given of assessment of the longevity of individuals in a whole population on this basis of information on the longevity of defined age groups, and assessment of the longevity of a selected group on the basis of information concerning the whole population. Using the geometrical truncated distribution an example is given of assessment of the trapping probability of an individual and assessment in the numbers of the given population for a short period.

When investigating biological populations by means of statistical methods we often have to do with the distribution of a random variable (by means of which the given biological phenomenon is described) defined in a certain range of the values of this variable. This is the so-called truncated distribution. The characters of this distribution, such as expected value and variance, depend on the parameters of the "total" distribution (defined for all values of the random variable) by means of which the given phenomenon is described. These parameters, characterizing the whole biological population, are estimated on the basis of a sample representing a given group of this population. The reverse situation may of course exist. Knowing the parameters of the distribution of the whole population we may wish to define the numerical characteristics (mean and variance) of part of this population. A sample representing the whole population then provides us with estimates of the unknown parameters on which the characteristics of the truncated distribution depend.

We encounter situations of the first type far more often than those of the second type. The period, interesting from the ecological aspect, during which the given observations are carried out, is frequently too short to permit of obtaining the total empirical distribution of the analysed random variable. Consequently if the connection is given between the moments of the truncated distribution and the parameters on which the moments of total distribution depend, this makes it possible to solve the types of problems presented above. In the discussion which follows we have limited ourselves to the exponential and geometrical distributions commonly used in statistical elaboration of many ecological questions. Those interested in normal distribution will find connections between the first moments of total and truncated distributions in the study by Cramer (1958).

In this study the values of the mean and variance of the distributions analysed have been replaced by the corresponding values from the sample. In order to indicate that the corresponding mean or variance had been calculated for the value of the variable from a to b, indexes a and b were placed by these values. When $a = 0$, $b = \infty$, the corresponding means and variances are moments of total distribution and we give them without indexes.

EXPONENTIAL DISTRIBUTION

The random variable with an exponential distribution has the following density of forms:

$$f(t) = \begin{cases} \mu\, e^{-\mu t} & \text{for } t \geqslant 0 \\ 0 & \text{for the remaining values } t \end{cases} \tag{1}$$

Mean and variance are respectively:

$$\bar{t} = \frac{1}{\mu}$$

$$S^2(t) = \frac{1}{\mu^2} \tag{2}$$

Random variable for $a \leqslant t \leqslant b$ has the density

$$f_1(t) = \frac{\mu\, e^{-\mu t}}{e^{-a\mu} - e^{-b\mu}} \tag{3}$$

For the remaining values t, $f_1(t) = 0$.
Mean and variance are respectively:

$$\begin{gathered} \bar{t}_{a,b} = \frac{1}{\mu} + a - \frac{(b-a)\, e^{-\mu(b-a)}}{1 - e^{-\mu(b-a)}} \\ S^2_{a,b} = \frac{1}{\mu^2} + (b-a)^2 - \frac{(b-a)^2}{1 - e^{-\mu(b-a)}} - \left[\frac{(b-a)\, e^{-\mu(b-a)}}{1 - e^{-\mu(b-a)}}\right]^2 \end{gathered} \tag{4}$$

The possibility of using the exponential truncated distribution can be illustrated by the following example: let us take the average longevity of individuals living under given ecological conditions as 10 months. We are interested in determining the average age at the time of the death of those individuals, which died between the second and eighth month of their lives. If we can assume that the distribution of longevity is exponential, with the density given by equation (1), then the distribution of longevity for the analysed group of rodents has a density defined by equation (3) and the required mean is defined by equation (4), in both equations $a = 2$ and $b = 8$.

In order to facilitate the solution of the last equation we use the relation presented in Figure 1 between the value of parameter μ and value $(\bar{t}_{a,b} - a) = g_1(k)$, where $\bar{t}_{a,b}$ is defined by means of equation (4), in which $(b - a) = k$. This relation was given for $k = 2, 3 \ldots 10$. For the example given above we have: $a = 2$, $b = 8$, $k = 6$, $\bar{t} = 10$, and thus in agreement with (2) $\mu = 0.1$.

