Ann. Inst. Statist. Math.
31 (1979), Part A, 325-348

A NEW STATISTICAL METHOD TO ESTIMATE THE SIZE OF ANIMAL POPULATION*

—ESTIMATION OF POPULATION SIZE OF HARE—

CHIKIO HAYASHI, TSUTOMU KOMAZAWA AND FUMI HAYASHI

(Received Apr. 19, 1977; revised Apr. 6, 1979)

1. Introduction

Several statistical methods to estimate the size of wild animal population are discussed with their own characteristics in [1], where the following methods are especially investigated, (i) method by capture (hare hunting), (ii) method by capture-recapture, (iii) method by traces, (iv) method by pellets (droppings) and (v) method by amount of damage in forest. The well-known method in mathematical statistics to estimate the size of wild animal population is the capture-recapture method. It has been confirmed in the field experiment on wild rabbit population in an islet that this method may be at least impracticable in field survey of hare population, however it may be interesting as a theory on desk [2]. As for the following points, the method will not be applicable.

(1) Complete randomization does not hold. The condition of p_{ij} being always constant for all i and j is quite unrealistic where p_{ij} is defined as the probability that the ith animal is caught by the jth trap. However, the similar estimation, so far as the distribution of binomial type is described as likelihood function, may be realized if the density of traps in every sub-area is proportional to the density of wild animals in the very area even though the territories are found. This point is investigated in details in [2]. But this is not generally known before survey.

(2) The capture of live animals is, generally speaking, very difficult. Especially the capture of a number of live animals is impracticable at least in the case of hare. The high precision is not secured if the proposed number is small.

---

* The present paper is a revised and enlarged one of the report presented to PT-section of JIBP (Japanese International Biological Program). The Section 8, sampling variance of $\bar{d}$ is newly added although the remaining sections are revised ones of the articles already published.

Thus we have reached to the conclusion that the capture-recapture method is rather a fantastic story in hare population in Japan. In spite of the fact this method seems to be useful in estimating the distribution of the range of movement of wild animal, though it is not applicable to estimate the size of population [1]. The methods of (iv) and (v) mentioned above require further examination, because they are not available so far as the fundamental field techniques are not developed [9].

The capture method by hare hunting is a usual method in sample survey techniques. Sample spots are drawn by the space-time sampling by independent equal probability system. That is to say, the hare hunting is done in sample spots at the designated time and the number of captured animals are counted. The estimator is made by these numbers and the variance of the estimator is calculated on a time-space sampling scheme. However, this method is laborious and very much expensive. On this point, the capture method is impracticable however it may be possible in the sense of statistics.

In the present paper, the trace method will be presented in estimating the population size of hare. This method is based on the information of tracks of hares in fresh snow. Methodologically new ideas are shown in the estimation of the size by (i) the estimation of total length of tracks in an area based on a probability model for reduction of measurement errors (Section 4) and (ii) the estimation of mean track length of a single hare based on a probability model (Section 7). The characteristics of this method are expressed as the combination of a probability model with both the sample survey and computer simulation.

2. Fundamental idea

The present trace method is useful in winter at snowy districts. Here tracks of hare on fresh snow are utilized. The hare is nocturnal and it is reported that hares are active in the interval from 10 P.M. to 4 A.M., especially from 11 P.M. to 3 A.M. [3]. The trace is marked on snow during the night. Next day, we can observe the tracks during the previous night and measure them, where the tracks are easily distinguishable from those on earlier nights by the changing pattern of traces in the absense of further snow. If this condition is fulfilled, the statistical method mentioned below is applicable. The question is to estimate the number of hares in the field which we call A. Theoretically speaking, we assume that A is closed and hares are only in A and there is no death and birth of the hares during the period of survey, i.e. the size is preserved to be constant. Practically speaking, the assumption that area A is closed, says that A is so large that the imbalance of the number of in-going and out-going hares in the field A is almost disregarded in comparison with the total number of hares inhabitting in A and the size is approximately unchangeable.

As an experimental field, we fix a field which have a fairly large area of about 5600 hectares and another which has an area of about 250 hectares. First, we estimate the total length of tracks in A in one night. Let X be the total length of tracks in A and x be the estimator of that total length which is obtained by a sampling method. Let L_i be the running length of the ith hare in the same one night. We take $L=\sum_{i=1}^{N} L_i/N$ where N is the total (size) of hare population in A, i.e. L is the mean value of the running lengths of hares in A in one night*. Let $\bar{d}$ be the estimator of L. It holds clearly $X=NL$, i.e. $N=X/L$. Thus we make $x/\bar{d}$ as an estimator of N. This is a ratio estimate. We have x and $\bar{d}$ as shown below as random variables, the means and variances of which are $\mathrm{E}(x)$, $\mathrm{E}(\bar{d})$ and σ_x^2, $\sigma_{\bar{d}}^2$ respectively. We use an unbiased estimator x of X, i.e. $\mathrm{E}(x)=X$ in rigorous sense and use an estimator $\bar{d}$ which is regarded as unbiased as much as possible, i.e. $\mathrm{E}(\bar{d})=L$ holds approximately. The reason why we mention this is as followings; the estimate of X is based on a usual sampling method in a rigorous sense, however the estimate of L is not always based on a formal sampling theory and the method is only adopted which is regarded practically as unbiased. The details will be described in Sections 4–7.

The mean square error of the estimator $\hat{N}=x/\bar{d}$ is given in a well-known form as

$$\tau_{\hat{N}}^2/N^2 \doteqdot \sigma_x^2/X^2+\sigma_{\bar{d}}^2/L^2$$

where $\tau_{\hat{N}}^2$ is the mean square error of $\hat{N}$ and $\tau_{\hat{N}}^2=\mathrm{E}(\hat{N}-X/L)^2=\mathrm{E}(\hat{N}-N)^2$, because $\hat{N}$ and $\bar{d}$ are regarded as independent.

Thus the population size of hare can be calculated from the total length of tracks and the mean value of track of a hare in one night which are estimated by separate field surveys under the same conditions during the same period.

---

* This expression is recognizable for the fixed one night. In theoretical treatment this idea is reasonable. In a field experiment the tracks in several nights are usually surveyed. So, strictly speaking, L_i is to be regarded as a random variable in the period of survey however it may be complicated. The measured value of L_i is regarded as a realization of that random variable. Let $\tilde{L}_i$ be the mean value of that random variable. It is desirable to estimate $\tilde{L}=\sum_{i=1}^{N} \tilde{L}_i/N$ based on the time-space sampling scheme. This holds for the estimation of X as mentioned later. If the mean values for X and L in the period of survey are considered, the following idea is applied on the same principle. For simplicity, the case for that one night is treated as below.

