ESSAY I. INTRODUCTION, CONSTRUCTION, AND FUNDAMENTAL PROPERTIES

0. INTRODUCTION. In this first essay, our subject is introduced in a setting general enough to cover its uses in the remainder of the work. Then the fundamental properties and results needed later are developed and proved from scratch, making only minimal use of the "general theory of processes," as presented for example in C. Dellacherie [5]. In the later material, which prepares the method developed here for application in various more specialized situations, it is inevitable that there be more reference to, and reliance on, the results of the Strasbourg school as developed in Volumes I-XII of the Strasbourg Seminaire de Probabilites [14], in C. Dellacherie [5], in C. Dellacherie and P.-A. Meyer [4], and in R. K. Getoor [8]. Yet it should be emphasized that the prediction process is not simply another chapter in this development. Rather it is a largely new method. It could be developed in the framework of the above, but whatever would be gained in brevity and completeness would be offset, at least for the reader who is less than fully familiar with the Strasbourg developments, by the prerequisites. Consequently, we have tried to proceed here in such a way as to be understood by the less initiated reader, and yet not to be considered infantile by the initiated. For the reader who is familiar with the Strasbourg work, and wants to get an idea of what the prediction process means in that setting, the second essay below may be read as an introduction. It does not depend on the more general theory to be developed. The aim here is not to incorporate the prediction process into any general theory of stochastic processes, but to develop it as an independent entity.

Having gone this far in setting our work apart from that of the Strasbourg group, we must hasten to give credit where due. In the first place, the present work borrows unsparingly from the papers of P.-A. Meyer [12], of M. Yor and P.-A. Meyer [13], and of M. Yor [15], on the technical side. The proof of the Markov property of the prediction process, which was difficult (and possibly incomplete) in Knight [9], is derived in these papers from a stronger identity holding pathwise on the probability space, and we follow their method. Again, the very definition of the process in [12] avoids the necessity of completing the σ-fields (until a later stage), and we adopt this improvement. The measurability of the dependence of the process on the initial measure, too, is due to these authors. On this score, we have not hesitated to profit from their mistakes, as described in [12] and [13]. Further, the basic role of the set H_0 of "non-branching points" is due to P.-A. Meyer ([12, Proposition 2]). Finally and perhaps most importantly, we adopt a new idea of M. Yor [15] to the effect that one need not only predict the future of the specified process in order to get a homogeneous Markov process of prediction probabilities. One may just as well predict the futures of any countable number of other processes at the same time. The only essential precondition is that the future of the specified process (the process which generates the "known past") must be included in the future to be predicted. This, in our opinion, places the prediction process of [9] into an entirely new dimension.

Meanwhile, in regard to our use of the Strasbourg ideas and formalism, we would emphasize the distinction between σ-fields on a probability space, such as F_t^o, F_t^μ etc., and σ-fields of a product space in which time is one coordinate, such as the optional or previsible σ-fields. It is often very convenient to use σ-fields of the latter type, and for a complete understanding of many results, they are probably unavoidable. On the other hand, while σ-fields of the former type are needed to express the state-of-affairs as it actually exists at a given t, σ-fields of the latter type are needed rather to define various kinds of processes, usually as an auxiliary, and they can always be circumvented at a cost of sacrificing some degree of completeness. Thus, one will not go essentially wrong in the present work, if one substitutes right-continuous and left-continuous adapted process for "optional" and "previsible" process, respectively, and limit of a strictly increasing sequence of stopping times for "previsible stopping time". In particular, while the section theorems are used freely in establishing results for all t, no use is made of the corresponding projections of a measurable process, although they are heavily implicated in the results.

To give a general preview of the applications treated in subsequent essays, some of them (such as the Lévy system of a martingale) may be stated without any reference to the prediction process, and when possible such formulations are included. For these, the prediction aspect is needed only for the proofs. For most results, however, the prediction process is a necessary part of the formulation of the idea or problem involved. The central purpose is thus to elaborate, and by implication more or less to phase in, the prediction process as a feature of the general theory of stochastic processes. Once the reader becomes adept at thinking in terms of this process, other applications will suggest themselves immediately according to the context, or so we have found. For example, very little is done here in the way of using the prediction process itself in the manner of probabilistic potential theory. In this way, many stopping times of the given process would become first passage times of the prediction process, but the interconnection of the two processes remains largely unstudied. It might be of interest to follow such a direction farther. Even less (if possible) has been done in connecting the present work with stochastic integration, a medium in which the author is not highly proficient. Accordingly, such matters are left aside in favor of applications in which we can feel confident at least that a correct beginning has been made.

