AD-A052 319 FLORIDA STATE UNIV TALLAHASSEE DEPT OF STATISTICS F/G 12/1
STOCHASTIC MODELS FOR THE RANDOM LOCATION OF INDIVIDUALS IN A H--ETC(U)
MAR 78 N A LANGBERG, R JOHNSON, R A BRADLEY N00014-76-C-0608
UNCLASSIFIED FSU-STATISTICS-M455 NL

| OF |

AD
A052319

END
DATE
FILMED
5-78
DDC

The Florida State University

Department of Statistics

Tallahassee, Florida
32306

FLORIDA · STATE · UNIVERSITY ·
1857

(11)

STOCHASTIC MODELS FOR THE RANDOM LOCATION OF INDIVIDUALS IN A HABITAT

by

Naftali A. Langberg, Roger Johnson, and Ralph A. Bradley

FSU Statistics Report M455
ONR Technical Report No. 132

March, 1978
The Florida State University
Department of Statistics
Tallahassee, Florida 32306

Research supported by the Army, Navy and Air Force under Office of Naval Research Contract No. N00014-76-C-0608.

Stochastic Models for the Random Location of Individuals in a Habitat

by

Naftali A. Langberg, Roger Johnson, and Ralph A. Bradley

ABSTRACT

Very little had been done in the past to develop coherent stochastic structures for the location of individuals in a region. Such a structure can provide sound probabilistic bases for a variety of inference procedures presented in the literature and clear the way to the development of consistent statistical methodology. Various probabilistic models for the random locations of individuals in a region are considered.

Models for both infinite and finite regions are developed. In the literature, certain marginal distributions are used to characterize the numbers of individuals in subregions. Relationships between these marginal distributions and our probabilistic models are discussed.

Key words: Stochastic processes, stationarity, independent increments, Poisson Process, Poisson, Binomial and Negative Binomial distributions.

1. Introduction and Summary.

Trees in a forest, stars in a galaxy and bacteria on a Petridish are a few examples wherein there is interest in the spatial distributions of individuals over regions. In pursuit of better understanding of these distributions, attempts have been made to use probabilistic and statistical tools [Eberhardt (1967), Pilou (1969), Pollard (1971), Patil and Stiteler (1974), Cox (1976), Cox and Lewis (1976)]. Very little effort has been made to develop coherent statistical structures for the locations of individuals in a region. The main purpose of this paper is to propose such structures in order to avoid inconsistencies that can be found in the literature, to provide a sound probabilistic basis for a variety of inference procedures that are used, and to permit the development of better and consistent statistical methodology in the area.

In Sections 2, 3 and 4, we show that, under certain conditions, the collection of random variables representing the numbers of individuals in subsets of a region, can be regarded as a stochastic process. All the models for the stochastic process presented here have a stationarity property defined in (2.1) below. In Section 2 we present Simple Random Models. The distribution of the number of individuals in a region is considered to be simple random if, in addition to stationarity, it has a property of independent increments defined in (2.2) below. We prove that simple random models are uniquely determined by a positive number and a sequence of real numbers in [0, 1] summing to unity. In a number of the references cited above, Poisson and Negative Binomial distributions have been used to describe the number of individuals in subsets of a region. We conclude Section 2 by showing that these two distributions are marginal distributions resulting from particular simple random models.

In Section 3 we introduce models that are stationary, but do not necessarily have independent increments. We show that a Negative Binomial distribution proposed by Patil and Stiteler (1974) to characterize the random behavior of the numbers of individuals in subsets of a region is a marginal distribution resulting from two different stochastic models that do not have independent increments.

Section 4 is devoted to the development of several different stochastic models when the region under consideration has finite Lebesgue measure. It is shown that the Binomial distribution suggested by Eberhardt (1967) can be derived from all of these models.

2. Simple Random Models.

Some basic notation is needed to define spatial stochastic models. Throughout the paper, R will denote a Euclidian space or a Borel subset of a Euclidian space with finite Lebesgue measure, λ will be the Lebesgue measure on R, X(A) will be the number of individuals in a subset $A \varepsilon R$, and $\sum$ will represent the set of all Borel subsets of R with finite Lebesgue measures.

Individuals are said to be located in R in a simple random way if defined propositions (2.1) and (2.2) hold.

