Wang Zikun Yang Xiangqun

Birth and Death Processes and Markov Chains

Springer-Verlag

Science Press

Wang Zikun Yang Xiangqun

Birth and Death Processes and Markov Chains

Springer-Verlag
Berlin Heidelberg New York
London Paris Tokyo
Hong Kong Barcelona
Budapest

Science Press
Beijing

Wang Zikun
Beijing Normal University
Beijing
The People's Republic of China

Yang Xiangqun
Xiangtan University
Xiangtan, Hunan province
The People's Republic of China

Revised edition of the original Chinese edition published by Science Press Beijing 1980 as the fifth volume in the Series in Pure and Applied Mathematics.

Distribution rights throughout the world, excluding the People's Republic of China, granted to Springer-Verlag Berlin Heidelberg New York London Paris Tokyo Hong Kong Barcelona Budapest

Mathematics Subject Classification (1980): 60A-XX 60G-XX 60J-XX

ISBN 3-540-10820-3 Springer-Verlag Berlin Heidelberg New York
ISBN 0-387-10820-3 Springer-Verlag New York Berlin Heidelberg
ISBN 7-03-002250-5 /O • 422 Science Press Beijing

Typesetting: Science Press, Beijing. The People's Republic of China
Printing: C&C Joint Printing co., (H. K.) Ltd.
Binding: K. Triltsch, Würzburg
41/3140-543210

Preface

The objective of this book is to describe the fundamental theory of birth-death processes and Markov chains and to present the developments in this field in recent years. The so-called Markov chain here refers to a Markov process which has continuous time parameters, countably many states and is time-homogeneous. Chains of this kind are important not only because its comparatively complete and unified theory can be used for reference in general Markov chains and other stochastic processes, but also because there is a steady increase in the applications to natural sciences and practical problems such as physics, biology, chemistry, programming theory and queueing theory. For these the reader is referred to the works by K. L. Chung, Hou Zhen-ting, Guo Qing-feng, Bharucha-Reid, quoted at the end of this book.

Birth and death processes belong to a particular kind of the Markov chain. Despite the abundance of relevant works and reference materials, it seems that thus far there has been no systematic monograph available to expound them. Some outstanding scholars such as D. G. Kendall, G. E. H. Reuter and W. Feller, especially S. Karlin and J. McGregor, have done a great deal of thorough and important study in this field. They have, in general, resorted to the methods of analytic mathematics. The authors regret that they have not been able to make an extensive study of this area. However, they have solved certain problems encountered. Probabilistic methods are mainly used in Chapters V and VI to construct all the birth and death processes. Starting from the investigation of the trajectory of a motion, an intuitive form is derived and then paved with vigorous proofs in mathematical computation and measure theory. The advantage of this method is that the probabilistic meaning is fairly clear although the whole procedure may seem a little lengthy. In Chapters VII and VIII all the birth-death processes and bilateral birth-death processes are constructed mainly via analysis.

Chapter I is preliminary. Chapters II and III deal with the analytic properties and the trajectory behaviours of Markov chains which are mainly due to K. L. Chung, R. L. Dobrushin, J. L. Doob, W. Feller, A. N. Kolmogorov and P. Lévy. Chapter IV discusses some special topics and Chapters V to VIII deal with birth and death processes. These last five chapters are basically achievements of recent research made by Chinese scholars, including the authors. For details, see Annotations on the History of the Contents of Each Section after Appendix II.

The authors wish to express their sincere gratitude to Professor R. L. Dobrushin, who motivated their interest in birth and death processes. They are also indebted to professors Wu Rong, Liu Wen and Yang Zhen-ming, for reading over the manuscripts carefully and offering many valuable suggestions, particularly to Zhang Run-chu and Zhang Shu-dong for their tremendous work in preparing this English version.

Wang Zikun
Yang Xiangqun

November 1991

CONTENTS

Chapter I

General Concepts of Stochastic Processes

§1.1 Definition of Stochastic Processes

(I) Probability spaces. Let $\Omega=(\omega)$ be a space of points ω and $\mathfrak{F}$ a class of some subsets A of Ω . $\mathfrak{F}$ is called a *σ-algebra* if it possesses the following properties:

1) $\Omega\in\mathfrak{F}$.