1. THE PREDICTION PROCESS OF A RIGHT-CONTINUOUS PROCESS WITH LEFT LIMITS.

We use the following standard notation for measurability.

1) If $\mathcal{F}$ and $\mathcal{G}$ are σ-fields on spaces F and G, a random variable or function $X : F \to G$ is $\mathcal{F}/\mathcal{G}$ - measurable, or $X \in \mathcal{F}/\mathcal{G}$, if $X^{-1}(S) \in \mathcal{F}$ for $S \in \mathcal{G}$, and when X is real or extended-real valued and $\mathcal{G}$ is the corresponding Borel σ-field, we write simply $X \in \mathcal{F}$.
2) $b(\mathcal{F})$ denotes the bounded, real-valued, $\mathcal{F}$ - measurable functions; $b^+(\mathcal{F})$ denotes the non-negative elements of $b(\mathcal{F})$. Further, we denote the extended real line $[-\infty, \infty]$ by $\overline{R}$, with Borel sets $\overline{\mathcal{B}}$ and the product space $(X_{n=1}^{\infty} \overline{R}, X_{n=1}^{\infty} \overline{\mathcal{B}})$ by $(\overline{R}_\infty, \overline{\mathcal{B}}_\infty)$.

We begin with the following measurable space.

DEFINITION 1.1. Let Ω denote the space of all sequences $w(t) = (w_1(t), w_2(t), \ldots, w_n(t), \ldots)$ of right-continuous extended-real-valued functions of $t \geq 0$, with left limits for $t > 0$. Let $\mathcal{G}_t^o$ denote the σ-field generated by all $w_n(s), s \leq t, n \geq 1$, and let $\mathcal{F}_t^o$ denote that generated by all $w_{2n}(s)$, $n \geq 1$, so that $\mathcal{F}_t^o \subset \mathcal{G}_t^o$. We set $X = (w_{2n}(t), n \geq 1)$ on Ω, and $\mathcal{F}^o = \bigvee_t \mathcal{F}_t^o$, $\mathcal{G}^o = \bigvee_t \mathcal{G}_t^o$. Thus X_t has right-continuous paths with left limits in $X_{n=1}^{\infty} \overline{R}_{2n}$ with the product topology. Finally, we set $\theta_t w(s) = w(t+s)$ on Ω, and denote by P a fixed probability on $(\Omega, \mathcal{G}^o)$.

Before going further, we give a brief rationale for selecting this as the initial structure. In setting up a prediction process, we require basically two things. The first is a process which generates the conditioning σ-fields (in this case, the process X_t) and the second is a definition of the futures which are to be predicted (in this case, $\theta_t^{-1}\mathcal{G}^o$), which must contain those of X_t (namely $\theta_t^{-1}\mathcal{F}^o$) . Once we define the process X_t and the futures $\theta_t^{-1}\mathcal{G}^o$, there may be some latitude as to exactly how these futures are to be generated, but it seems to be necessary that they be generated by processes in order to write then with shift operators in the form $\theta_t^{-1}G^o$. This being granted, the remainder of our set-up represents a compromise between the more general assumption of [9], where X_t was only a measurable process, and the more familiar requirements of the applications, in all of which X_t is right-continuous with left limits (abbrev. r. c. l. l.). Since X_t is assumed r. c. l. l., it is logical that the "unobserved" processes $w_{2n-1}(t)$ $n \geq 1$, are also r. c. l. l.

It should be pointed out that the choice of real-valued processes is only a matter of convenience. If the actual process has values in a locally compact metric space, or even a metrizable Lusin space (homeomorphic to a Borel subset of a compact metrizable space) we can obtain the above situation by considering the processes composed with a sequence of uniformly continuous functions separating points. Similarly, if the actual process X_t is real-valued, we may take $P\{w_{2n}(t) = 0, n > 1\} = 1$ and replace Ω by the corresponding subset, and so forth. It is easy to see that our set-up is P-indistinguishable from the canonical space of right-continuous paths with left limits in the product of any two metrizable Lusin spaces, but we prefer the more explicit situation.