For the point with co-ordinates $(\mu, \bar{t}_{a,b} - a) = (0.1, \bar{t}_{a,b} - a)$ located on curve $g_1(6)$ which is marked with the figure 6 in Figure 1, we look for the value of co-ordinate $(\bar{t}_{a,b} - a)$. The required value is the value situated on the ordinate axis, corresponding to value $\mu = 0.1$ from the abscissa axis, and thus equal to 2.7. Consequently $\bar{t}_{2,8} - 2 = 2.7$, hence $\bar{t}_{2,8} = 4.7$. We therefore obtained the result that the average age at the time of death of those rodents which died between the second and eighth month of life was 4.7 months.

In order to solve the reverse problem, namely from a knowledge of the average longevity of a given group of individuals, we want to find the average longevity of the whole population. Value $\bar{t}_{a,b}$ is given, and we wish to find the

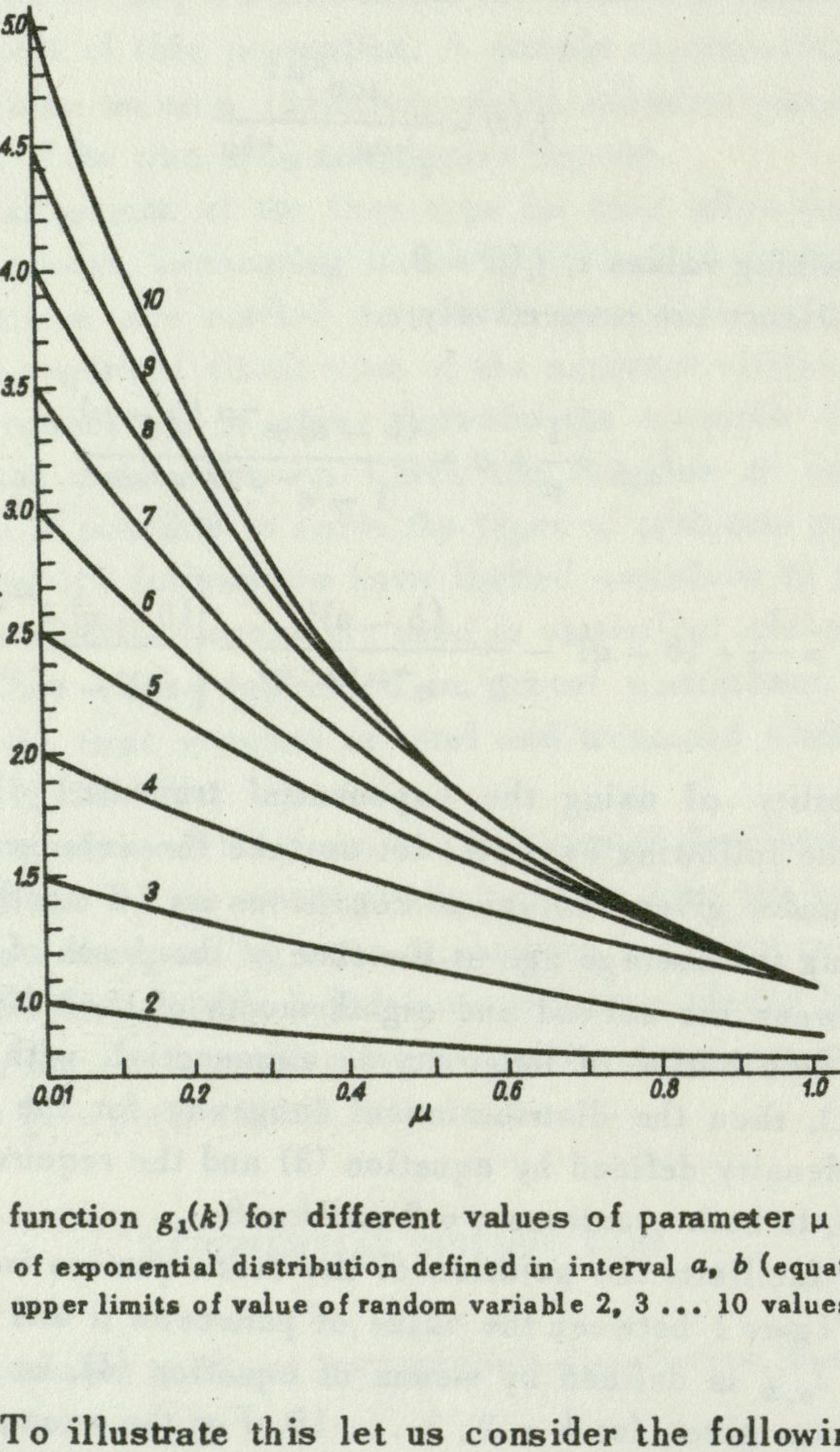

Fig. 1. Values of function $g_1(k)$ for different values of parameter μ $g_1(k) = \bar{t}_{a,b} - a$ $\bar{t}_{a,b}$ – expected value of exponential distribution defined in interval a, b (equation 4) a, b – lower and upper limits of value of random variable 2, 3 ... 10 values k