2) If $A\in\mathfrak{F}$, then $\overline{A}=\Omega\setminus A\in\mathfrak{F}$.

3) If $A_n\in\mathfrak{F}$, $n=1,2,\cdots$, then $\bigcup_{n=1}^{\infty}A_n\in\mathfrak{F}$.

A set function P defined on the σ-algebra $\mathfrak{F}$ is called a *probability* if P satisfies the following conditions:

1° $P(A)\geq 0$ for every $A\in\mathfrak{F}$.

2° $P(\Omega)=1$.

3° If $A_n\in\mathfrak{F}$, $n=1,2,\cdots$, $A_nA_m=\phi$, $m\neq n$(ϕ denoting the empty set), then

$$P(\bigcup_{n=1}^{\infty}A_n)=\sum_{n=1}^{\infty}P(A_n).$$

We call the triple $(\Omega,\mathfrak{F},P)$ a *probability space*, a point ω in Ω an *elementary event*, Ω the *space of elementary events*, a set A in $\mathfrak{F}$ an *event*, and $P(A)$ *the probability of* A .

Example 1. Assume that $\Omega=(1,2,\cdots,n)$, $\mathfrak{F}$ is the collection of all the subsets of Ω, and $p(A)=\dfrac{k}{n}$, where k is the number of the points in A.

Example 2. Assume that $\Omega=(0,1,2,\cdots)$, i.e., the set of all nonnegative integers, $\mathfrak{F}$ is the collection of all the subsets of Ω, and $P(A)=\sum_{k\in A}\dfrac{\lambda^k}{k!}e^{-\lambda}$, where $\lambda>0$ is a constant.

Example 3. Assume that $\Omega=[0,1]$, i.e., the set of all numbers between 0 and 1, $\mathfrak{F}$ refers to the σ-algebra consisting of all the Borel sets of Ω, and $P(A)$ is equal to the Lebesgue measure of A.

All the $(\Omega,\mathfrak{F},P)$ in the above three examples are probability spaces.

Sometimes for convenience, we need to suppose that the probability space $(\Omega,\mathfrak{F},P)$ is complete. By "complete" here we mean that if $P(A)=0$ and $B\subset A$, then $B\in\mathfrak{F}$ and hence $P(B)=0$. That is to say, each subset B contained in the set A with probability 0 is also an event with probability 0. Unless other-

wise stated hereafter, this condition is always assumed to be satisfied.

(II) Random variables. Let $x(\omega)$ be a real-valued function defined on Ω. If for each real number λ, we have

$$(\omega: x(\omega) \leq \lambda) \in \mathfrak{F},$$

then $x(\omega)$ is called a *random variable*. Put

$$F(\lambda)=P(x \leq \lambda), \quad \lambda \in R_1=(-\infty,\infty), \tag{1}$$

where $(x \leq \lambda)$ represents the set of points ω satisfying the condition $x(\omega) \leq \lambda$, namely, $(x \leq \lambda)=(\omega: x(\omega) \leq \lambda)$. We call $F(\lambda)$ the *distribution function* of $x(\omega)$. Clearly, $F(\lambda)$ is nondecreasing and right continuous. From now on unless otherwise stated, the probability of the event $(\omega: x(\omega)=\pm\infty)$ always denotes zero, hence

$$\lim_{\lambda \to -\infty} F(\lambda)=0, \quad \lim_{\lambda \to \infty} F(\lambda)=1.$$

The n random variables $x_1(\omega), \cdots, x_n(\omega)$ defined on the same probability space $(\Omega,\mathfrak{F},P)$ form an n-dimensional *random vector* $X(\omega)$:

$$X(\omega)=(x_1(\omega), \cdots, x_n(\omega)), \tag{2}$$

and the function of n variables $(\lambda_1, \cdots, \lambda_n) \in R_n$ (the n-dimensional real space)

$$F(\lambda_1, \cdots, \lambda_n)=P(x_1(\omega) \leq \lambda_1, \cdots, x_n(\omega) \leq \lambda_n) \tag{3}$$

is called the *n-dimensional distribution function* of $X(\omega)$. From (3), it can be seen that $F(\lambda_1, \cdots, \lambda_n)$ possesses the following properties:

a. For each λ_j, it is a nondecreasing and right continuous function.