A property of (Ω, G^o) which is needed in setting up the prediction process is the existence of regular conditional probabilities, given any subfield. For this it is of course sufficient that (Ω, G^o) is the Borel space of a metrizable Lusin space, i.e., a "measurable Lusin space" in the language of Dellacherie-Meyer [4, Chap. III, Definition 16]. There are many different topologies under which the present (Ω, G^o) becomes a measurable Lusin space. It suffices to write $\Omega = X_{n=1}^{\infty} \Omega_n$, then to give each Ω_n a Lusin topology as (a copy of) the space of all extended-real-valued right-continuous paths with left limits, and finally to give Ω the product topology. In the present work, we specialize on one particular such topology, a transplant to the present context of the one used in Knight [9]. This turns out again to be quite natural, and to have some rather unique advantages. In brief, this is the topology of scaled weak convergence of sample paths. This topology is metrizable in such a way that the completion of Ω is the space of all sequences of equivalence classes of measurable functions (with respect to Lebesgue measure). The completion is then a compact metric space, which we denote by $\bar{\Omega}$, and Ω is embedded in $\bar{\Omega}$ as a Borel subset. For some purposes, $\bar{\Omega}$ is a more natural space than Ω, and a few results will concern $\bar{\Omega}$ explicitly. The prediction process can be constructed on $\bar{\Omega}$ in complete analogy to Ω, but for simplicity we leave this to the reader (see also [9] and [12]).

Before beginning the construction, one more remark on its essential nature may provide orientation. It it generally accepted that a stochastic process, in application, is a model of a phenomenon which develops according to laws of probability. But there is no such agreement as to the nature of probability itself. Some authors (including such renowned figures as Laplace and Einstein) seem to have doubted that probability even exists in an absolute physical sense. However, it seems unlikely that anyone can doubt that probability does exist in a mental sense, as a way of thinking. If only because one does not know the entire future, it is clear that probabilistic thinking is an alternative possible procedure in many situations. Indeed, it may be the only one possible. Consequently, it can scarcely be doubted that stochastic processes do exist in some useful sense, if only, perhaps, in the minds of men. Furthermore, even if objective probabilities do exist entirely apart from subjective ones, it cannot be considered unimportant to study the more subjective aspects of probability. As with many other branches of mathematics, one is in a better position to make applications of probability to the physical world once one understands fully the mental presuppositions which are involved in the applications. Indeed, a large part of mathematics consists precisely in cultivating and developing the necessary mental operations, and one of the fundamental requirements for knowing how to apply mathematics lies in distinguishing what is a physical fact from what is only part of the mental reasoning. Thus, in stochastic processes as elsewhere in mathematics, it is important and useful to understand what one is doing mentally.

Coming, now, to the case of the prediction process, in much the same way as the probability distributions govern the development of a stochastic process, so the prediction process governs, or models, the development of these probabilities themselves. The prediction process, then, is a process of conditional probabilities associated with a given or assumed stochastic process. The given information will be that of the "past" (or observed part) of the given process, and the probabilities will be the conditional probabilities of the "future" (or unobserved part). In this way, the prediction process becomes at first an auxiliary (or second level) stochastic process associated with the given process. But the remarkable advantages of the method appear only when we consider this as a process per se, and define the original process in terms of it instead of conversely. This last step constitutes, in a sense, the main theme of the present work. The first step, however, is definition of the prediction process of the given X_t, and this is our immediate task.

We set $\rho(x) = \pi^{-1}(\pi/2 + \arctan x)$, $-\infty \le x \le \infty$, and consider the sequential process on Ω

$Y(t) = (Y_n(t)) = (\int_0^t e^{-s}\rho(w_n(s))\,ds,\ 1 \le n)$. Since

$\frac{d^+Y(t)}{dt^+} = (e^{-t}\rho(w_n(t)),\ 1 \le n)$ it is clear that $\{Y(s),\ s < t\}$ generates the σ-field $G^o_{t-} = V_{s<t}\, G^o_s$. In fact, the same is true of $\{Y(r),\ r$ rational, $r < t\}$, since the right derivatives at the rationals determine the right-continuous functions $\rho(w_n(s))$. In particular, G^o is generated by $Y(r)$, r rational, and hence by the countable collection of random variables $Y_n(r) = \int_0^r e^{-s}\rho(w_n(s))\,ds$. This countability is essential to the method, which relies on martingale convergence a.s. ("almost surely", i.e., with probability one) at a critical place. The random variables $Y_n(r)$ are analogous to those of [9, Def. 1. 1. 1], and also to those of the set V of [13, Lemma 1].

We note that $0 \le Y_n(r) \le 1$, and that each $Y_n(t)$ satisfies a uniform positive Lipschitz condition of order 1 :
: $0 \le Y_n(t+s) - Y_n(t) \le e^{-t}s \le s$. In particular, convergence of $Y_n(r)$ for each rational $r \ge 0$ is equivalent to uniform convergence.