Stationarity: For every set $A \varepsilon \sum$, the distribution of the random variable X(A) depends only on $\lambda(A)$. (2.1)

Independent increments: If $A_1, \ldots, A_m$ represent any m disjoint sets in $\sum$, $m > 1$, the random variables, $X(A_1), \ldots, X(A_m)$ are independent. (2.2)

The objective of this section is to show that under conditions (2.1) and (2.2), the collection, $\{X(A)\}_{A\varepsilon\sum}$ is a stochastic process determined by a positive number and a sequence of numbers in [0, 1] summing to unity.

Let q be an arbitrary point in R, $Y_q(t)$, the number of individuals in a sphere centered at q having Lebesgue measure t, $t \geq 0$, $Y_q(0) \equiv 0$, A_1, A_2, sets in $\sum$, and k_1, k_2 two nonnegative integers. Then

$$P[X(A_1) = k_1, X(A_2) = k_2] = \sum_{\substack{j+r=k_1 \\ j+\ell=k_2}} P[X(A_1 \cap A_2) = j, X(A_1 \cap \bar{A}_2) = r, X(\bar{A}_1 \cap A_2) = \ell]. \quad (2.3)$$

From (2.1), (2.2) and (2.3) it follows that

$$P[X(A_1) = k_1, X(A_2) = k_2] = \sum_{\substack{j+r=k_1 \\ j+\ell=k_2}} P[Y_q(t_1) = j]\, P[Y_q(t_2) = r]\, P[Y_q(t_3) = \ell], \quad (2.4)$$

where t_1, t_2 and t_3 are $\lambda(A_1 \cap A_2)$, $\lambda(A_1 \cap \overline{A}_2)$ and $\lambda(\overline{A}_1 \cap A_2)$ respectively. Without difficulty, equations (2.3) and (2.4) can be extended to any finite number of sets in $\sum$. The formal proof is omitted but the general result is stated in the following theorem.

Theorem 2.1. Under conditions (2.1) and (2.2), the joint distribution of any finite collection of random variables from $\{X(A)\}_{A\in\sum}$ is determined uniquely by a joint distribution of a finite collection of random variables from $\{Y_q(t)\}_{t\geq 0}$.

To achieve our objective, it suffices to show that, under (2.1) and (2.2), the collection $\{Y_q(t)\}_{t\geq 0}$ is a stochastic process determined by a positive number and a probabilistic sequence.

Conditions (2.1) and (2.2) respectively insure that $\{Y_q(t)\}_{t\geq 0}$ is stationary and has independent increments. We state formally:
For every two nonnegative numbers t and s, $Y_q(t)$ and $Y_q(t + s) - Y_q(s)$ are identically distributed. For every m real numbers $0 \leq t_1 \leq t_2 \leq \ldots \leq t_m$, the random variables $Y_q(t_1)$, $Y_q(t_2) - Y_q(t_1)$, ..., $Y_q(t_m) - Y_q(t_{m-1})$ are independent. Applying a well known characterization of stationary processes with independent increments, [Khintchine (1960), 36.], we obtain the following:

Theorem 2.2. Under conditions (2.1) and (2.2), there exist a Poisson process $\{N(t)\}_{t\geq 0}$ and a sequence of i.i.d. random variables $\{Z_i\}_{i=1,2,\ldots}$ on the positive integers, independent of $\{N(t)\}_{t\geq 0}$, such that the processes $\{Y_q(t)\}_{t\geq 0}$ and $\{\sum_{i=1}^{N(t)} Z_i\}_{t\geq 0}$ are identical.

It is a consequence of Theorem 2.2 that, the collection $\{Y_q(t)\}_{t\geq 0}$ is a stochastic process, determined by the constant $E[N(1)]$ and the sequence $\{P(Z_1 = i)\}_{i=1,2,\ldots}$.

Two marginal distributions for $X(A)$, used frequently in the literature, arise from special cases of simple random models; they are the Poisson and the Negative Binomial. To show it, we give the two definitions:

X has a Negative Binomial distribution with parameters a and b, $(a,\ b > 0)$ if $P(X = i) = \dfrac{\Gamma(a + i)}{i!\Gamma(a)} \left(\dfrac{a}{a + b}\right)^a \left(\dfrac{b}{a + b}\right)^i$, $i = 0, 1, 2, \ldots$, where $\Gamma(\alpha) = \int_0^\infty e^{-x}x^{\alpha-1}dx$, $\alpha > 0$.