b. $\lim_{\lambda_j \to -\infty} F(\lambda_1, \cdots, \lambda_n)=0 \ (j=1, \cdots, n)$,

$$\lim_{\lambda_1, \cdots \lambda_n \to \infty} F(\lambda_1, \cdots, \lambda_n)=1.$$

c. If $\lambda_j < \mu_j$, $j=1, \cdots, n$, then

$$F(\mu_1, \cdots, \mu_n)-\sum_{j=1}^{n} F(\mu_1, \cdots, \mu_{j-1}, \lambda_j, \mu_{j+1}, \cdots, \mu_n)$$
$$+\sum_{j,k=1}^{n} F(\mu_1, \cdots, \mu_{j-1}, \lambda_j, \mu_{j+1}, \cdots, \mu_{k-1}, \lambda_k, \mu_{k+1}, \cdots, \mu_n)$$
$$-\cdots+(-1)^n F(\lambda_1, \cdots, \lambda_n) \geq 0.$$

(The intuitive meaning of this condition is clear for $n=2$. In general, the right member of the above formula represents the probability for $x(\omega)$ to take values in the cuboid of the n-dimensional space R_n, and hence it is greater than or equal to zero; the cuboid is $(\lambda_1,\mu_1] \times (\lambda_2,\mu_2] \times \cdots \times (\lambda_n,\mu_n]$, i.e. the set of such points in R_n that their j-th coordinates are situated in $(\lambda_j,\mu_j], j=1, \cdots, n$).

Now we can define a distribution function without relation to random variables. A function $F(\lambda_1, \cdots, \lambda_n)(\lambda_j \in R_1, j=1, \cdots, n)$ of n variables with properties a, b and c, is called a *distribution function of n variables*. $\mathfrak{B}_n$ denotes the σ-algebra consisting of all the Borel sets of the n-dimensional space R_n,

then from measure theory we know that $F(\lambda_1,\cdots,\lambda_n)$ generates a probability measure $F(A)$ on $\mathfrak{B}_n$:

$$F(A)=\int_A dF(\lambda_1,\cdots,\lambda_n) \quad (A\in\mathfrak{B}_n).$$

$F(A),(A\in\mathfrak{B}_n)$ is called the *n-dimensional distribution* generated by $F(\lambda_1,\cdots,\lambda_n)$. In particular, if $F(\lambda_1,\cdots,\lambda_n)$ is generated by (3), then $F(A)$ is called the distribution of $x(\omega)$.

(III) Stochastic processes. Let T be a subset of R_1, for instance, $T=[0,\infty)$ or $T=(0,1,2,\cdots)$. If for each $t\in T$ there corresponds a random variable $x_t(\omega)$, then the collection $X(\omega)$ of random variables

$$X(\omega)=\{x_t(\omega), t\in T\}$$

is called a *stochastic process*, or simply a *process*. Sometimes it is written as $\{x(t,\omega), t\in T\}$, or $\{x_t, t\in T\}$, or $\{x(t), t\in T\}$, or $X(\omega)$, or X.

In particular, X is reduced to an n-dimensional random vector for $T=(1,2,\cdots,n)$. Just as we defined the distribution function for X, we can also define finite-dimensional distribution functions of a stochastic process. For any $t_j\in T$, $j=1,\cdots,n$, put

$$F_{t_1,\cdots,t_n}(\lambda_1,\cdots,\lambda_n)=P(x_{t_1}\le\lambda_1,\cdots,x_{t_n}\le\lambda_n), \tag{4}$$

which is the distribution function of $x_{t_1}(\omega),\cdots,x_{t_n}(\omega)$. As n varies in all positive integers and t_j varies in T, the collection of distribution functions of several variables is obtained as follows:

$$F=\{F_{t_1,\cdots,t_n}(\lambda_1,\cdots,\lambda_n),\ n=1,2,\cdots \\ t_j\in T,\ j=1,\cdots,n\}. \tag{5}$$

In addition, F is called the *family of finite-dimensional distribution functions* of the stochastic process X. From (4) it is easily seen that F satisfies the following two conditions (consistency conditions):