average value $\bar{t}$. To illustrate this let us consider the following example: the average length of life in the experimental area of those rodents which did not live longer than 5 months was 2.2 months. What is the average length of life of the whole population?

In order to calculate this mean value we need the value of parameter, which we calculated from equation (4). We make use as previously of the relations given in Figure 1.

For the above example $a = 0$, $b = 5$, $k = 5$, $\bar{t}_{0,5} = 2.2$.

Value $\mu = 0.15$ corresponds on the abscissa axis to value $\bar{t}_{0.5} - a = 2.2$, which we find on the ordinate axis. Finally, in accordance with equation (2), the average length of life is 6.7 months.

GEOMETRICAL DISTRIBUTION

Let us assume the random variable X has a geometrical distribution with probability function

$$P\,(X = k) = p \cdot g^{k-1} \qquad k = 1, 2, 3 \ldots \tag{5}$$

mean

$$\bar{x} = \frac{1}{p}$$

and variance

$$S^2 = \frac{g}{p^2} \quad \text{where } g = 1 - p \tag{6}$$

The distribution of variable X defined for value k from the range $a \leqslant k \leqslant b$ takes the following form

$$P_1\,(X = k) = \frac{p}{g^a\left(1 - g^{b-a+1}\right)} \cdot g^k \tag{7}$$

with mean

$$\bar{x}_{a,b} = \frac{g}{p} + a - \frac{(b - a + 1)\, g^{b-a+1}}{1 - g^{b-a+1}} \tag{7a}$$

and variance

$$S^2_{a,b} = \frac{g}{p^2} + (b - a + 1)^2 - \frac{(b - a + 1)^2}{1 - g^{b-a+1}} - \left[\frac{(b - a + 1)\, g^{b-a+1}}{1 - g^{b-a+1}}\right]^2 \tag{7b}$$

The geometrical distribution can be used in the solution of many ecological problems. In the present study we have limited ourselves to giving two examples of the application of this distribution, or more exactly of the application of the truncated geometrical distribution.

The first distribution refers to the trapping probability of marked individuals which belong to the population studied. If trapping probability (p) were constant, then after being caught in a trap situated in the area the given individual will be caught again on the following trapping day (after the expiry of the trapping unit) with probability p, caught again on the second trapping day (after expiry of two trapping units), with probability pg. In all, after the expiry of k trapping units with probability pg^{k-1}.

Therefore the number of trapping units (days, weeks) which elapse from the the given catch to the next or the interval of time between two successive captures has the geometrical distribution given in equation (5). The average number of days (weeks) elapse from the capture analysed to the next is then equal to the reciprocal of trapping probability (equation 6).