3. Estimation of X

The fundamental idea to estimate X is shown as below. Suppose that A is divided into M sub-areas $S_1, S_2, \cdots, S_M$ the area of which is constant i.e. equal to U, for which $S=\sum_{g=1}^{M} S_g = MU$ holds. The shape of sub-area may not be always square and is dictated according to the needs of field survey. The sample of size m is drawn with equal probability without replacement. The sample of size m is expressed as $(s_1, s_2, \cdots, s_m)$. The length of tracks of hares is measured in every s by surveying. Let x_i be the length of tracks expressed in meters in s_i, $i=1, 2, \cdots, m$. $x=\bar{x}M=M\sum_{i=1}^{m} x_i/m$ is taken as the estimator of X, where $\bar{x}$ is the sample mean. Of course $\bar{x}$ is unbiased and the variance of x is $\sigma_x^2=\{(M-m)/(M-1)\}\cdot(\sigma^2/m)\cdot M^2$ where $\sigma^2=\sum_{g=1}^{M}(x_g-\bar{X})^2/M$, $\bar{X}=\sum_{g=1}^{M} x_g/M$ where x_g is the length of tracks in the gth sub-area. If the measurements of the length in sample spots had to be made in different days, the time-space sampling system with independent equal probability would be adopted. The ith spot s_i is surveyed in the jth day assigned to s_i with equal probability and independently in the interval between T_1 and T_2. In this case, the estimator for X is considered, where X is $\sum_{t=T_1}^{T_2} X(t)/(T_2-T_1)$ and $X(t)$ is the total length of tracks of hare in A on the tth day. This X is the mean value with respect to space and time. Let x_{ij_i} be the total length of tracks of hares in the ith spot s_i on the j_ith day. x_{ij_i} is obtained by surveying. We take $x=\bar{x}M=M\sum_{i=1}^{m} x_{ij_i}/m$ as the estimator for X mentioned above. Now suppose that E means the mean value with respect to space and time. Here we have $\mathrm{E}(x)=X$ and the variance σ_x^2 of x is as follows under the condition of sampling mentioned above.

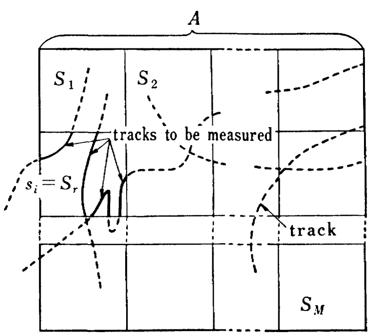


Fig. 1

$\sigma_x^2=(\sigma_\alpha^2+\sigma_\beta^2)M^2/m$ except for the coefficient due to finite population where σ_α^2 is the variance between spots and σ_β^2 is the variance between days, i.e. $\sigma_\alpha^2=\sum_{i=1}^{M}(\tilde{X}_i-\bar{X})^2/M$ and $\sigma_\beta^2=(1/M)\sum_{i=1}^{M}\left\{\sum_{t=T_1}^{T_2}(X_i(t)-\tilde{X}_i)^2/(T_2-T_1)\right\}$ where $\tilde{X}_i$ is the mean value of $X_i(t)$ with respect to time and $\bar{X}$ is the mean value of X with respect to space and time and $X_i(t)$ means the length of tracks in the ith spot at the time t and is equal to x_{ij_i} obtained by survey, t being the j_ith day. The estimate of $\sigma_\alpha^2+\sigma_\beta^2$ is easily obtained by $V^2=\sum_{i=1}^{m}(x_{ij_i}-\bar{x})/(m-1)$ using the obtained data, if M is large.

The fundamental sampling scheme is mentioned above, however the work of measurement of x_{ij_i} is rather laborious. So several methods by easier field work without measurement error were presented based on the mathematical model using geometrical probability [1], [4]. Thus we fixed a method called INTGEP (Intersection Points Counting Method Based on Geometrical Probability).

4. INTGEP

A rectangle is taken as S, the base line of which is l meters on the level and another side of which is h meters on the level. The l and h are taken comparatively small enough for the tracks in the rectangle to be regarded as straight line, for example $l=10$ meters and $h=2$ meters. If h is small and equal to 2 meters, we can obtain the reliable data by simple observation without rigorous measurement even in complicated ground condition except rare ambiguous cases in which rigorous measurement is done. Three models are presented.

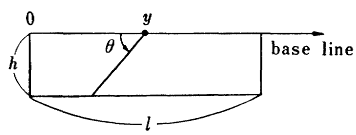


Fig. 2

Let $\overline{\varphi(l, h)}$ be the mean length of a track in the rectangular area mentioned above, the length of the base line being l meters.

(Model A) Assumption; A point on the base line is drawn with equal probability, i.e. probability element dy/l and then a line is drawn in the rectangular area with an angle determined with equal probability, i.e. with the probability element of $d\theta/\pi$. The probability distribution function of the length of that line contained in the rectangular area is required and its mean value is calculated with its variance.

$$\overline{\varphi(l, h)}=\frac{1}{\pi l}\left[l^2\log\left\{\sqrt{\left(\frac{h}{l}\right)^2+1}+\frac{h}{l}\right\}-h(\sqrt{l^2+h^2}-h)\right.$$
$$\left.-2lh\log\left\{\sqrt{\left(\frac{l}{h}\right)^2+1}-\frac{l}{h}\right\}\right]$$

$$\sigma_\varphi^2=\frac{2lh}{\pi}-\overline{\varphi(l, h)}^2 .$$

(Model B) Assumption; A point on the rectangle is drawn with equal probability (probability element $dy/2(l+h)$) and then a line is drawn in the rectangular area with an angle determined with equal probability (probability element $d\theta/\pi$). The probability distribution function of the length of the line which is contained in the rectangular area and meets the base line is required, and its mean value $\overline{\varphi(l, h)^B}$ is calculated with its variance. In this case the mean value $\overline{\varphi(l, h)^N}$ of the length of the line which is contained in the rectangular area and does not meet the base line is also calculated with the probability of the occurrence of those events.

Using $\overline{\varphi(l, h)^B}$, $\overline{\varphi(l, h)^N}$ with the probabilities of occurrence of those events and the observed data, the required $\overline{\varphi(l, h)}$ is calculated.

Other several models were investigated. The following model turned out to be the best by the field survey.

(Model C) The assumption is the same as for the model B. The probability distribution function of the length of that line contained in the rectangular area is obtained and its mean value is calculated with its variance.

$$\begin{aligned}\overline{\varphi(l, h)} = \frac{1}{\pi(l+h)} \{ & l(l+2h) \log (\sqrt{1+(h/l)^2}+h/l) \\ & +h(h+2l) \log (\sqrt{1+(l/h)^2}+l/h) \\ & -h(\sqrt{l^2+h^2}-h)-l(\sqrt{l^2+h^2}-l)\}\end{aligned}$$

$$\sigma_\varphi^2 = \frac{2lh}{\pi} - \overline{\varphi(l, h)}^2 .$$

The difference between the model B and the model C is found in the field measurement. In the model B, only the numbers of the points of the intersection of a base line and tracks of hares are counted, whereas in the model C, all the numbers of the points of the intersection of a rectangle and tracks of hares are counted.

Thus the estimate of the total of the length in the rectangular area of $l \times h$ square meters, i.e. x_{ij_i} is obtained using the model C with increased accuracy, where the estimator is $\{\overline{\varphi(l, h)} \cdot q_i/2\}$, q_i being the total number of the points of the intersection of the ith rectangle and tracks. The estimate of the total length of tracks in A is obtained by summing up the estimator mentioned above with respect to i divided by m (the size of spots) and multiplied by M. The variance of the estimate becomes smaller with the increase of the number of tracks in the sample areas of size m.

The best model was confirmed to be C on the data-check (See Section 5). The field work is shown as below in detail.