X has a logarithmic distribution with parameter θ $(0 < \theta < 1)$, if $P(X = i) = -\theta^i/i\ell n(1 - \theta)$, $i = 1, 2, \ldots$.

If Z_1 of Theorem 2.2 is a degenerate random variable at 1, or has the logarithmic distribution defined above, it follows from Theorem 2.2 that, for every set $A \in \sum$, the distribution of $X(A)$ is respectively Poisson, with mean $\lambda(A)\ E[N(1)]$, or Negative Binomial with parameters $a = \dfrac{E[N(1)]}{\ell n(1 - \theta)} \lambda(A)$ and $b = \dfrac{E[N(1)]}{\ell n(1 - \theta)} \cdot \dfrac{\theta}{1 - \theta} \lambda(A)$.

3. A Generalization of the Simple Random Models.

In this section we present two methods of developing a general stationary stochastic structure for the random quantities $\{X(A)\}_{A\varepsilon\sum}$. This is done in two steps. Firstly, we present joint distributions of $X(A_1), \ldots, X(A_m)$, where $A_1, \ldots, A_m$ are any m disjoint sets of $\sum$ and $m = 1, 2, \ldots$. We then show that this set of finite dimensional distributions determine uniquely the stochastic structure of $\{X(A)\}_{A\varepsilon\sum}$.

For construction purposes only we introduce some notation. $\{M(t)\}_{t\geq 0}$ to be a stochastic process on the nonnegative integers, stochastically increasing in t, that is, $M(t) \leq M(s)$ for $0 \leq t \leq s$ and $\{W_i\}_{i=1,2,\ldots}$ a sequence of random variables with positive integer values, independent of $\{M(t)\}_{t\geq 0}$. We assume that $\{M(t)\}_{t\geq 0}$ and $\{W_i\}_{i=1,2,\ldots}$ depend on parameters θ_1 and θ_2 respectively, ranging in parameter spaces Θ_1 and Θ_2, where Θ_2 is a collection of infinite sequences, Θ_1 and Θ_2 being probability spaces with F and G the respective probability measures.

Let $A_1, \ldots, A_m$ be disjoint sets in $\sum$ with respective Lebesgue measures $t_1, \ldots, t_m$ and let $k_1, \ldots, k_m$ be nonnegative integers. We define the joint probability in two ways.

$$P[X(A_i) = k_i,\ i = 1, \ldots, m] = \int_{\Theta_1\times\Theta_2}\int \prod_{i=1}^{m} P[\sum_{j=1}^{M(t_i)} W_j = k_i]dF(\theta_1)dG(\theta_2) \tag{3.1}$$

or

$$P[X(A_i) = k_i,\ i = 1, \ldots, m] = \int_{\Theta_1\times\Theta_2}\int P[\sum_{j=1}^{M(t_i)} W_j = k_i,\ i = 1, \ldots, m]dF(\theta_1)dG(\theta_2). \tag{3.2}$$

The consistency of the set of joint distributions of finite collections of random variables from $\{X(A)\}_{A\varepsilon\sum}$, generated by disjoint sets and defined by (3.1) or (3.2), follows from the stochastic structure imposed on $< \{M(t)\}_{t\geq 0}, \{W_i\}_{i=1,2,\ldots} >$. The stationarity condition, stated in (2.1), is clearly satisfied, since definitions (3.1) and (3.2) vary only with the Lebesgue measures of the respective sets. The generalization of equation (2.3) to any finite number of sets in $\sum$ provides a way to extend definitions (3.1) and (3.2) to any finite number of sets in $\sum$. These extensions preserve the required consistency and stationarity of $\{X(A)\}_{A\varepsilon\sum}$. We have proved the following:

<u>Theorem 3.1</u>. If the joint distributions of every finite number of random variables in $\{X(A)\}_{A\varepsilon\sum}$ are given by the extensions of (3.1) or (3.2), then $\{X(A)\}_{A\varepsilon\sum}$ is a stationary stochastic process.

It is important to note that (3.1) and (3.2) yield the same marginal distributions for $X(A)$, $A \varepsilon \sum$, but they define different processes. Statistical methodologies based on the two models may be quite different and this has been ignored in the literature.