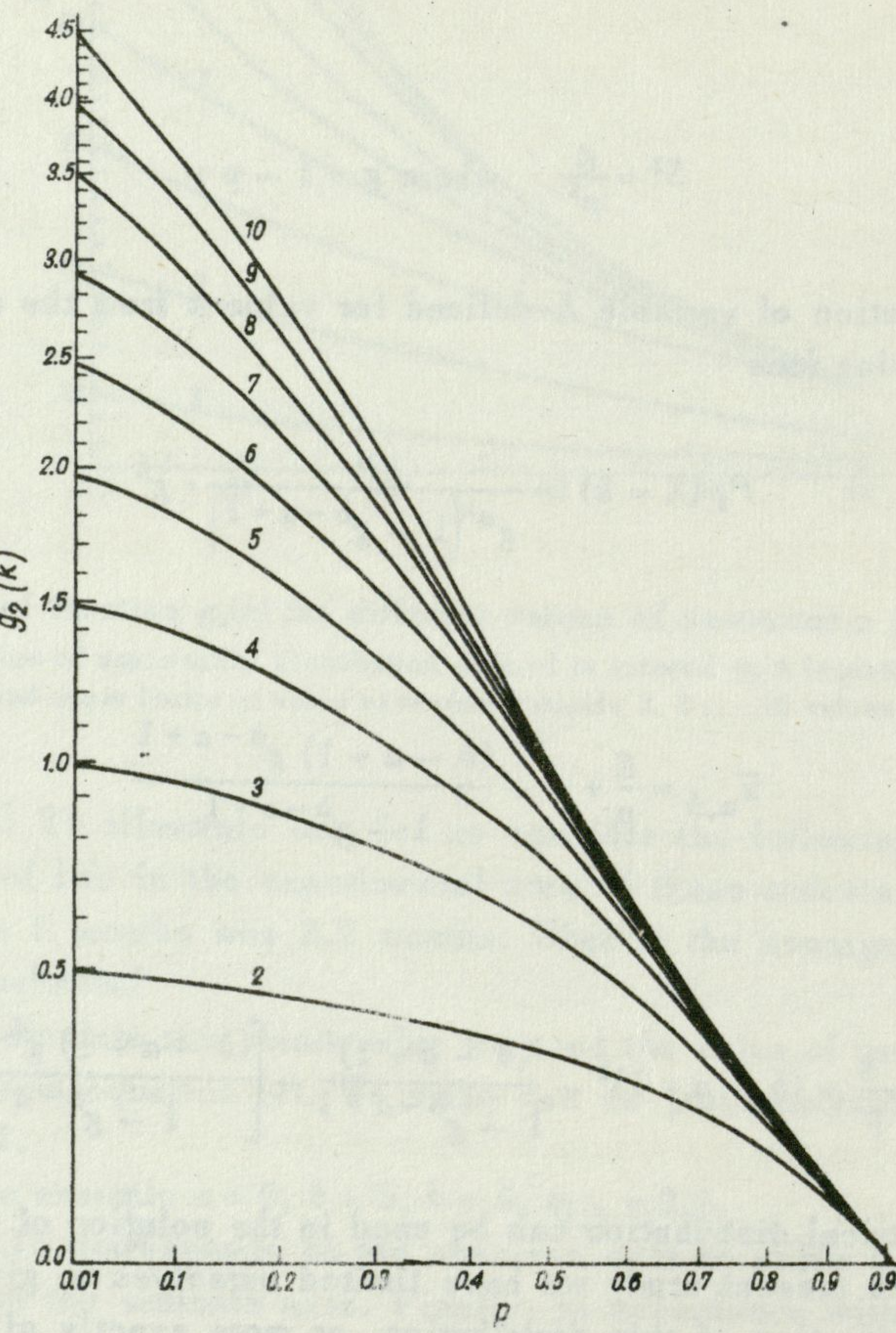

Fig. 2. Values of function $g_2(k) = \bar{x}_{a,b} - a$ for different values of parameter p
$\bar{x}_{a,b}$ — expected value of geometrical distribution defined in interval a, b (equation 7a), a, b — lower and upper limits of value of random variable, 2, 3 ... 10 values k

If, however, the period for which we wish to define trapping probability p is so short that not all individuals caught during this period "manage" to be caught in it again, then the reciprocal of the average interval between captures for these individuals which "succeeded" in being caught again during the study period cannot be taken as an estimate of capture probability p. If we did so we should overestimate value p (mean interval of time between captures is under-estimated). In such a case it is necessary to use the truncated distribution given in formula (7), and assess value p using equation (7a), where the mean interval between captures $\bar{x}_{a,b}$ is obtained empirically, and value p is calculated using equation (7a). As in the case of the exponential distribution, when solving equation (7a), use was made of the enclosed diagrams (Fig. 2) in which the relation is given between the value of parameter p and the value of the expression $(\bar{x}_{a,b} - a) = g_2\ (k)$, where $\bar{x}_{a,b}$ is defined by means of equation (7a), in which $(b - a + 1) = k$. The relation was given for $k = 2, 3 \ldots 10$.