See Fig. 3. Walking on the base line from the starting point to the end point (the distance between them is l meters (10 meters on the

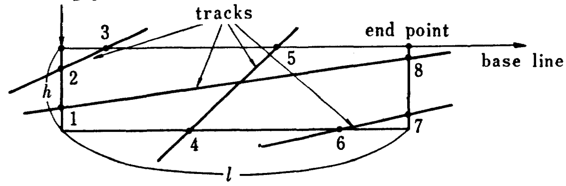
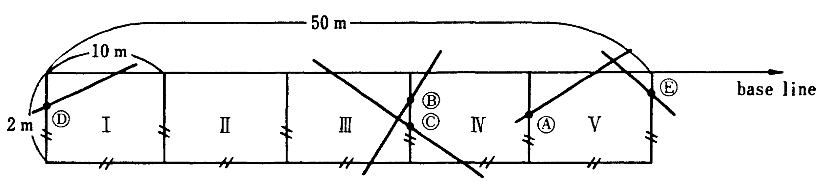


Fig. 3

level)), we only count the number of intersection points. In Fig. 3, we count 8 points. As mentioned previously, the points 1, 2, 3, 4, 5, 6, 7 and 8 will be generally obtained by simple observation of trained field workers by setting an imaginary rectangle without rigorous measurement when $h=2$ meters. If they can not decide, the measurement must be done, for example the points 1, 2, 4, 6, 7 and 8 may not be decided whether these are on the side of the distance of 2 meters from the base line, or not. The intersection points divided by 2 is the number of tracks of hares, i.e. $8/2=4$ is the number of tracks. Thus the length of tracks in the rectangle is estimated to be equal to $4\cdot\overline{\varphi(10, 2)}$.

In the field survey, we adopt the following methods. The starting points are drawn with equal probability in the sense of area on the level, i.e. the lattice points with equal interval are put on the map and the sample lattice points are drawn with equal probability. As a next step, the direction of the base line is drawn with equal probability in the sense of $d\theta/2\pi$ being constant where θ means declination. Conventionally an azimuth angle is determined with the same probability of 1/360. Thus a rectangular area to be surveyed is practically determined. Such rectangular areas are $s_1, s_2, \cdots, s_m$ as mentioned in Section 3. The x's are determined as mentioned above only by counting the intersection points. However x's are not always equal to the actual length as x's are random variables, the number of sample spots and the number of tracks are larger, the precision of the estimated total length are higher. The precision is easily calculated in the form of variance, the coefficient of variation of the estimated total length of tracks in area A being $\{\sigma_\varphi/\overline{\varphi(l, h)}\}/\sqrt{(\text{the total number of tracks})}$ under some trivial



Mark // means imaginary sides.
Ⓓ and Ⓔ are counted once. Ⓑ, Ⓒ and Ⓐ are counted twice, i.e.
Ⓑ and Ⓒ being counted in III and IV, Ⓐ being in IV and V.

Fig. 4

statistical assumption.

Usually, we take five successive rectangles of (10 meters × 2 meters) instead of only one rectangle as in the form of Fig. 4.

The loss of information has been verified to be small in the investigation of field survey mentioned in Section 5. We call the method, based on the model C and the field work mentioned above, INTGEP.

5. Verification of the model C by data

The field survey of tracks of hares (Lepus brachyurus lyoni) was done in the north west area of 5577 hectares (under 500 meters above see level except streets) in the Island of Sado, Niigata prefecture. 110 spots (rectangular areas of $l=50$ meters and $h=5$ meters) are drawn with equal probability, i.e. the starting points and the declinations were drawn at random as mentioned in Section 4. The survey was done in the period between the end of January and the beginning of February in 1968. The tracks in the rectangles were fully measured by surveying. As a result of the survey, the data were obtained as in Table 1.

Table 1

the percentage of the spots where tracks were found	47/110=43%
the maximum of the number of tracks found in a spot	59
the maximum length of tracks found in a spot	360.4 m (in $50\,m \times 5\,m = 250\,m^2$)
mean length of tracks in a spot	22.1 m
mean length of tracks in 1 hectare	885 m
mean length of tracks in a spot in the spots where the tracks were found	51.8 m (in $50\,m \times 5\,m = 250\,m^2$)

The confidence interval of the estimated value of 885 meters in 1 hectare is calculated to be ± 372 meters under the confidence level of 95%.

We investigate the model C using these data. The $\overline{\varphi(l, h)}$ of the model C is calculated for l's and h's, the graph of which is shown in Fig. 5.

If we take $l=10$ meters and $h=2$ meters, $\overline{\varphi(l, h)}$ of the model C is 2.95 meters in $10\,m \times 2\,m = 20\,m^2$.

We constructed the spot S (rectangle, $10\,m \times 2\,m = 20\,m^2$, and successive five rectangles) by using the survey maps made available. The estimate of the model A was very poor. The estimates were stable and nearly constant according to the ways of construction of the spots

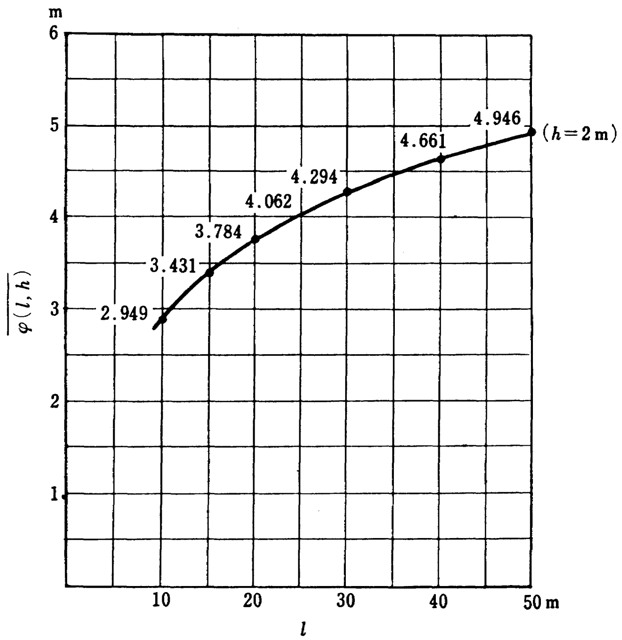


Fig. 5 $\overline{\varphi(l, h)}$ by the model C.

Table 2 The length of tracks estimated in 1 hectare

	Model B	Model C	True (data)
	1015 meters	881 meters**	885 meters*
Confidence interval of estimate under the confidence level of 95%	—	47 meters	— *
Number of tracks observed		657	

* No variance because the estimate by model was made to estimate this value. This meaning is different from the estimation of the length of tracks in A mentioned previously.

** This is given by two-sides-estimate (see Fig. 6). The estimate only on the one side (base line) measurement is 888 meters and the confidence interval is 66 meters under the confidence level of 95%.

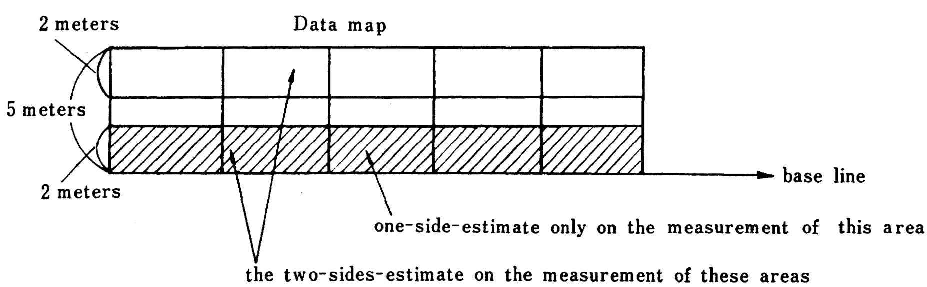


Fig. 6

in the model B and the model C. The result of the comparison is shown (Table 2).