A. For each permutation $(\alpha_1,\cdots,\alpha_n)$ of $(1,\cdots,n)$,

$$F_{t_1,\cdots,t_n}(\lambda_1,\cdots,\lambda_n)=F_{t\alpha_1,\cdots,t\alpha_n}(\lambda_{\alpha_1},\cdots,\lambda_{\alpha_n}).$$

B. If $m<n$, then

$$F_{t_1,\cdots,t_m}(\lambda_1,\cdots,\lambda_m)=\lim_{\lambda_{m+1},\cdots,\lambda_n\to\infty} F_{t_1,\cdots,t_n}(\lambda_1,\cdots,\lambda_n).$$

Now we are going to deal with the converse problem. In the above discussion, we first gave a stochastic process X, then a family of consistent finite-dimensional distribution functions is obtained. Now conversely, suppose what we give first is a parameter set T and a family (5) of finite-dimensional distribution functions satisfying the consistency conditions. The question is: Does there exist such a stochastic process that its family of finite-dimensional distribution functions exactly coincides with F ? The answer is "yes", or more precisely, we have the following theorem.

Theorem 1. *Given a parameter set T and a family of finite-dimensional distribution functions satisfying the consistency conditions, then there must exist a probability space $(\Omega,\mathfrak{F},P)$ and a stochastic process $X(\omega)=\{x_t(\omega), t\in T\}$ defined on it such that for each natural number n and arbitrary $\lambda_j\in R_1$, $t_j\in T$, $j=1,\cdots,n$,*

$$F_{t_1,\cdots,t_n}(\lambda_1,\cdots,\lambda_n)=P(x_{t_1}\leq\lambda_1,\cdots,x_{t_n}\leq\lambda_n). \tag{6}$$

Proof. Let $\Omega=R_T$, hence $\omega=\lambda(\cdot)$, where $\lambda(\cdot)$ stands for a real-valued function $\lambda(t), t\in T$ defined on T, $\mathfrak{F}=\mathfrak{B}_T$. Here $\mathfrak{B}_T$ expresses the minimal σ-algebra containing all those sets in R_T that have the form $(\lambda(\cdot):\lambda(t)\leq c)$ with arbitrary $t\in T$ and $c\in R_1$. From the Kolmogorov extension theorem and the consistency assumption, it follows that F generates a unique probability measure P_F defined on $\mathfrak{B}_T$, which satisfies

$$P_F(\lambda(\cdot):\lambda(t_1)\leq\lambda_1,\cdots,\lambda(t_n)\leq\lambda_n)=F_{t_1,\cdots,t_n}(\lambda_1,\cdots,\lambda_n). \tag{7}$$

Put $P=P_F$. Finally, we define

$$x_t(\omega)=\lambda(t), \text{ if } \omega=\lambda(\cdot) \tag{8}$$

In other words, $x_t(\omega)$ denotes a t-coordinate function, namely, the value of x_t at $\omega=\lambda(\cdot)$ is equal to the value $\lambda(t)$ of $\lambda(\cdot)$ at t. It can be easily seen that $(R_T,\mathfrak{B}_T,P_F)$ and $\{x_t(\omega), t\in T\}$ defined by (8) satisfy the requirement (6) of the theorem. In fact, it follows from (8) and (7) that

$$\begin{aligned}&P_F(x_{t_1}(\omega)\leq\lambda_1,\cdots,x_{t_n}(\omega)\leq\lambda_n)\\&\quad=P_F(\lambda(\cdot):\lambda(t_1)\leq\lambda_1,\cdots,\lambda(t_n)\leq\lambda_n)=F_{t_1,\cdots,t_n}(\lambda_1,\cdots,\lambda_n).\end{aligned}$$ #

(IV) Several basic concepts.

(a) A stochastic process $\{x_t(\omega), t\in T\}$ may be regarded as a function of two variables (t,ω) with $t\in T$ and $\omega\in\Omega$. As stated above, if t is fixed and $\{x_t(\omega), t\in T\}$ is considered to be a function of ω, we get a random variable $x_t(\omega)$. If ω is fixed and $\{x_t(\omega), t\in T\}$ is regarded as a function of t, we obtain a function $x_t(\omega)$ defined on T, and call it the *sample function* or *trajectory* corresponding to the elementary event ω.