We will be concerned with the uniformly closed algebra of functions generated on Ω by the $Y_n(r)$. Explicitly, this may be generated as follows. For each $m \ge 1$, let $f_{m,j}(x_1,\dots,x_m)$, $1 \le j$, be a sequence of continuous functions on $[0,1]^m$ which is uniformly dense in the set of all such functions. Then the algebra in question is the uniform (linear) closure of all the random variables $f_{m,j}(Y_1(r_1),\dots,Y_m(r_m))$, for all positive rationals $r_1,\dots,r_m$, $1 \le j$, $1 \le m$. This is easily checked by first fixing m and $r_1,\dots,r_m$, noting that the range of $(Y_1(r_1),\dots,Y_m(r_m))$ is compact, and then using the uniform continuity of $f_{m,j}$ in conjunction with the Stone-Weierstrass approximation theorem.
REMARK. More generally, if we enumerate the $Y_i(r_j)$, and choose any countable collection of continuous functions on the Hilbert cube $X_{n=1}^{\infty}[0,1]$ which is uniformly dense in the set of all such functions, the composition of these with the sequence $Y_i(r_j)$ can replace the above particular choice.

We let U denote this algebra, and summarize the needed function - theoretic properties as follows.

THEOREM 1.2 a) The topology on Ω generated by U is metrizable by the metric

$$d(w^a,w^b) = \sum_n 2^{-n}||Y_n^a - Y_n^b|| \ ,$$

where $||f(t)|| = \sup_t |f(t)|$. The completion of Ω in this metric is a compact metric space $\bar{\Omega}$, whose elements are identified with all sequences of (equivalence classes mod Lebesgue-null sets of) measurable functions $w_n(t) : R^+ \to \bar{R}$, with the same definition of d . With this identification, Ω is Borel in $\bar{\Omega}$, and the Borel field on Ω is G^o .

b) The map $(t,w) \to \theta_t w$ is continuous from $[0,\infty) \times \bar{\Omega} \to \bar{\Omega}$.

PROOF. We have already noted that convergence in the topology generated by U is the same as uniform convergence of each Y_n^a, proving the first assertion. On the other hand, since this convergence is equivalent to weak convergence of each Y_n^a considered as a distribution function, the completion is a closed subset of the space of all sequences of distributions of mass ≤ 1 on $[0,\infty)$, which is compact under weak convergence by Helly's Theorem. Hence $\bar{\Omega}$ is a compact metric space. An element of $\bar{\Omega}$ is given by a sequence of uniform limits of Y_n's i.e. by a sequence of non-decreasing continuous functions of Lipschitz constant 1. Such functions being absolutely continuous, we may identify them as integrals of their a.e. - derivatives $\rho(w_n(t)) \leq 1$, and $w_n(t)$ is identified by applying ρ^{-1} .

Conversely, given any sequence w_n of measurable functions, the functions $\rho(w_n)$ are bounded by 0 and 1 and measurable. For such functions, convergence in the metric d is simply convergence in finite time intervals in the weak topology $\sigma(L_1,L_\infty)$, i.e. convergence of $\int(\rho w_n(t))f(t)dt$ for bounded measurable f with compact support (or equivalently, only for continuous f, which are dense in L_1). The closure of the continuous functions ρw_n is all measurable functions bounded by 0 and 1, since it contains the L_1-closure in finite time intervals. Therefore, the completion includes all measurable w_n as asserted.

Finally, an approximation by Riemann sums shows that $\int_0^t \rho(w(s))ds$ is G^o-measurable on Ω for each t . It follows that $d(w^a,w^b)$ is G^o-measurable on Ω for fixed $w^b \in \bar{\Omega}$, and therefore the inclusion mapping $(\Omega,G^o) \to (\bar{\Omega},\bar{G}^o)$ is Borel, where $\bar{G}^o$ denotes the Borel field of $\bar{\Omega}$. By a well-known theorem [4, III, Theorem 21] it follows that $\Omega \in \bar{G}^o$ and $G^o = \bar{G}^o|\Omega$, as asserted.

Turning to b), we have

$$|Y_n(r,\theta_{t+\varepsilon}w^a) - Y_n(r,\theta_t w^b)| = |e^{t+\varepsilon}\int_{t+\varepsilon}^{t+r+\varepsilon} e^{-s}\ \rho(w_n^a(s))ds$$
$$- e^t\int_t^{t+r} e^{-s}\ \rho(w_n^b(s))ds| \le (e^{\varepsilon}-1)\int_{t+\varepsilon}^{t+r} e^{(t-s)}\rho(w_n^a(s))ds$$
$$+ e^t|\int_{t+\varepsilon}^{t+r} e^{-s}(\rho(w_n^a(s)) - \rho(w_n^b(s)))ds|$$
$$+ \int_{t+r}^{t+r+\varepsilon}\rho(w_n^a(s))ds + \int_t^{t+\varepsilon}\rho(w_n^b(s))ds$$
$$\le (e^{\varepsilon}-1) + 2e^t||Y_n^a - Y_n^b|| + 2\varepsilon\ .$$

This easily implies b).