Thus, the model C was confirmed to be useful, as it shows a desirable goodness of fit for the data, whereas the model B was turned out to give poorer fit. The details are described in [4].

In Sections 3-5, we presented the reliable method of estimating the total length X of tracks in a field survey by INTGEP, based on the model C.

6. Estimation of L —COC method—

We have had two ideas to estimate L. The COC method (Collar of Coloring Matter) and the RST method (Randomly Selected Trace method) [5], [6]. Here the COC method will be shown.

First alive hares are caught by a box-trap (Fig. 7) deviced by our colleagues, Dr. Ueda and Dr. Shibata in the Experimental Station of Forestry, Hokkaido Branch.

Hares are caught at night. Next morning, we attach a collar filled with coloring matter to them (See Fig. 8), deviced by C. Hayashi, Dr.

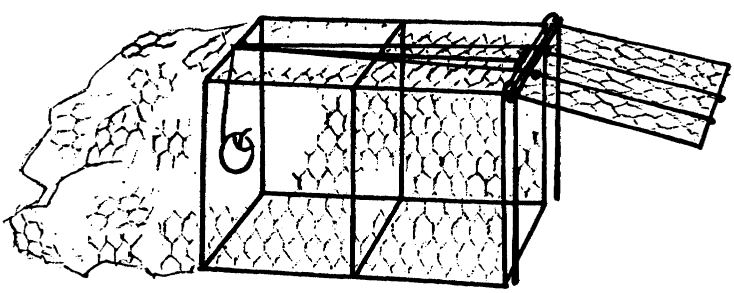
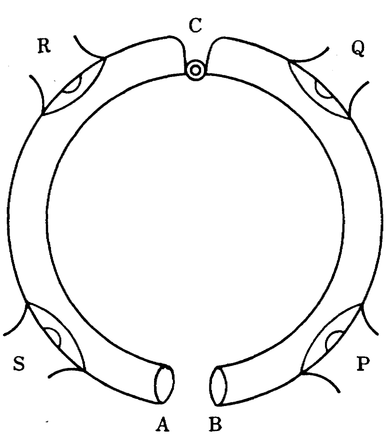


Fig. 7 Box-trap.



A, B: Mouth, C: Hinge, P, Q, R, S: holes, diameter of which is 3 mm, with protecter and meshes of stainless steel. The rubber stoppers are used in A and B and vinyl tape is used to fix A and B.

Fig. 8 Collar of coloring matter.

J. Toyoshima and Dr. T. Niwaguchi [5].

This collar is made of aluminum tube the thickness and the inner diameter of which are 0.75 millimeters and 9.5 millimeters respectively. The inner periphery of this tube is 160~170 millimeters in length. The weight is 22.5~23 grams including two rubber stoppers of 0.5 grams. The coloring matter included in this tube is 10 grams in weight. The total weight of this collar is about 33 grams. The weight is regarded as neglisible comparing with the body weight of hares (2300 grams~4000 grams) without influencing the activity.

The coloring matter consists of 70 units of Eosin Yellow, 30 units of Silic Acid and 2 units of Calcium Phosphate ($Ca_3(PO_4)_2$) in weights.

A bit of the coloring matter drops on the snow from P, Q, R, S as hare runs, stops and moves, and leads us when we trace. Thus we can identify the hare in question.

We attach the collar to the hare caught during the previous night and release the hare in the morning. Next morning we follow the tracks by the color spots on the snow. We measure the distance in level between the first rest place in the day-time and the second rest place (See Fig. 9). This is L_i. Repeating these processes we can make an estimator $\bar{d}$ of L.

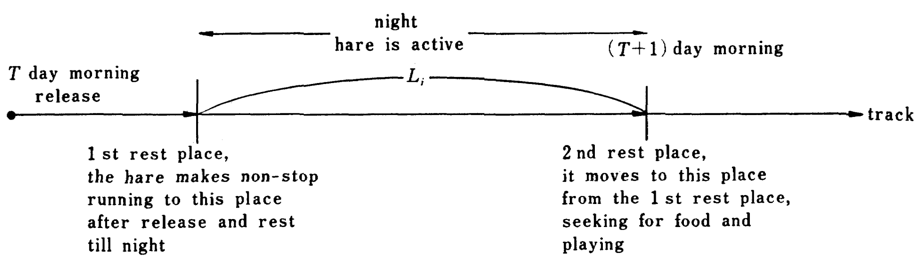


Fig. 9

[Data]

The data of hare (Lepus timidus ainu) on this method were obtained in Nopporo, near Sapporo in Hokkaido by Dr. Y. Shibata, in the period between the middle of February and the middle of March. Nopporo is stretched in a flat plane with shallow valleys [5]. The distance to the first rest place, i.e. the distance of non-stop running from the release point is fairly stable (Fig. 10). The mean value is 476 meters.

$\bar{d}$ is 1373 meters where $\bar{d}=(1813+318+1341+610+1805+1693+1675+937+1311+1745+1857)/11$, the number being the distance in level measured in meters, 11 being the sample size. The standard deviation of 508 meters of the L_i is fairly large, the coefficient of variation being 0.37.

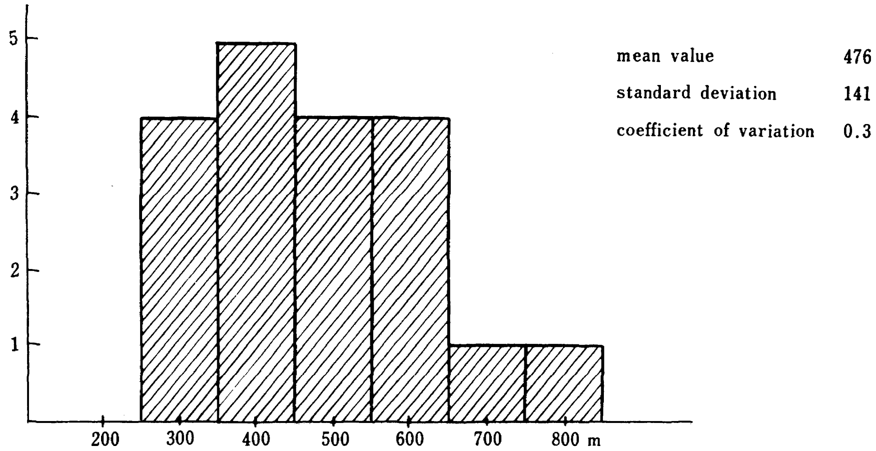


Fig. 10

7. Estimation of L —RST method—

We trace the natural tracks (See Fig. 11). From any point, A group and B group trace the track measuring the distance forward and backward directions of a hare's running track until the 2nd rest place and the 1st rest place are found respectively. Often it crosses with other tracks. However, we continue to measure, if we can find out the track