(b) Let $\Xi=\{\xi(\omega)\}$ be a collection of some random variables $\xi(\omega)$ and consider the ω-set $(\omega:\xi(\omega)\leq\lambda)$. We get a collection $\{(\xi(\omega)\leq\lambda)\}$ of the subsets of Ω as $\xi(\omega)$ varies in Ξ and λ in R_1. The minimal σ-algebra containing this collection of subsets is denoted by $\mathfrak{F}\{\Xi\}$ and called the *σ-algebra generated by Ξ*. Therefore $\mathfrak{F}\{x_t, t\in T\}$ is the σ-algebra generated by the stochastic process $\{x_t(\omega), t\in T\}$.

(c) The two stochastic processes $\{x_t(\omega), t\in T\}$ and $\{\xi_t(\omega), t\in T\}$ defined on the same probability space $(\Omega,\mathfrak{F},P)$ are said to be *equivalent*, if for each fixed $t\in T$,

$$P(x_t(\omega)=\xi_t(\omega))=1. \tag{9}$$

It follows from (9) that for finitely or denumerably many $t_i\in T$, $i=1,2,\cdots$,

$$P(x_{t_i}(\omega)=\xi_{t_i}(\omega),\ i=1,2,\cdots)=1. \tag{10}$$

This shows that two stochastic processes equivalent to each other have the same family of finite-dimensional distribution functions.

(d) A stochastic process $\{x_t(\omega), t \in T\}$ (here T being an interval) is said to be *stochastically continuous at* $t_0 \in T$ if

$$P \lim_{t \to t_0} x_t(\omega) = x_{t_0}(\omega), \tag{11}$$

where $P\lim$ stands for the limit in a sense of convergence in measure P. If the process is stochastically continuous at every $t_0 \in T$, then we say that it is *stochastically continuous.* Replacing $t \to t_0$ by $t \to t_0 + 0$ (or $t \to t_0 - 0$), we obtain the definition of *stochastic continuity from the right* (or *left*).

(e) Whenever we say hereafter that almost all (or with probability one) sample functions possess a certain property A, we mean: There exists a Ω_0, $P(\Omega_0)=1$ such that for each $\omega \in \Omega_0$, the sample function $x(\cdot, \omega)$ possesses the property A ("$\cdot$" standing for the mobile coordinate on T). For instance, almost all sample functions with right lower semicontinuous (the property A) mean that there exists a set Ω_0 having the probability equal to one such that for $\omega \in \Omega_0$ we have $\underline{\lim} x(s, \omega) = x(t, \omega)$ for each $t \in T$. We have to distinguish this concept from the following: almost all sample functions are right lower semi-continuous at a fixed t. The latter means only

$$P(\omega: \lim_{s \to t} x(s, \omega) = x(t, \omega)) = 1$$

whereas the former implies a stronger conclusion:

$$P(\omega: \lim_{s \to t} x(s, \omega) = x(t, \omega), \text{ for all } t \in T) = 1.$$

(f) If all the random variables forming a stochastic process $\{x_t, t \in T\}$ take on values in the same set $I(\subset \overline{R}_1)$, we say that I is the *state space* of this process and each element i in I is related to a *state.* State spaces, generally speaking, are not unique, because any set containing I is also related to a state space. We call I the *minimal state space* if I is a state space and for each $i \in I$ there exists a $t \in T$ satisfying $P(x_t = i) > 0$. The state space is always referred henceforth to the minimal one unless otherwise stated. Sometimes, a state space is called a phase space denoted by either E or I.

(V) Previously, only stochastic processes that take on real values were discussed. If $x_t(\omega) = y_t(\omega) + iz_t(\omega)$, $\{y_t(\omega), t \in T\}$ and $\{z_t(\omega), t \in T\}$ being two real-valued stochastic processes defined on the same probability space, we then call $\{x_t(\omega), t \in T\}$ a *complex-valued stochastic process.* The processes to be discussed hereafter, unless otherwise stated, all pertain to real-valued ones.