The way in which parameter p is estimated on the basis of the enclosed diagrams has been illustrated by the following example:

From among individuals which "succeeded" in being caught twice in the period analysed only those individuals which remained in the area for at least 5 trapping units (5 days) after the first capture were taken for analysis of the time interval between succesive trappings. Out of this group of 92 individuals, 40 individuals were caught for the second time on the following trapping day (after the lapse of one day), 24 individuals after 2 days, 14 individuals after 3 days, 9 individuals after 4 days and 5 individuals after 5 days. Therefore individuals are caught on the average after

$$\frac{1{\cdot}40 + 2{\cdot}24 + 3{\cdot}14 + 4{\cdot}9 + 5{\cdot}5}{92} = \frac{191}{92} = 2.076 \text{ days.}$$

Since in accordance with the symbols in the text

$$a = 1, \qquad b = 5, \qquad \bar{x}_{1.5} = 2.076, \qquad \bar{x}_{1.5} - a = 1.076, \qquad k = 5$$

therefore for the curve $g_2(5)$ (Fig. 2) point 0.40 on the abscissa axis correponds to point 1.076 on the ordinate axis. This is the value of the estimator of the unknown parameter p.

Estimation of parameter p could also be analysed for those individuals only which were caught again after two days, three, four and five days. The average interval between captures is then

$$\frac{2{\cdot}24 + 3{\cdot}14 + 4{\cdot}9 + 5{\cdot}5}{52} = \frac{151}{52} = 2.9$$

since $a = 2$, $b = 5$, $\bar{x}_{2.5} = 2.9$ then $\bar{x}_{2.5} - a = 0.9$, $k = 4$. For the curve $g_2(4)$ in Figure 2 point 0.4 on the abscissa axis corresponds to point 0.9 on the ordinate axis.

The truncated geometrical distribution can also be applied to estimation of the number of rodents present in an experimental area for a certain period. It is assumed that trapping probability of a rodent is constant and is p, and that the group of rodents consisting of N individuals does not alter (in respect of numbers) as the result of some individuals disappearing and is not increased by the arrival of new individuals. This period must be short enough to satisfy these assumptions. Rodents from group N "reveal" their presence (are trapped) on successive trapping days. If we consider the successive day during the period in which the individual is caught as the variable, then this variable has the distribution given in equation (5). In order to satisfy the assumptions of the model, estimation of N is made on the basis of trapping results from the first few days of trapping during the study period. We then obtain the truncated distribution given in equation (7). Individuals caught from day a to day b are caught on the average on day $\bar{x}_{a,b}$. Value $\bar{x}_{a,b}$ is thus the arithmetical mean of the intervals (or trapping days) between the day the traps are set up and the day on which the individual living in this area is trapped. Calculations are made for those individuals which were caught not earlier than after a days but before the expiry of b days. Examination of N is made by the graphic (Grodziński, Pucek, Ryszkowski 1966) and analytical method (Zippin 1956, Janion, Ryszkowski, Wierzbowska 1968). Within the first method there are two variants of estimation of parameter p and N, often reciprocally mistaken. On this account the graphic method of estimating the parameters required has been presented in detail in this study.

Let us indicate by N the number of individuals living for a given short period, by p – trapping probability, then the number of individuals (Y_t) which should be caught (or marked) on successive t day of trapping (see equation 5) is:

$$Y_t = N p g^{t-1} \quad \text{where } g = 1 - p \tag{8}$$

Up to day t inclusive X_t individuals should be caught (or marked), where

$$X_t = N (1 - g^t) \tag{9}$$

and before day t the following should be caught

$$X'_t = N (1 - g^{t-1}) \tag{10}$$

The above equations show that

$$Y_t = -\frac{p}{g} X_t + \frac{p}{g} N \qquad t = 1, 2, 3 \ldots \tag{11}$$

and

$$Y_t = -pX'_t + Np \tag{12}$$

Estimation of parameter p and N is made by the graphic method (Fig. 3 – A, B). Using the relation given in Figure 3 – A, we obtain the value of parameter p making use of the fact that $\mathrm{tg}\,\alpha = \frac{p}{g}$, where $g = 1 - p$ and therefore

$$p = \frac{\mathrm{tg}\,\alpha}{\mathrm{tg}\,\alpha + 1}$$

where the value of the angle α is obtained empirically.