REMARK. This topology is somewhat artificial in that it depends on the choice of the function ρ . The artificiality disappears, however, if we begin with a process Y_t having values in a metrizable Lusin space $E \subset \bar{E}$ compact, and consider $w_n(t) = f_n(Y_t)$ where (f_n) is a uniformly dense subset of the continuous functions on $\bar{E}$. Then the topology of Ω reduces to that of weak convergence of sojourn time measures $\mu(t,A) = \int_0^t I_A(Y_s)ds$ for the process Y .

We turn now to the state space of the prediction process.

DEFINITION 1.2. Let $(H, \mathcal{H})$ be the set of all probability measures on $(\Omega, \mathcal{G}^o)$, with the σ-field generated by $\{h(S),\ S \in \mathcal{G}^o\}$ as functions on H . We give H the topology of weak-* convergence with respect to convergence in the topology of $\bar{\Omega}$. Let P^h and E^h, $h \in H$, denote the probability and expectation determined over $\mathcal{G}^o$ by h .

PROPOSITION 1.3. H is a separable metrizable space, with respect to which $\mathcal{H}$ is the Borel σ-field. The metric may be so defined that H is Borel in its completion $\bar{H}$, the compact metrizable space of all probabilities on $\bar{\Omega}$.

PROOF. Since Ω is embedded in the compact $\bar{\Omega}$, there is a uniformly dense sequence f_n in the uniformly continuous functions on Ω . The functions $E^h f_n$ then induce the topology of H, which is clearly metrizable with completion $\bar{H}$. Since $E^h f$ is then measurable for continuous f, by a monotone class argument $E^h f$ is measurable for bounded, $\mathcal{G}^o$-measurable f . Hence $h(S) = E^h I_S$ is measurable for $S \in \mathcal{G}^o$, as asserted. Finally, since $H = \{h \in \bar{H} : h(\Omega) = 1\}$, H is Borel in $\bar{H}$.

For the construction of the prediction process we introduce a fixed sequence of continuous functions which are suitably bounded, but the outcome is entirely free of which such sequence is used.

NOTATION 1.4. Let $0 \le f_n \le 1$ be any fixed sequence of continuous functions on $\bar{\Omega}$ whose uniform linear closure is all continuous functions. For example, the $f_{m,j}(Y_1(r_1),\ldots,Y_m(r_m))$ of Theorem 1.2 suffice, when extended by continuity to $\bar{\Omega}$. We also need

LEMMA 1.5. For $\lambda > 0$, and bounded measurable $f \ge 0$, the expressions $f_{\lambda,h}(t) = e^{-\lambda t} E^h(\int_0^\infty e^{-\lambda s} f \cdot \theta_{t+s}\, ds \mid F_t^o)$ are P^h-supermartingales in t for every $h \in H$.

PROOF. This is a familiar computation, due to G. Hunt.

$$
\begin{aligned}
&E^h(f_{\lambda,h}(t_1 + t_2) \mid F^o_{t_1}) \\
&= e^{-\lambda(t_1+t_2)} E^h(\int_0^\infty e^{-\lambda s} f \cdot \theta_{t_1+t_2+s}\, ds \mid F^o_{t_1}) \\
&= e^{-\lambda t_1} E^h(\int_0^\infty e^{-\lambda(t_2+s)} f \cdot \theta_{t_1+(t_2+s)}\, ds \mid F^o_{t_1}) \\
&\le f_{\lambda,h}(t_1) .
\end{aligned}
$$

In order to use martingale convergence with Lemma 1.5, we first choose for each rational $r > 0$ a regular conditional probability $W_r^h(S)$, $S \in G^o$, of $\theta_r^{-1} S$ given F_r^o. In fact, we choose W_r^h to be $H \times F_r^o$ - measurable in (h,w) as is possible by a well-known construction of J. L. Doob (using the fact that F_r^o is countably generated -- for the method, see also Theorem 1.4.1 of [9]). Thus we may be more precise in Lemma 1.5 for $f = f_n$ by setting

$$f_{\lambda,n,h}(r) = e^{-\lambda t} E^{W_r^h} \int_0^\infty e^{-\lambda s} f_n \cdot \theta_s\, ds ,$$

and we now assume this particular choice.

Next, we prepare one more lemma.

LEMMA 1.6. For any $t \ge 0$, $h \in H$, and $w \in \Omega$, existence of the limits along the rationals r

$$\lim_{r \to t+} f_{\lambda,n,h}(r)$$

for all n and all rational $\lambda > 0$ is equivalent to the existence of

$$\lim_{r \to t+} W_r^h$$

in the topology of $\bar{H}$.