Actually, the definition of the stochastic process can be generalized as follows: Given a probability space $(\Omega, \mathfrak{F}, P)$ and another measurable space $(E, \mathfrak{B})$ $(E = (e)$ denoting a set of points e, $\mathfrak{B}$ signifying a σ-algebra consisting

of some subsets of E , and E together with $\mathfrak{B}$ being called a measurable space), a variable $x(\omega)$ defined on Ω and taking on values in E is called a random variable if $(\omega: x(\omega) \in A) \in \mathfrak{F}$ for each set $A \in \mathfrak{B}$. Now let a parameter set T be given. If for any $t \in T$, it corresponds to a random variable $x_t(\omega)$ as stated above, $\{x_t(\omega), t \in T\}$ is then called a stochastic process taking on values in $(E,\mathfrak{B})$. In particular, a real-valued stochastic process is obtained when $(E,\mathfrak{B})$ reduces to $(R_1, \mathfrak{B}_1)$ (the set of real numbers and the collection of all its Borel sets). An n-dimensional stochastic process is obtained if $(E,\mathfrak{B})$ reduces to $(R_n,\mathfrak{B}_n)$($\mathfrak{B}_n$ is the collection of all the Borel sets of R_n).

We say that a set is discrete if it contains at most denumerably many elements. There may occur the four cases below according as T and E are discrete or continuous:

1° Both T and E are discrete.
2° T is discrete while E is continuous.
3° T is continuous while E is discrete.
4° Both T and E are continuous.

A stochastic process is also called a stochastic sequence for a discrete T.

§1.2 Separability of Stochastic Processes

(I) Let $\{\xi_t(\omega), t \in T\}$ be a stochastic process defined on $(\Omega,\mathfrak{F},P)$. Recall that we have already made $(\mathfrak{F},P)$ complete. In practical problems, we often need to discuss some ω-sets involving a number of nondenumerable t. For example, there is need for investigating the probability of

$$A=\{\omega: |\xi_t(\omega)| \leq \lambda, \text{ for all } t \in T\}, \tag{1}$$

where $\lambda \in R_1$. If T is neither a denumerable nor a finite set, then, since

$$A=\bigcap_{t \in T}(|\xi_t(\omega)| \leq \lambda),$$

the set A, as an intersection of nondenumerably many events, is generally not an event, i.e., in general $A \overline{\in} \mathfrak{F}$ and hence there is no probability of A to speak of.

Thus a difficulty arises: on the one hand, we need in practice to investigate the probability of A; on the other hand, it cannot even be ensured theoretically that A has a probability.

Similarly, the ω-sets

$B \equiv \{\omega$: the sample function $x_t(\omega)$ is continuous on T, $T=[0,\infty)\}$,
$C \equiv \{\omega$: the sample function $x_t(\omega)$ is a monotonic nondecreasing function on $T\}$

and so forth may not necessarily be events either.

One way to overcome this difficulty is to assume that the process possesses separability (see the definition below). Making use of the separability, we can reduce the study of a certain property A, which relates to all the para-

meters t, to that of the corresponding property involving only a number of denumerable parameters.

To simplify the notation, let T be an interval in R_1; in fact, the conclusions below hold for an arbitrary $T \subset R_1$, provided some evident modifications are made.

Let $x(t), t \in T$ be an arbitrary function which may take $\pm\infty$ as its values. Denote by X_T the two-dimensional set $\{(t, x(t)), t \in T\}$ (whose graph is a plane curve). Again let R be an arbitray denumerable subset in T, dense in T, and write $X_R = \{(r, x(r)), r \in R\}$, which is also a two-dimensional set. It is obvious that $X_R \subset X_T$.

Denote by $\overline{X}_R$ the closure of X_R in the usual distance[1], and hence $\overline{X}_R$ consists of X_R and all the limit points of X_R.