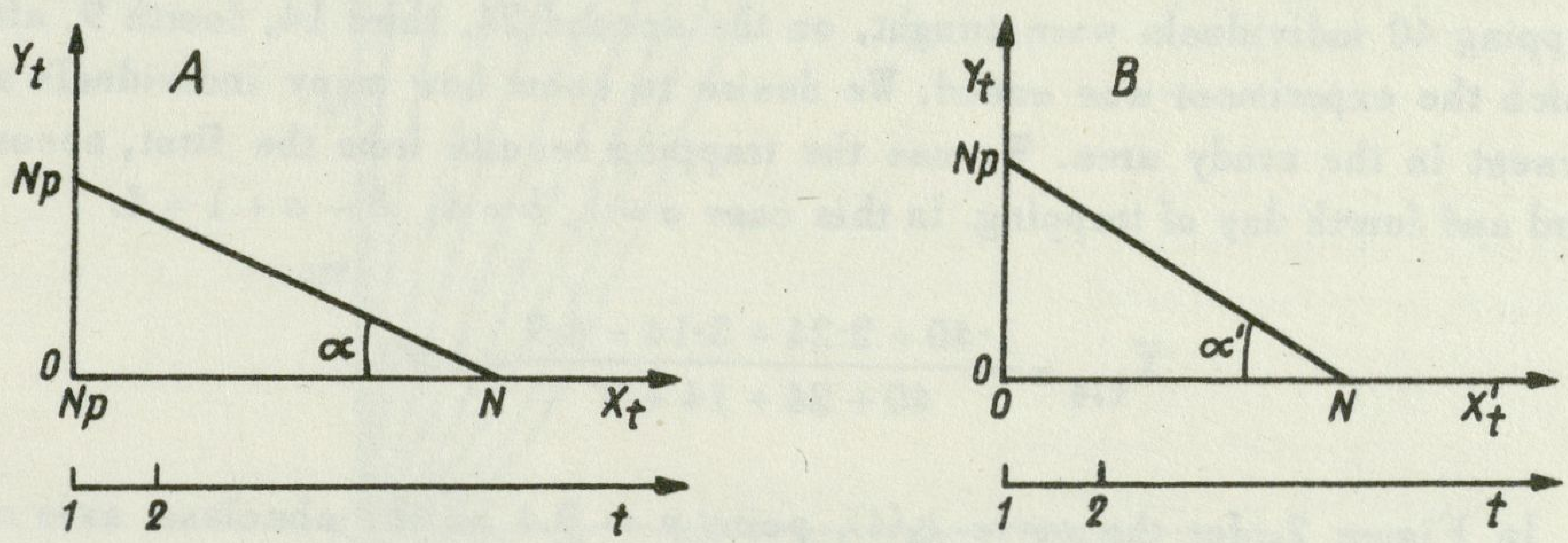

Fig. 3. Graph method of assessing parameters p and N
p – probability of capture, N – number of individuals

Knowing parameter p (and therefore g as well) we can estimate value of parameter N, making use of the fact that the distance (r) of the intersection point of line Y_t from the beginning of the system is equal to N $(1 - p)$, and therefore $N = \frac{r}{g}$. A simpler way of estimating the two parameters can be obtained by making use of the relation presented in Figure 3 – B. $p = \mathrm{tg}\,\alpha'$, while N is a point on axis X'_t, on which line Y_t intersects axis X'_t. This point is N units distant from the beginning of the system. It is clear from the above that it is more convenient to use the relation presented in Figure 3 – B.

The analytical method for estimating N on the basis of the truncated geometrical distribution was presented in the study by Janion, Ryszkowski, Wierzbowska (1968). Estimator N' of the N number of individuals present in the study area is given in the following form:

$$N' = \frac{N_{a,b}}{g^{a-1}\,(1 - g^{b-a+1})} \tag{13}$$

where $N_{a,b}$ is the number of individuals caught from day a to day b, $g = 1 - p$. Parameter g occurring in this equation is estimated on the basis of the distribution of number of days elapsing up to capture of an individual (see equation 7 and 7a) and then its value is introduced into equation (13) in order to calculate value N'.

The study by Janion, Ryszkowski, Wierzbowska (1968) contains tables facilitating the appropriate calculations. In the present study these tables have been enlarged and given in the form of graphs (Fig. 2, 4). In order to illustrate the way in which these graphs are used an example referred to earlier on this study will be given below.