In Section 2, we proved that simple random models depend on a positive number say μ, and a probabilistic sequence, say $\{P_i\}_{i=1,2,\ldots}$. Now we show that those models are particular cases of the stochastic processes presented by (3.2). To do so, let $\{M(t)\}_{t\geq 0}$ be a Poisson process with parameter μ, and $\{W_i\}_{i=1,2,\ldots}$ be an i.i.d. sequence of random variables given by $P(W_1 = i) = P_i$, $i = 1, 2, \ldots$. In addition we take Θ_1 and Θ_2 to be sets containing only μ and the sequence $\{P_i\}_{i=1,2,\ldots}$ respectively. Now the extension of (3.2) reduces directly to a simple random model.

Let's assume that the sequence of discrete random variables $\{W_i\}_{i=1,2,...}$ is degenerate at 1, Θ_1 is $[0, \infty)$, Θ_2 is only the sequence (1, 0, 0, ...), $\{M(t)\}_{t\geq 0}$ is Poisson, and F is given by

$$F(x) = \begin{cases} \dfrac{\beta^\alpha}{\Gamma(\alpha)} \displaystyle\int_0^x e^{-\beta y} y^{\alpha-1} dy & x > 0 \\ 0 & x \leq 0 \end{cases} \qquad \alpha, \beta > 0. \tag{3.3}$$

Then the marginal distribution of X(A), A ε $\sum$, computed by (3.1) or (3.2), is negative Binomial with $a = \alpha$ and $b = \frac{\alpha}{\beta} \lambda(A)$. This marginal distribution for X(A) was assumed by Eberhardt (1967) and by Patil and Stiteler (1974).

4. Models for finite regions.

In the previous two sections, we presented stochastic models when individuals were randomly located in an infinite region. In this section, five different models are developed for situations wherein individuals are located in a Borel subset R_f of a finite Lebesgue measure in R.

Let $\sum_f$ be the set of all Borel sets contained in R_f and $X_f(A)$, the number of individuals in set A, $A \in \sum_f$. Our objective is to develop a stochastic structure for the collection $\{X_f(A)\}_{A\in\sum_f}$. One way to acheive this objective is to select a stochastic process developed in Section 3 for the infinite region R, say $\{X(A)\}_{A\in\sum}$, and condition it by an event related to the random variable $X(R_f)$. To be more formal, let $A_1, \ldots, A_m$ be sets in $\sum_f$ and let $k_1, \ldots, k_m$ be nonnegative integers. We define the desired probability

$$P[X_f(A_i) = k_i,\ i = 1, \ldots, m] = P[X(A_i) = k_i,\ i = 1, \ldots, m | X(R_f)]. \tag{4.1}$$

The consistency and stationarity of the process $\{X_f(A)\}_{A\in\sum_f}$ defined in (4.1) are self-evident.

For the four remaining alternative ways of developing a stochastic structure for $\{X_f(A)\}_{A\in\sum_f}$, we refer to the process $\{M(t)\}_{t\geq 0}$, the sequence $\{W_i\}_{i = 1, 2, \ldots}$, the sets Θ_1 and Θ_2, and the probability measures F and G which were introduced in Section 3. We define first the joint distributions of $X_f(A_1), \ldots, X_f(A_m)$, where $A_1, \ldots, A_m$ are disjoint sets in $\sum_f$ and $m = 1, 2, \ldots,$ then extend these definitions to any finite collection of sets in $\sum_f$. Since the extension technique has been used twice, it is omitted from our present discussion. Now let $A_1, \ldots, A_m$ be disjoint subsets of R_f,

$\lambda(R_f) = t_0$, $\lambda(A_i) = t_i$, $i = 1, \ldots, m$, and let $k_1, \ldots, k_m$ be nonnegative integers. We define the desired probability in the following four ways:

$$P[X_f(A_i) = k_i,\ i = 1, \ldots, m] = \int\limits_{\Theta_1 \times \Theta_2}\int \prod_{i=1}^{m} P\Big(\sum_{j=1}^{m(t_i)} W_j = k_i \Big| \sum_{j=1}^{M(t_0)} W_j\Big) dF(\theta_1) dG(\theta_2). \quad (4.2)$$

$$P[X_f(A_i) = k_i,\ i = 1, \ldots, m] = \int\limits_{\Theta_1 \times \Theta_2}\int P\Big(\sum_{j=1}^{M(t_i)} W_j = k_i,\ i = 1, \ldots, m \Big| \sum_{j=1}^{M(t_0)} W_j\Big) dF(\theta_1) dG(\theta_2). \quad (4.3)$$

$$P[X_f(A_i) = k_i,\ i = 1, \ldots, m] = \int\limits_{\Theta_1 \times \Theta_2}\int \prod_{i=1}^{m} P\Big(\sum_{j=1}^{M(t_i)} W_j = k_i \Big| M(t_0)\Big) dF(\theta_1) dG(\theta_2). \quad (4.4)$$

$$P[X_f(A_i) = k_i,\ i = 1, \ldots, m] = \int\limits_{\Theta_1 \times \Theta_2}\int P\Big(\sum_{j=1}^{M(t_i)} W_j = k_i,\ i = 1, \ldots, m \Big| M(t_0)\Big) dF(\theta_1) dG(\theta_2). \quad (4.5)$$