Definition 1. A function $x(t), t \in T$, is said to be *separable relative to* R if $X_T \subset \overline{X}_R$, that is, for every $t \in T$ there exists a sequence $\{r_i\} \subset R$ (r_i may be equal to r_j) such that we have simultaneously

$$r_i \to t, \quad x(r_i) \to x(t).$$

The R here is called a *separable set of the function.*

Definition 2. A stochastic process $\{x_t(\omega), t \in T\}$ is said to be *separable relative to* R if there exists a null set N such that the sample function $x_t(\omega)(t \in T)$ is separable relative to R for every $\omega \in \overline{N}$. In this case, R is called a *separable set of the process* and N, an *exceptional set.*

A stochastic process is said to be *separable* if there exists a denumerable subset R, which is everywhere dense in T, such that the process is separable relative to R.

A stochastic process is said to be *well-separable* if it is separable relative to an arbitrary set R as stated above.

Example 1. A continuous function is separable relative to the set R of rational points in T, and actually it is also well-separable.

Example 2. Let $s \in T$, s being an arbitrary irrational point, and a function $x(t)=0, t \in T \setminus s$, $x(s)=1$. Then this function is not separable relative to the set R of all the rational points in T; however, it is separable relative to $R \cup \{s\}$.

Example 3. Denote by F the set of rational points. Then the following function is inseparable relative to F:

$$x(t) = \begin{cases} 1, & \text{if } t \in F, \\ 0, & \text{if } t \in \overline{F}. \end{cases}$$

Take arbitrarily a denumerable set E of irrational points, which is dense in R_1. Then the above function is separable with respect to $F \cup E$.

---

1) Namely the distance between the two points $P_1=(x_1, y_1)$ and $P_2=(x_2, y_2)$ is given by $d(P_1, P_2)=\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}$.

Obviously, if a process $\{\xi_t(\omega), t \in T\}$ is separable with respect to R, then the set A in (1) and the event

$$A' \equiv \{\omega: |\xi_r(\omega)| \leq \lambda, \text{ for all } r \in R\} = \bigcap_{r \in R} (|\xi_r(\omega)| \leq \lambda) \in \mathfrak{F}$$

differ at most by a null set (which is a subset of N). Hence, owing to the completeness of $(\mathfrak{F}, P)$, A is an event too.

(II) **Theorem 1.** *For any stochastic process* $\{\xi_t(\omega), t \in T\}$ *defined on* $(\Omega, \mathfrak{F}, P)$, *there must exist a separable and equivalent stochastic process* $\{x_t(\omega), t \in T\}$.

This theorem indicates that, although a given process $\{\xi(t), t \in T\}$ is unnecessarily separable, there must exist a separable representative in its equivalent processes. Therefore, for a given family of consistent finite-dimensional distributions, by Theorem 1 in § 1.1 and Theorem 1 here, there must exist a separable process whose family of finite-dimensional distributions coincides with the given one. In other words, so long as the treated problem only involves a family of finite-dimensional distributions, we may assume that the considered process is separable. Now let us first prove the following lemma.

Lemma 1. *For any two intervals* J *and* G, $J \subset T$, *there exists a sequence* $\{s_n\} \subset J$ *such that for each fixed* $t \in J$,

$$P(\xi_t \in G, \ \xi_{s_n} \overline{\in} G, \ n=1,2,\cdots) = 0. \tag{2}$$

Proof. Choose $\{s_n\}$ by induction. Take an arbitrary point $s_1 \in J$. If $s_1, \cdots, s_n$ have been chosen already from J, we put

$$P_n = \sup_{t \in J} P(\xi_t \in G, \ \xi_{s_1} \overline{\in} G, \cdots, \xi_{s_n} \overline{\in} G). \tag{3}$$

Hence there must exist a point $s_{n+1} \in J$ satisfying

$$P(\xi_{s_{n+1}} \in G, \ \xi_{s_1} \overline{\in} G, \cdots, \xi_{s_n} \overline{\in} G) \geq P_n(1 - \frac{1}{n}). \tag{4}$$