The trapping of individuals gave the following results: on the first day of trapping 40 individuals were caught, on the second 24, third 14, fourth 9, after which the experiment was ended. We desire to know how many individuals are present in the study area. We use the trapping results from the first, second, third and fourth day of trapping. In this case $a = 1$, $b = 4$, $b - a + 1 = 4$.

$$\overline{X}_{1.4} = \frac{1{\cdot}40 + 2{\cdot}24 + 3{\cdot}14 + 4{\cdot}9}{40 + 24 + 14 + 9} = 1.90$$

In Figure 2, for the curve $g_2(4)$, point $p = 0.4$ on the abscissa axis corresponds to point $(\overline{x}_{1.4} - 1)$ on the ordinate axis. The value of parameter p obtained is introduced into equation (13). In our case $N_{a,b} = N_{1.4} = 87$, $1 - p = 0.6$ $g^{a-1} = 1$, $1 - g^{b-a+1} = 1 - g^4$. The value g^4 is read from Figure 4, where for curve $g^k = g^4$, point 0.13 on the ordinate axis corresponds to point $g = 1 - p = 0.6$ on obscissa axis. There $1 - g^4 = 0.87$. Finally

$$N = \frac{87}{0.87} = 100$$

Discussion of the possibilities of applying the truncated geometrical distribution could be further extended, giving more examples of its application.

This study, however, was aimed at showing the connections occurring between moments of total and truncated distributions, and the examples given served as numerical illustrations.