PROOF. By definition of the weak-* topology, existence of the last limit is equivalent to that of

$$\lim_{r\to t+} E^{W_r^h} f_n$$

for all n. Now by Fubini's Theorem we have

$$f_{\lambda,n,h}(r) = e^{-\lambda r}\int_0^\infty e^{-\lambda s}\, E^{W_r^h}(f_n \cdot \theta_s)ds\ ,$$

and by Theorem 1.2 b) we know that $E^{W_r^h}(f_n \cdot \theta_s)$ is continuous in s, uniformly in r. Clearly it is bounded by 1. Thus $f_{\lambda,n,h}(r)$ is uniformly continuous in $\lambda > \varepsilon > 0$, uniformly in r, for each n and ε. Convergence of $f_{\lambda,n,h}(r)$ as $r \to t+$ for all $\lambda > 0$ implies, by the continuity theorem for Laplace transforms, convergence of the measures

$$(E^{W_r^h} f_n \cdot \theta_s)ds\ .$$

By a simple use of equicontinuity of these densities in s, this is equivalent to convergence of

$$E^{W_r^h} f_n \cdot \theta_s$$

for each s. But at $s = 0$ this implies convergence of W_r^h in H, as n varies. Conversely, since each $f_n \cdot \theta_s$ is continuous on $\bar{\Omega}$, convergence of W_r^h implies that of

$$E^{W_r^h} f_n \cdot \theta_s$$

for each s. Hence by the dominated convergence theorem, we obtain the existence of limits of $f_{\lambda,n,h}(r)$ and the proof is complete.

We can now give the definition of the prediction process for fixed $h \in H$.

DEFINITION 1.6. Let $T_h = \sup\{t : \text{for } 0 \le s < t \text{ the limits } W_{s\pm}^h = \lim_{r\to s\pm} W_r^h \text{ both exist and are in } H\}$. We define the prediction process of h by

$$Z_t^h = \begin{cases} \lim_{r \to t+} W_r^h & \text{on } \{t < T_h\} \\ h & \text{on } \{t \geq T_h\} \end{cases} .$$

In discussing optional and previsible stopping times, it is convenient to use the σ-fields F_{t+}^h consisting of F_{t+}^o augmented by all h-null sets in the h-completion of F^o. Furthermore, there is no loss of generality in the following theorem to assume $0 < T < \infty$ for previsible stopping times T, since we may replace any T by $T_n = T \wedge n$ on $\{T > 0\}$ and $T_n = n$ on $\{T = 0\}$, and let $n \to \infty$ (on $\{T = 0\}$ the corresponding form of the assertion is trivial).

THEOREM 1.7. a) For $h \in H$, $P^h\{T_h = \infty\} = 1$, and Z_t^h is F_{t+}^o-measurable for each t. It is right-continuous, with left limits Z_{t-}^h in H except perhaps at $T_h < \infty$, and it is $H \times B \times F^o$ measurable in (h,t,w).

b) For every F_{t+}^h-optional stopping time $T < \infty$, we have

$$P^h(\theta_T^{-1} S | F_{T+}^h) = Z_T^h(S) \ , \quad S \in G^o . \tag{1.7b}$$

c) For every F_{t+}^h-previsible stopping time $0 < T < \infty$ we have

$$P^h(\theta_T^{-1} S | F_{T-}^h) = Z_{T-}^h(S) \ , \quad S \in G^o , \tag{1.7 c}$$

where we set $Z_{T_h-}^h = h$ if the left limit at $T_h < \infty$ does not exist.

d) The processes Z_t^h and Z_{t-}^h are respectively F_{t+}^h-optional and F_{t+}^h-previsible, and either of these facts together with (1.7b) or (1.7c) respectively, determines Z_t^h or Z_{t-}^h uniquely up to an h-null set for all $t > 0$ (≥ 0 if we set $Z_{0-}^h = h$).

REMARKS. It follows from [4, VI, 5] that Z_t^h is even F_{t+}^o-optional.