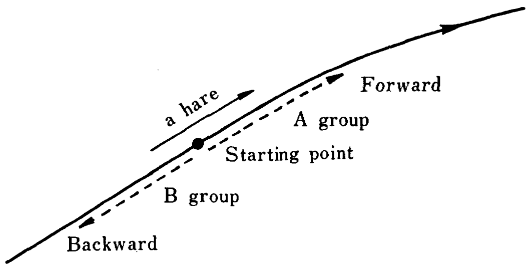


Fig. 11

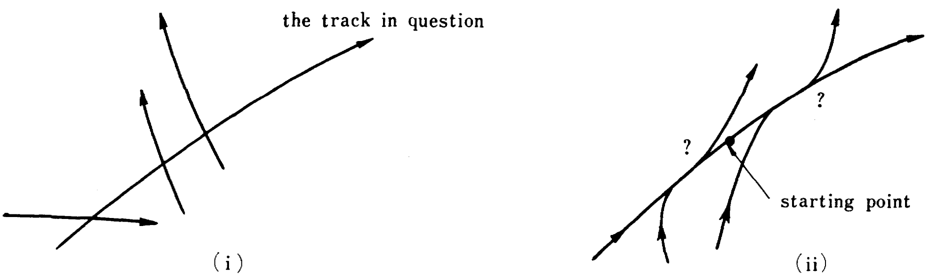


Fig. 12

which we have been following (See Fig. 12 (i)). But in the case of (ii) in Fig. 12, it is impossible to find the track in question. At the point of intersection of two tracks where we can not decide which track is for hare in question, we select the track to survey at random (with equal probability). We continue this process, at the points of intersection mentioned above, to find the starting point of a trace and the end point of the (or another) trace. If the track in question is traced, the departure point of a hare corresponds to the 1st rest place and the end point corresponds to 2nd rest place.

Thus we obtain the distance d' on the level which is not always equal to the running length of a particular hare. The method to estimate the length L from d' is required. For this purpose, we set three models, i.e. the model I, the model II and the model III. An assumption is introduced that the points of intersection are put on all tracks independently with equal probability and matched at random and further we assume that every track does not form any loop and every track has at most only one point of intersection with the other particular track (every track may intersect with many other tracks). This does not mean the fact of actual crosses as far as the track being measured is distinguishable however it may be crossed actually, i.e. the cross in the assumption means only the cross where the track in question can not be decided, and the cross where it can be decided is not regarded as cross in the assumption. Thus the assumption is not so severely restricted.

[Model I] All tracks T_i $(i=1, 2, \cdots)$ to intersect are equal to L (>0) in length with the same number K (>0) of the points of intersection.

[Model II] All tracks T_i $(i=1, 2, \cdots)$ to intersect are equal to L (>0) in length with the random number K_i (>0) of the points of intersection which obeys to some probability distribution law of a mean K and a variance σ_K^2.

[Model III] The ith track T_i is a two-dimensional random variable the components of which are a positive random variable L_i that obeys to some probability distribution law of a mean L and a variance σ_L^2 and a positive random variable K_i that obeys to some probability distribution law of a mean K and a variance σ_K^2, L_i and K_i being independent to each other. We write $T_i(L_i, K_i)$. All random variables are assumed to be mutually independent*.

* There may be a question that the assumption of independence between L_i and K_i is not natural. Theoretically speaking this idea is considered to be reasonable. But the mathematical expression of this "dependence" requires some additional assumptions which can not be practically verified. This is the reason why we adopt the simplest assumption as a first step. Of course this point must be reconsidered if the data show a large gap between the theory.

In the model I, we can calculate the mean of d', $\mathrm{E}(d')$, by recurrence formula in elementary probability calculus.

$$\mathrm{E}(d')=\frac{L}{K+1}\left\{1+\frac{2K}{K+1}\left(\frac{K}{1-(1/2)^K}-1\right)\right\}, \qquad K\geqq 2.$$

We have $d'=L$ in the case $K=1$ and $\mathrm{E}(d')\to 2L$ as $K\to\infty$. It is very interesting and gives an efficient information that $\mathrm{E}(d')/L$ is monotonely increasing but does not exceed 2 for any K.

In the model II and the model III, we obtain $\mathrm{E}(d')/L$ and $\{\mathrm{E}(d'^2)-\mathrm{E}(d')^2\}/L^2=\sigma_{d'}^2/L^2$ by computer simulation (1000 trials in each case) in

Table 3 $\mathrm{E}(d')/L$ ($L=1000$) by computer simulation

(1000 trials in each case of K)

K	M I		M II**		M III**		Theoretical mean*
	mean	standard deviation	mean	standard deviation	mean	standard deviation	
1	1.003	0.264	—	—	—	—	1.000
2	1.064	0.448	—	—	—	—	1.074
4	1.249	0.663	1.203	0.735	1.218	0.693	1.245
6	1.419	0.891	1.327	0.878	1.359	0.894	1.391
8	1.482	0.999	1.515	1.013	1.449	1.012	1.500
10	1.593	1.157	1.556	1.135	1.575	1.180	1.580
15	1.678	1.141	1.709	1.257	1.711	1.274	1.703
20	1.771	1.285	1.755	1.350	1.768	1.345	1.771
40	1.873	1.281	1.881	1.425	1.923	1.377	1.880

* theoretical formula; $\frac{1}{K+1}\left\{1+\frac{2K}{K+1}\left(\frac{K}{1-(1/2)^K}-1\right)\right\}$.

** In these models, computer simulation was done only for $K\geqq 4$, because of the discrete realization of the Gaussian distribution of the range of 6.

Observed number of points of intersection

(1000 trials in each case of K)

K	M I		M II		M III	
	mean	standard deviation	mean	standard deviation	mean	standard deviation
1	1.000	0				
2	2.219	0.938				
4	5.157	2.895	4.895	2.897	4.935	2.914
6	8.883	5.656	8.171	5.070	8.460	5.752
8	12.372	8.343	12.475	8.726	12.157	8.502
10	16.368	11.218	16.014	11.832	16.139	11.323
15	25.766	17.669	26.162	19.389	26.155	17.731
20	36.428	26.350	35.935	26.628	36.347	25.731
40	71.183	53.028	79.889	57.961	77.187	55.252

the case where the probability distributions are Gaussian, i.e. L_i obeys to $N(L, \sigma_L^2)$ and K_i obeys to $N(K, \sigma_K^2)$ for all i. In practice a symmetric discrete distribution except 0 is used as approximation of Gaussian distributions the coefficient of variation of L being 1/6 and the range of K being 6. Thus we find an important information that $\{E(d')/L\}$'s in three models are approximately equal and also almost equal to that of theoretical formula by the model I (see Table 3 and Fig. 13). We also have the mean and variance of the number of the points of intersection in three models by computer simulation (see Table 3). The details are described in [7].