The stochastic processes determined by (4.2)-(4.5) have been constructed to be both consistent and stationary. Definitions (4.2) and (4.3) or (4.4) and (4.5), yeild identical marginal distributions for $X_f(A)$, $A \in \sum_f$, but define different processes.

If we let $\{M(t)\}_{t \geq 0}$ be a Poisson process, $\{W_i\}_{i = 1, 2, \ldots}$, a sequence of random variables degenerate at 1, Θ_1 and Θ_2 singleton sets, then the marginal distribution of $X_f(A)$, $A \in \sum_f$ according to each of the last four definitions are Binomial with $X(R_f)$ corresponding to the number of Bernulli trials and $\frac{\lambda(A)}{\lambda(R_f)}$ corresponding to the probability of succes in a single trial.

5. Concluding Remarks.

The selection of models for stochastic processes for particular applications and the development of pertinent statistical methodologies have not been addressed in this paper. We have demonstrated that models proposed may be used to yield marginal distributions assumed in the literature; they may also be used to avoid unwarranted assumptions and inconsistencies that arise. We propose in subsequent work to use the general models of this paper to devise improved statistical methodologies for problems involving the location of individuals in a habitat.

REFERENCES

[1] Cox, T. F. (1976). Using new conditional distance method. Biometrika, 63, 493-497.

[2] Cox T. F. and Lewis T. (1976). A conditional distance ratio method for analyzing spatial patterns. Biometrika, 63, 483-491.

[3] Eberhardt, L. L. (1967). Some developments in distance sampling. Biometrika, 23, 207-216.

[4] Khintchine, A. Y. (1960). Mathematical Methods in the Theory of Queueing, Hafner Publishing Company, New York.

[5] Patil G. P. and W. M. Stitler (1974). Concepts of aggregation and their quantification: a critical review with some new results and applications. Res. Popul. Ecol., 15 238-253.

[6] Pielu, E. C. (1969). An Introduction to Mathematical Ecology, John Wiley and Sons, New York.

[7] Pollard, J. H. (1971). On distance estimators of density in randomly distributed forests. Biometrika, 27, 991-1002.

(14) FSU-STATISTICS-M455
TR-132-ONR

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1. REPORT NUMBER FSU No. M455 ONR No. 132	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. (6) TITLE Stochastic Models for the Random Location of Individuals in a Habitat.		5. TYPE OF REPORT & PERIOD COVERED (9) Technical Rep[redacted]rt, (10) 6. ~~PERFORMING ORG. REPORT NUMBER~~ FSU Statistics Report M455
7. AUTHOR(s) Naftali A. Langberg Roger Johnson Ralph A. Bradley		8. CONTRACT OR GRANT NUMBER(s) ONR No. (15) N00014-76-C-0608
9. PERFORMING ORGANIZATION NAME & ADDRESS The Florida State University Department of Statistics Tallahassee, Florida 32306		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (12) 16p.
11. CONTROLLING OFFICE NAME & ADDRESS Office of Naval Research Statistics and Probability Program Arlington, Virginia 22217		12. REPORT DATE (11) Mar[redacted]78 13. ~~NUMBER OF PAGES~~ 12
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS

Stochastic processes, stationarity, independent increments, Poisson Process, Poisson, Binomial and Negative Binomial distributions.

20. ABSTRACT

Very little had been done in the past to develop coherent stochastic structures for the location of individuals in a region. Such a structure can provide sound probabilistic bases for a variety of inference procedures presented in the literature and clear the way to the development of consistent statistical methodology. Various probabilistic models for the random locations of individuals in a region are considered.

Models for both infinite and finite regions are developed. In the literature, certain marginal distributions are used to characterize the numbers of individuals in subregions. Relationships between these marginal distributions and our probabilistic models are discussed.

400 277