PROOF. By Lemma 1.6 and the classical supermartingale convergence theorem of Doob (continuous parameter version) we know that $P^h\{\lim_{r \to s\pm} W_r^h = W_{s\pm}^h$ exist for all $s\} = 1$. Unfortunately, there seems to be no way to deduce from this that the limits are concentrated on Ω (hence are in H) except by first proving parts b)-d) for $W_{s\pm}^h$ in place of $Z_{s\pm}^h$ (this is the price we pay for using Ω instead of $\bar{\Omega}$). Accordingly, let T be any finite F_{t+}^h-stopping time, and let $T_k = (m+1)2^{-k}$ on $\{m2^{-k} \leq T < (m+1)2^{-k}\}$ for all $m \geq 0$, as usual. Then by Theorem 1.2 b), and martingale convergence of conditional expectations, we have

(1.8)
$$E^h\left(\int_0^\infty e^{-\ s} f_n \cdot \theta_{T+s} ds \,|\, F^h_{T+}\right)$$
$$= \lim_{k\to\infty} E^h\left(\int_0^\infty e^{-\lambda s} f_n \cdot \theta_{T_k+s} ds \,|\, F^h_{T_k+}\right)$$
$$= \lim_{k\to\infty} E^{W^h_{T_k}}\left(\int_0^\infty e^{-\lambda s} f_n \cdot \theta_s ds\right)$$
$$= E^{W^h_{T+}} \int_0^\infty e^{-\lambda s} f_n \cdot \theta_s ds \ .$$

By monotone class argument using linear combinations of the f_n, it follows that W^h_{T+} defines a regular conditional probability on $\theta^{-1}_T G^o$ given F^h_{T+}, and in particular $W^h_{T+}(\Omega) = 1$ a.s. Since W^h_{t+}(set $\equiv 0$ where it does not exist for all t) is F^h_{t+}-optional, the optional section theorem [4, IV, 84] shows that $P^h\{W^h_{t+}(\Omega) = 1$ for all $t\}$ $= 1$.

Turning to W^h_{t-}, let $0 < T < \infty$ be an F^h_{t+}-previsible stopping time. By [4, IV, 71 and 77] this is equivalent to the existence of an increasing sequence (T_k) of stopping times with $T_k < T$ and $\lim_{k\to\infty} T_k = T < \infty$. Then by (1.8) and Hunt's Lemma [4, V, 45] we have

(1.9)
$$E^h\left(\int_0^\infty e^{-\lambda s} f_n \cdot \theta_{T+s} \,|\, \bigvee_k F^h_{T_k}\right)$$
$$= \lim_{k\to\infty}\left(\int_0^\infty e^{-\lambda s} f_n \cdot \theta_{T_k+s}\, ds \,|\, F^h_{T_k}\right)$$
$$= \lim_{k\to\infty} E^{W^h_{T_k+}} \int_0^\infty e^{-\lambda s} f_n \cdot \theta_s\, ds$$
$$= E^{W^h_{T-}} \int_0^\infty e^{-\lambda s} f_n \cdot \theta_s\, ds \ .$$

But by [4, IV, 56 b) and d)] we have $\bigvee_k F^h_{T_k} = F^h_{T-}$, and so by monotone class argument W^h_{T-} defines a regular conditional probability on $\theta^{-1}_T G^o$ given F^h_{T-} . Since W^h_{t-} (set $\equiv 0$ where it does not exist for all t) is F^h_{t+}-previsible, the previsible section theorem [4, IV, 85] shows that $P^h\{W^h_{t-}(\Omega) = 1$ for all $t > 0\} = 1$.

Combining the above results, it follows immediately that $P^h\{\tau_h = \infty\} = 1$ and we have (1.7b) and (1.7c). It is clear that T_h is even an $\mathcal{F}^o_{t+}$ stopping time, and obviously Z^h_t is right continuous, with left limits except perhaps at $T_h < \infty$. It now follows by [5, IV, T 27] that Z^h_t is $\mathcal{F}^h_{t+}$-optional. To see that Z^h_{t-} is $\mathcal{F}^h_{t+}$-previsible it suffices to note that it is a measurable process h-equivalent to the previsible process W^h_{t-}, since $\mathcal{F}^h_{t+}$ contains all subsets of h-null sets. In view of the two sections theorems, this completes the proof of d). Finally, the joint measurability assertion in a) follows from the $\mathcal{H} \times \mathcal{F}^o_r$-measurability of W^h_r since $I_{[0,T_h)}(t)$ is $\mathcal{H} \times \mathcal{B} \times \mathcal{F}^o$-measurable and

$$Z^h_t = W^h_{t+}\, I_{[0,T_h)} + h\, I_{[T_h,\infty)} .$$

In fact, for later use we may state

COROLLARY 1.7. For $\varepsilon > 0$ and $t > 0$, Z^h_s is $\mathcal{H} \times \mathcal{B}_{[0,t]} \times \mathcal{F}^o_{t+\varepsilon}$-measurable on $0 \leq s \leq t$, where $\mathcal{B}_{[0,t]}$ are the Borel sets of $[0,t]$.

REMARK. This follows immediately by the same method, since $W^h_{s+}\, I_{[0,T_h \wedge (t+\varepsilon))}$ is $\mathcal{F}^o_{t+\varepsilon}$-measurable for each s. It does <u>not</u> follow, however, that Corollary 1.7 holds if $\mathcal{F}^o_{t+\varepsilon}$ is replaced by $\mathcal{F}^o_{t+}$.