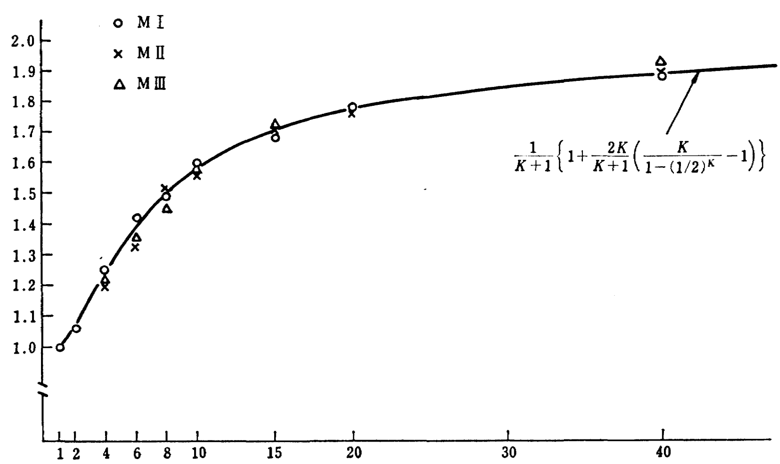


Fig. 13

But, as this value is based on the unknown K, the estimate of $E(d')/L$ can not be obtained. The estimate of $E(d')/L$ must be based on the observed number k of intersecting points in the sense mentioned previously. We adopt the Bayesian estimate of $E(d')/L$ from the observed number of randomly intersected points in the model III. The prior distributions of K are assumed. A posteriori distribution of $E(d')/L$ on the prior distribution is calculated for observed number k of intersecting points, $k=1, 2, \cdots$, by computer simulation in [8].

Suppose that the prior distributions are

$A_1(U)$: equal distribution on the range of U

$A_2(U)$: equilateral triangle distribution on the range of U

$A_3(U)$: right-angled triangle distribution of J type on the range of U

$A_4(U)$: right-angled triangle distribution of L type on the range of U

where 6 and 10 are taken for U, which are considered to be reasonable for the actual data. Let $P(K|k)$ be the Bayesian probability function on K on observed number, k, $\varphi(k|K)$ be the probability function of k on K and $\Psi(K)$ be prior distribution of K, i.e. $A(U)$'s. Here we have

$$P(K|k)=\frac{\varphi(k|K)\Psi(K)}{\sum_K \varphi(k|K)\Psi(K)}$$

which is calculated by computer simulation according to $\Psi(K)$. Also we have $\mathrm{E}(d')/L=r(K)$ as the function of K by computer simulation as mentioned previously. Thus we have the mean value of $\mathrm{E}(d')/L$ as the function of k

$$\sum_K r(K)P(K|k)=R(k)$$

and the variance

$$\sigma(k)^2=\sigma_w(k)^2+\sigma_b(k)^2$$

where $\sigma_w(k)^2=\sum_K \sigma(r(K))^2\cdot P(K|k)$, $\sigma_b(k)^2=\sum_K (r(K)-R(k))^2\cdot P(K|k)$ and $\sigma(r(K))^2$ means the variance of d'/L for a constant K. The mean and the variance are shown in Tables 4 and 5. In practical calculation, the missing values in Table 3 for $K\leqq 3$ in M III were set by extraporation curve as in Fig. 13.

From these tables, we make a table of $\mathrm{E}(d')/L$ to be used reasonably and practically as those values are not so different according to prior distributions. This is shown in Table 6 which is arranged from $A_1(6)$ for $k\leqq 3$ and from $A_1(10)$ for $k\geqq 5$ and from the arithmetic mean

Table 4 $R(k)$

Prior distribution / Observed k	$A_1(6)$	$A_2(6)$	$A_3(6)$	$A_4(6)$	$A_1(10)$	$A_2(10)$	$A_3(10)$	$A_4(10)$
1	1.035	1.047	1.076	1.018	1.046	1.082	1.117	1.028
2	1.080	1.080	1.121	1.049	1.115	1.138	1.196	1.075
3	1.125	1.113	1.158	1.085	1.176	1.180	1.246	1.124
4	1.161	1.143	1.182	1.125	1.222	1.210	1.276	1.171
5	1.186	1.163	1.201	1.150	1.265	1.240	1.309	1.209
6	1.200	1.180	1.210	1.174	1.294	1.256	1.333	1.235
7	1.208	1.192	1.215	1.186	1.315	1.271	1.351	1.253
8	1.210	1.192	1.217	1.187	1.322	1.278	1.356	1.261
9	1.219	1.206	1.222	1.205	1.349	1.292	1.379	1.282
10	1.221	1.210	1.223	1.209	1.357	1.300	1.385	1.291
15	1.222	1.214	1.224	1.213	1.394	1.335	1.415	1.329
20	1.228	1.231	1.228	1.231	1.427	1.365	1.442	1.364

Table 5 Variance of d'/L, $\sigma(k)^2$

Prior distribution / Observed k	$A_1(6)$	$A_2(6)$	$A_3(6)$	$A_4(6)$	$A_1(10)$	$A_2(10)$	$A_3(10)$	$A_4(10)$
1	0.194	0.236	0.299	0.151	0.222	0.314	0.395	0.177
2	0.310	0.314	0.400	0.238	0.389	0.437	0.566	0.297
3	0.405	0.379	0.472	0.320	0.518	0.519	0.663	0.406
4	0.473	0.434	0.515	0.401	0.607	0.574	0.718	0.496
5	0.526	0.472	0.558	0.450	0.694	0.636	0.783	0.576
6	0.552	0.502	0.573	0.493	0.752	0.669	0.831	0.626
7	0.565	0.521	0.582	0.512	0.791	0.696	0.865	0.660
8	0.574	0.526	0.591	0.517	0.808	0.713	0.875	0.680
9	0.594	0.554	0.605	0.552	0.863	0.745	0.923	0.723
10	0.598	0.561	0.607	0.559	0.878	0.760	0.934	0.741
15	0.609	0.571	0.616	0.570	0.950	0.831	0.991	0.819
20	0.634	0.631	0.634	0.631	1.016	0.892	1.045	0.891

of $A_1(6)$ and $A_1(10)$ for $k=4$.

The values are under random fluctuation and so had better to be smoothed. The smoothed values are called adjusted values in the table and to be adopted in practice.

Thus an estimate of L is given by d i.e. d' divided by the adjusted value in Table 6 according to observed k. The sample mean $\bar{d}$ of d' is obtained with the variance by field data of many tracks and Table 6.

Table 6

k	$R(k)$	Smoothing (adjusted value)	Variance of d'/L, $\sigma(k)^2$ Smoothing (adjusted value)
1	1.04	1.02	0.18
2	1.08	1.08	0.31
3	1.13	1.14	0.44
4	1.19	1.20	0.56
5	1.27	1.25	0.67
6	1.29	1.29	0.74
7	1.32	1.31	0.79
8	1.32	1.33	0.83
9	1.35	1.35	0.86
10	1.36	1.36	0.88
15	1.39	1.40	0.96

If $k=0$, $E(d')/L=1$ is taken the variance of which is $(1/6)^2$ according to the model.

[Data]

The data of hare (Lepus brachyurus lyoni) by the RST method in the north-west part of Sado Island, Niigata Prefecture, a steep and mountainous district with many valleys, were given by Prof. J. Toyoshima et al. in the period between the end of January and the beginning of March, detail of which were published in [5] and [8]. The estimate of L is found to be smaller than those shown in Section 6. This seems to be due to the difference in kind of hares and in the configuration of ground. Lepus timidus ainu is heavier and stronger than Lepus brachyurus lyoni. The field in Hokkaido is a flat field with simple configuration and that in Sado is mountainous with complicated configuration of ground.