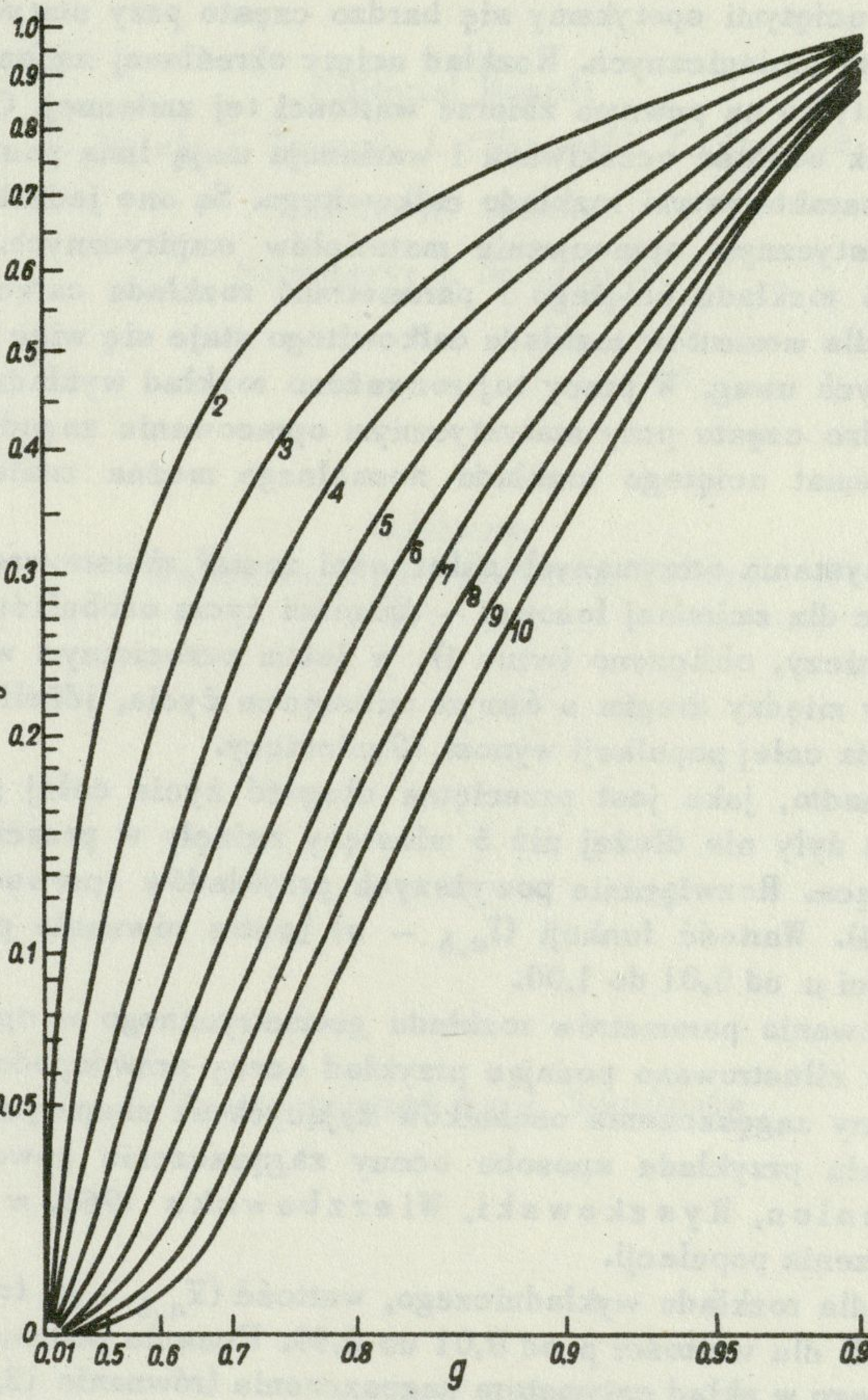

Fig. 4. Values of functions g^k for different values of parameter g; 2, 3 ... 10 values k

REFERENCES

1. Cramer, H., 1958 — Metody matematyczne w statystyce — Warszawa, 558 pp.
2. Grodziński, W., Pucek, Z., Ryszkowski, L. 1966 — Estimation of rodent numbers by means of prebaiting and intensive removal — Acta theriol. 11: 297–314.
3. Janion, S. M., Ryszkowski, L., Wierzbowska, T., 1968 — Estimate of numbers of rodents with variable probability of capture — Acta theriol. 13: 285–294.
4. Zippin, C. 1956 — An evaluation of removal method of estimating animal populations — Biometrics, 12: 163–189.

ROZKŁADY UCIĘTE I ICH ZASTOSOWANIE W EKOLOGII

Streszczenie

Z rozkładami uciętymi spotykamy się bardzo często przy statystycznym opracowaniu wielu zagadnień biologicznych. Rozkład ucięty określonej zmiennej losowej jest to rozkład określony tylko na pewnym zbiorze wartości tej zmiennej. Charakterystyki rozkładu uciętego, jak wartość oczekiwana i wariancja mają inną postać funkcyjną, inną niż analogiczne charakterystyki rozkładu całkowitego. Są one jednak mylone przez biologów przy statystycznym opracowaniu materiałów empirycznych. Podanie związku między momentami rozkładu uciętego i parametrami rozkładu całkowitego oraz analogicznego związku dla momentów rozkładu całkowitego staje się więc konieczne z punktu widzenia powyższych uwag. W pracy tej rozważono rozkład wykładniczy i geometryczny stosowany bardzo często przy statystycznym opracowaniu zagadnień biologicznych. Rozważania na temat uciętego rozkładu normalnego można znaleźć w opracowaniu Cramera (1958).

Sposób wykorzystania otrzymanych zależności został zilustrowany za pomocą przykładów. Przyjmując dla zmiennej losowej – długości życia osobników badanej populacji – rozkład wykładniczy, obliczono (wzór 4), w jakim przeciętnym wieku były te osobniki, które zginęły między drugim a ósmym miesiącem życia, jeżeli wiadomo, że przeciętna długość życia całej populacji wynosi 10 miesięcy.

Obliczono ponadto, jaka jest przeciętna długość życia całej populacji, wiedząc, że osobniki, które żyły nie dłużej niż 5 miesięcy zginęły w przeciętnym wieku wynoszącym 2,2 miesiąca. Rozwiązanie powyższych przykładów sprowadza się do rozwiązania równania (4). Wartość funkcji $(\bar{t}_{a,b} - a)$ [patrz równanie podane na wykresie (fig. 1)] dla wartości μ od 0,01 do 1,00.

Sposób oszacowania parametrów rozkładu geometrycznego w oparciu o ucięty rozkład geometryczny zilustrowano podając przykład oceny prawdopodobieństwa złowienia osobnika oraz oceny zagęszczenia osobników żyjących na eksperymentalnej powierzchni. Przy omawianiu przykładu sposobu oceny zagęszczenia powołano się na prace Zippin 1956, Janion, Ryszkowski, Wierzbowska 1968, w których zajmowano się oceną zagęszczenia populacji.

Podobnie jak dla rozkładu wykładniczego, wartość $(\bar{x}_{a,b} - a)$ (równanie 7a) podano na wykresie (fig. 2) dla wartości p od 0,01 do 0,99. Ponadto podano wartość wyrażenia $(1 - g^k)$ wchodzącego w skład estymatora zagęszczenia (równanie 13, fig. 4).

AUTHOR'S ADDRESS:
Dr. Teresa Wierzbowska
Instytut Ekologii PAN,
Warszawa, ul. Nowy Świat 72,
Poland.