These events $G_n = (\xi_{s_{n+1}} \in G, \ \xi_{s_1} \overline{\in} G, \cdots, \xi_{s_n} \overline{\in} G)$ $(n=1,2,\cdots)$, however, are disjoint, thus

$$\sum_{n=1}^{\infty} P(G_n) \leq 1.$$

Consequently, the values $P_n(1-\frac{1}{n})$ of the right-hand side in (4) tend to zero. This shows that

$$\lim_{n \to \infty} P_n = 0. \tag{5}$$

Secondly, now that for every fixed t we have

$$\begin{aligned} &(\xi_t \in G, \ \xi_{s_i} \overline{\in} G, \ i=1,2,\cdots,n) \\ &\supset (\xi_t \in G, \ \xi_{s_i} \overline{\in} G, \ i=1,\cdots,n+1) \supset \cdots, \end{aligned}$$

the intersection of those events being the very event in (2). Therefore the proof of (2) is completed by (3) and (5). #

Proof of Theorem 1. We say that two intervals J and G $(J\subset T)$ with rational numbers as their endpoints are a "duad". All such duads form a denumerable set. For each duad (J,G), a sequence $\{s_n\}$ possessing the properties in Lemma 1 can be obtained. When such sequences and rational numbers in T are merged, a subset R denumerable and dense in T, is acquired. If some new points are added to $\{s_n\}$, the event in (2) will not enlarge. Therefore, R has the following property:

For each fixed $t\in T$ and each fixed duad (J,G), satisfying $t\in J$, it is found

$$P(\xi_t\in G,\ \xi_s\overline{\in} G, \text{ for all } s\in JR)=0. \tag{6}$$

Now fix t and denote by A_t the event that "there exists at least a duad $(J,G), t\in J$ such that $\xi_t\in G$ and $\xi_s\overline{\in} G$ hold for all $s\in JR$". Then it follows from (6) that

$$P(A_t)\leq\sum_{J,G}P(\xi_t\in G,\ \xi_s\overline{\in} G, \text{ for all } s\in JR)=0,$$

hence $P(\overline{A}_t)=1$. Below we fix $\omega\in\overline{A}_t$ arbitrarily and take any G such that $\xi_t(\omega)\in G$. For an arbitrary J containing t, by the definition of $\overline{A}_t$, there must exist an $s\in JR$ such that $\xi_s(\omega)\in G$; otherwise, this $\omega\in A_t$. Because of the arbitrariness of J, $\{u_j\}\subset R$ can be found as J lessens such that $u_j\to t$ and each $\xi_{u_j}(\omega)\in G$ as well.

Now take $G_n\supset G_{n+1}$ such that not only $\xi_t(\omega)\in G_n$ but also the lengths of G_n tend to zero. As stated above, for each G_n, $\{u_j^{(n)}\}\subset R$ can be found such that

$$u_j^{(n)}\to t\,(j\to\infty),\quad \xi_{u_j^{(n)}}\in G_n.$$

Choose a sequence $\{v_i\}\subset R$ as follows:

Put $v_1=u_1^{(1)}$, and v_n is an arbitrary $u_k^{(n)}$ satisfying $|u_k^{(n)}-t|<1/n$. Obviously, $v_n\to t, \xi_{v_n}(\omega)\to\xi_t(\omega)$, $(n\to\infty)$. This indicates that

$$(t,\xi_t(\omega))\in\overline{\Xi_R(\omega)}=\overline{((r,\xi_r(\omega)),r\in R)}.$$

Since $\omega\in\overline{A}_t$ is arbitrary, thus we have proved that

$$P((t,\xi_t(\omega))\in\overline{\Xi_R(\omega)})\geq P(\overline{A}_t)=1 \tag{7}$$

for any fixed $t\in T$.

Now construct a new process $\{x_t(\omega), t\in T\}$: for any $\omega\in\Omega$, when $t\in R$, let

$$\left.\begin{aligned} &x_t(\omega)=\xi_t(\omega);\\ &\text{when } t\overline{\in} R, \text{ let}\\ &\quad x_t(\omega)=\xi_t(\omega), \text{ if } (t,\xi_t(\omega))\in\overline{\Xi_R(\omega)},\\ &\quad\ =\delta_t(\omega), \text{ if } (t,\xi_t(\omega)\overline{\in}\ \overline{\Xi_R(\omega)}, \end{aligned}\right\} \tag{8}$$

where $\delta_t(\omega)$ should be chosen in such a way that $(t,\delta_t(\omega))\in\overline{\Xi_R(\omega)}$. Such a $\delta_t(\omega)$ can always be found by the following method: take arbitrarily a sequence