Table 7 (by adjusted value)

	data		estimate (in meters)
	distance in level surveyed	k	
1	1,573	0	1,573
2	458	2	424
3	682	0	682
4	475	0	475
5	1,499	2	1,388
6	540	1	529
7	575	1	564
8	1,122	0	1,122
9	458	0	458
10	1,973	7	1,506
11	901	0	901
12	684	7	522
13	446	0	446
14	387	1	379
15	1,321	0	1,321
16	645	0	645
17	1,193	7	911
18	689	2	638
19	857	1	840
mean			807

8. Sampling variance of $\bar{d}$

The estimator $\bar{d}$ of L is defined by $\sum_{i=1}^{n} d_i/n$ where d_i is the length of the ith track and n is the sample size of tracks. In the case where the COC-method is applied, $\bar{d}$ is an unbiased estimate and the sampling variance of $\bar{d}$ is obtained from the usual sampling theory so far as d_1, $d_2, \cdots, d_n$ are regarded as simple random sample from the population. In the case where the RST-method is applied, the problem is to be treated as below. We assume that every starting track to be surveyed is determined independently by equal probability from the track population and the RST-model is completely applicable and the distribution function of tracks L's is constant as a whole in the period of survey even though L_i (length of the ith hare) fluctuates from night to night. Let $d_i'(k)$ (d_i' for simplicity without confusion) be the length of the ith track of survey by the RST-method and d_i be $d_i'/R(k)$ when d_i' has k observed points of random selection.

$$\mathrm{E}\,(d_i)=\mathrm{E}\,(d_i')/R(k)=L\cdot R(k)/R(k)=L$$

where $R(k)$ is the mean value of $\mathrm{E}\,(d')/L$. The variance $\sigma^2_{d_i}$ of d_i is

$$\sigma^2_{d_i}=\mathrm{E}\,(d_i-L)^2=\sigma^2_{d_i'}/R(k)^2$$

where $\sigma^2_{d_i'}$ is obtained by the variance $\sigma(k)^2$ of d'/L in Table 6 multiplied by L^2. Thus the mean value $\bar{d}$ of d, $\bar{d}=\sum_{i=1}^{n} d_i/n$ is clearly unbiased and the variance is readily calculated from the variance $\sigma^2_{d_i}$ of d_i.

Strictly speaking, L (mean value of tracks) also fluctuates from night to night. Let L_t be the mean value of tracks for the tth night. So, we define $L=\sum_{t=1}^{T_2-T_1} L_t/(T_2-T_1)$ and the model III is assumed to be applicable for the tth night $t=1, 2, \cdots, T_2-T_1$. Considering along this idea, it is possible that estimate of L is done as mentioned above in the sense of the estimate of L_t leading to the estimate of it.

Next, we shall examine the assumption that every starting track of the survey is determined independently by equal probability i.e. every track is selected according to the distribution law of track. This assumption may not be possibly fulfilled and that of the selection probability being proportional to the length of the track, may be rather admissible*. Generally speaking, the case where the selection probability is between equal and proportional will be reasonablly recognized, though it is not actually verified. Then the difference between equal case and proportional case may be investigated because the bias is at most that due to the latter case, the former case being unbiased. If $K=0$, the expectation of d is as below in the case where the selection probability is proportional.

* If the track to be surveyed is selected in the way of INTGEP-method, i.e. the first track is selected by the starting point (in the case where it vanishes while being traced the second track is followed), the selection probability may be considered to be as mentioned above. This method is practically adopted.

$$\mathrm{E}(d)=\sum_{i=1}^{N} L_i p_i=\sum_{i=1}^{N} L_i \frac{L_i}{\sum_{i=1}^{N} L_i}=\frac{1}{NL}\sum_{i=1}^{N} L_i^2$$

$$=\frac{1}{NL}\sum_{i=1}^{N}(L_i^2-L^2+L^2)=\frac{\sigma_L^2}{L}+L=L(1+CV_L^2)$$

where p_i is the selection probability of L_i and equal to $L_i\Big/\sum_{i=1}^{N} L_i=L_i/NL$ and CV_L is the coefficient of variation of σ_L/L, $\sigma_L^2=\sum_{i=1}^{N}(L_i-L)^2/N$, $L=\sum_{i=1}^{N} L_i/N$. d is a biased estimate and the relative bias is the square of coefficient of variation. If the CV_L is small the bias becomes negligible. If $k\neq 0$, the bias is examined by computer simulation.

Here, the new model [Model IV] is presented.

[Model IV] Only the starting track is drawn by the probability proportional to the length of the track. Other conditions are quite the same with the model III.

Then we have the following Table 8 corresponding to Table 3. The bias turned out to be at most less than 8.0% for $K\geqq 4$. For $K=0$, theoretically the bias is $CV_L^2=(1/6)^2$. Table 9 is given for the relative bias.

The difference between that of [M IV] and that of [M III], i.e. the bias has the tendency to decrease with K as in Table 9 and this is reasonably recognized.

Using Table 8, we calculate the Bayesian estimate with Table 10 which corresponds to Tables 4 and 5. Then we have Table 11 corresponding to Table 6 on the basis of the same rule.

The difference between those in Table 6 and those in Table 11 of

Table 8 Computer simulation by M IV

K	$\mathrm{E}(d')/L$ ($L=1000$)		observed number of points of intersection	
	mean	standard deviation	mean	standard deviation
1	—	—	—	—
2	—	—	—	—
4	1.298	0.750	4.942	3.104
6	1.431	0.880	8.542	5.527
8	1.493	0.988	12.058	8.118
10	1.551	1.040	15.524	10.678
15	1.720	1.229	26.312	18.959
20	1.759	1.256	35.198	23.393
40	1.853	1.313	74.412	53.221

Table 9 Difference of $\mathrm{E}(d')/L$

K	$\frac{\text{M IV}-\text{M III}}{\text{M III}}$
4	0.066
6	0.053
8	0.030
10	−0.015
15	0.005
20	−0.005

mean, is fairly small as shown in Table 12. So, if we take, in practice, $\bar{d}/(1+B)$ as an estimate instead of $\bar{d}$ from Table 6 where B means degree of bias and is fairly smaller (for example when B is taken as 0.01 because B is evaluated to be at most 0.02) the bias is expected to be sufficiently small.

In conclusion, we can reasonably apply the values in Table 6 or Table 11 except for negligible fluctuation if the selection probability of starting track obeys to some smooth types between equal and proportional, for example proportional to (length)$^{\alpha}$, $1 \geqq \alpha \geqq 0$. The sampling

Table 10

k	E $(l')/L$		variance of d'/L	
	$A_1(6)$	$A_1(10)$	$A_1(6)$	$A_1(10)$
1	1.038	1.052	0.201	0.234
2	1.080	1.110	0.305	0.371
3	1.129	1.178	0.413	0.517
4	1.160	1.224	0.477	0.608
5	1.185	1.264	0.525	0.684
6	1.206	1.298	0.565	0.749
7	1.220	1.323	0.592	0.794
8	1.237	1.348	0.622	0.838
9	1.250	1.368	0.646	0.873
10	1.257	1.383	0.657	0.899
15	1.273	1.405	0.688	0.938
20	1.272	1.427	0.686	0.976

Table 11

k	E $(d')/L$	smoothing adjusted value	variance of d'/L smoothing adjusted value
1	1.04	1.02	0.19
2	1.08	1.09	0.33
3	1.13	1.15	0.46
4	1.20	1.20	0.57
5	1.26	1.25	0.67
6	1.30	1.30	0.75
7	1.32	1.33	0.80
8	1.35	1.35	0.85
9	1.37	1.37	0.88
10	1.38	1.38	0.90
15	1.41	1.41	0.94

Table 12 Difference of mean between Table 6 and Table 11 (smoothed)

k	difference (Table 6−Table 11)
1	0.00
2	−0.01
3	−0.01
4	0.00
5	0.00
6	−0.01
7	−0.02
8	−0.02
9	−0.02
10	−0.02
15	−0.01

variance of $\bar{d}$ in Table 7 is calculated to be 81^2 with the estimated mean 807.

9. Data and the estimation by theory

A large scale hare (Lepus brachyurus angustidens) hunting was done in the western part of Niigata Prefecture. This place is mountainous like our field in Sado Island and has 252 hectares. Ten days before hare hunting, the total length of hare tracks in one night was estimated by the INTGEP. The number of S ($10\,\mathrm{m}\times 2\,\mathrm{m}$) surveyed was 924. The estimated length for one hectare was 340 meters and that in this district was 85,640 meters. Here the estimate of the number was made, based on this estimated length by using the mean running length of a hare during one night $\bar{d}$ in the result of Sado Island as the configuration of the ground in this district shows a similar feature to that in Sado, and furthermore Lepus brachyurus angustidens is same in behavior as Lepus brachyurus lyoni. That is to say, the estimate is $85640/807=106$ where the figure of denominator is shown in the data of Section 7. The result by hare hunting was 96 in number and the estimate gives a desirable goodness of fit with the mean square error, the difference $96-106=-10$ being small if compared with 96.

The estimate of the population size of hare by the INTGEP and RST method seems to be effective. We have had one valid example of this method. Including the COC method, this method must be verified by many examples for hare and other wild animals to which this method is applicable.

10. Some discussions

The method mentioned in Sections 6 and 7 will be also examined by the following idea. The removal method by hare hunting with the INTGEP is applied in an area A (See Section 2). Suppose that the INTGEP is applied about one week before hare hunting and about one week after hare hunting. Let x_B and x_A be the estimated total length of tracks of hares in A at one night, before and after hare hunting respectively. Assume that the length of a hare in one night L is constant before and after that hare hunting and the data under the same environmental and weather conditions are obtained in the before- and after-survey. By hare hunting M hares are removed, N_B being the initial number of hare in A. Then we have, except for the sampling variance,

$$x_B=LN_B \qquad x_A=L(N_B-M)\ .$$

So

$$N_B = \frac{M}{(1-(x_A/x_B))}$$

holds. From this, we estimate the total number N by $\hat{N}_B = M/\{1-(x_A/x_B)\}$, the mean square of which is easily calculated because x_A and x_B are independent random variables obtained from random sampling survey. Also we can estimate L by $x_B/\hat{N}_B$ or $x_A/(\hat{N}_B - M)$, which is called the THH (length of Track estimated by Hare Hunting) method.

Here, five cases will be discussed comparing the estimate by the THH method with that by the COC method and the RST method. The estimate by each method is called the THH, COC and RST in abbreviation and the sign $\neq$ is used where the difference between them exceeds the sampling variability and the sign $=$ is used when the difference is with in the sampling variability.

(i) THH=COC≠RST
In this case, the mathematical assumption in RST method must be revised.

(ii) THH=RST≠COC
In this case, the collar and the catch by trap may influence the movement of hare and only particular individuals may be caught alive.

(iii) THH≠RST=COC
In this case, the assumption that L is constant before and after hare hunting must be revised. The event of hare hunting and the in-going or out-going of hares by disturbance of hunting or the decrease in number may influence the movement of hares.

(iv) THH=COC=RST
In this case, the methods are suitable and the problem will have been solved.

(v) THH≠COC≠RST
In this case, the methods must thoroughly be revised. The data check by hare hunting as in Section 9 is expected and the most suitable method whether COC or RST must be determined.

If the most suitable method is determined by hare hunting as in Section 9, the characteristics of other methods will be investigated comparing with that in this method. Thus, we shall have much more information on L and ecological knowledge on hare.

Acknowledgement

The authors express hearty thanks to their colleagues in the project for their cooperative research works, Dr. M. Isida, in the Institute of Statistical Mathematics, Prof. J. Toyoshima and Prof. K. Takata in Faculty of Agriculture, Niigata University, Dr. M. Ueda and Dr. Y. Shibata in the Experimental Station of Forestry, Hokkaido Branch and Dr. T. Niwaguchi, in the National Research Institute of Police Science, Tokyo. Also they would like to thank Prof. M. Morisita, Faculty of Science, Kyoto University, for his valuable suggestions who is an organizer of methodological group of population counting in JIBP-PT and to thank the referee for his useful comments on the careful reading of this paper.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Hayashi, C. et al. (1966). Estimation of size of mobile population-I —Sample survey for the estimation of the size of hares in a district—, *Proc. Inst. Statist. Math.*, **14**, 63-86.

[2] Hayashi, C. et al. (1968). Estimation of size of mobile population-II —Statistical survey of wild rabbits in an islet with revision of capture-recapture method—, *Proc. Inst. Statist. Math.*, **16**, 121-132, JIBP-PT No. 53.

[3] Toyoshima, J. et al. (1973). Ecological studies of hares (Lepus brachyurus angustidens Hollister) 2. Frequency of migration between 4 PM and 8 AM as to hares, *Research Report of Agriculture and Forestry*, No. 25, Niigata University.

[4] Hayashi, C. et al. (1969). Estimation of size of mobile population-III —Sample survey for the estimation of the size of hare (Lepus brachyurus lyoni Kishida) population in the Island of Sado in Niigata Prefecture—, *Proc. Inst. Statist. Math.*, **17**, 5-21, JIBP-PT No. 61.

[5] Hayashi, C. et al. (1972). Estimation of the size of mobile population-VII —Field survey of a running distance of a hare in one night—, *Proc. Inst. Statist. Math.*, **20**, 45-60, JIBP-PT No. 162.

[6] Hayashi, C. (1973). One method to count the number of hares in an area, *Hoppo-Ringyo*—Northern Forestry—(Sapporo), No. 292.

[7] Hayashi, C. et al. (1971). Estimation of the size of mobile population-V —Estimation of trace length of a hare in one night by randomly selected trace (RST) method and computer simulation—, *Proc. Inst. Statist. Math.*, **19**, 14-27, JIBP-PT No. 125.

[8] Hayashi, C. et al. (1972). Estimation of the size of mobile population-VIII —Bayesian estimation in RST-method based on the observed number of randomly selected points—, *Proc. Inst. Statist. Math.*, **20**, 117-120, JIBP-PT No. 163.

[9] Hayashi, C. (1969). One population-dynamics-method for estimating of population size of hare in a district —by quantifying the amount of damage in forest—, *Research Report of Hare Research Group ISM & NIA*, No. 3.

[10] Cormack, R. M. (1966). A test for equal catchability, *Biometrics*, **22**, 330-342.

N.B. The papers of the well-known capture-recapture method are in the appendix of [1]. The interesting discussion for the method is described in [10]. The general discussions on various capture-probability models are found in the paper by D. L. Otis, K. P. Brunham, G. C. White and D. P. Anderson, "Statistical Inference from Capture Data and Closed Animal Population, Wildlife Monographs No. 62, October 1978". Only the theoretical discussion has been published in the paper by C. Hayashi, "A New Statistical Approach to Estimating the Size of an Animal Population: The Case of a Hare Population, Math. Scientist, 1978, 3, 117-130".

(April, 1979)