We next examine how to recover the process $X_t = (w_{2n}(t))$ from Z^h_t. In principle, this is possible because $Z^h_t\{(w_{2n}(0)) = X_t\} = 1$ for each t, h-a.s.

DEFINITION 1.8. Let a mapping from H into $\bar{R}_\infty$ be defined by

$$\varphi(h) = (\rho^{-1}(E^h\, \rho(w_{2n}(0))))_{1 \leq n} .$$

It is easy to see that $\varphi(h)$ is $\mathcal{H}/\bar{\mathcal{B}}_\infty$-measurable. Now we have

THEOREM 1.9. For $h \in H$,

$$P^h\{\varphi(Z^h_t) = X_t \text{ for all } t \geq 0\} = 1 .$$

PROOF. Since φ is a Borel function and the components of X_t are right-continuous, both sides of the equality are $\mathcal{F}^h_t$-optional. By the usual section theorem, it suffices to prove that for each optional $T < \infty$ we have $P^h\{\varphi(Z^h_T) = X_T\} = 1$. But for $n \geq 1$, by Theorem 1.7 b) we have

$$\rho(w_{2n}(T)) = E^h(\rho(w_{2n}(0)) \cdot \theta_T \mid \mathcal{F}^h_{T+})$$

$$= E^{Z^h_T}\, \rho(w_{2n}(0)) , \quad P^h \text{ a.s.}$$

Applying ρ^{-1} to both sides, we obtain the identity for the components of X_T, completing the proof.

REMARK. It follows in particular that $\{Z_s^h, \ 0 \le s \le t\}$ generates a σ-field whose completion for P^h contains that of F_t^o. Consequently, by Theorem 1.7b), for any $0 \le s \le \dots < s_n$ and $B_1,\dots,B_n \in \bar{B}_\infty$, we have easily

$$P^h(\cap_{k=1}^n \{w(s_k) \in B_k\} | F_{t+}^o)$$

$$= P^h(\cap_{k=1}^n \{w(s_k) \in B_k\} | Z_s^h, \ s \le t) \ .$$

But then, by obvious monotone extension, we have $P^h(S|F_{t+}^o) = P^h(S|Z_s^h, \ s \le t), \ S \in G^o$. Gence it follows that the augmentation of $\sigma(Z_s^h, \ s \le t)$ by all P^h-null sets is F_{t+}^h.

We turn now to a basic homogeneity property first proved by P.-A. Meyer and M. Yor ([12] and [13]), which is also the key to their proof of the Markov property of Z_t^h. The proof we give is new in that it avoids Theorem 1 of [13], which was in the nature of an amendment to [12].

Here and in the sequel, we will use where convenient the following abbreviation.

NOTATION 1.10. Let $Z_t^h f$ denote $E^{Z_t^h} f$.

THEOREM 1.11. For each h and F_{t+}^o-stopping time $T < \infty$, we have

$$Z_{T+t}^h = Z_t^{Z_T^h} \cdot \theta_T \quad \text{for all} \quad t \ge 0, \quad P^h \text{ - a.s.}$$

where θ_T on the right does not apply to the superscript $Z_T^h(w)$.

PROOF. We first observe that, for $f \in U$, $Z_{T+t}^h f$ and

$$(Z_t^{Z_T^h} f) \cdot \theta_T$$

are right-continuous in t. Therefore, to prove Theorem 1.11 it suffices to show that these are equal for all f in the sequence (f_n) of Notation 1.4 and t in a countable dense set $\{t_k > 0\}$ such that, for each n, $Z_{T+t}^h f_n$ is continuous at $t = t_k$, P^h - a.s. Such t_k exist since $Z_{T+t}^h f_n$ is r.c.l.l. in t. Thus Z_{T+t}^h is continuous at $t = t_k$, P^h-a.s., and since $T + t_k$ is previsible Theorem 1.7a), b) show that $F_{T+t_k-}^o$ and $F_{T+t_k+}^o$ differ at most by P^h-null sets. Since we have

$$F_{T+t_k-}^o \subset F_{T+t_k}^o \subset F_{T+t_k+}^o$$

$F_{T+t_k}^o$ may be included in this equivalence.

Next, we note that for any t the two sides of Theorem 1.11 are F^o_{T+t+}-measurable. The left side is clearly so. As for the right side, by Corollary 1.7, for each t and $\varepsilon > 0$, Z^h_t is $(H \times F^o_{T+\varepsilon})/H$ - measurable and θ_T is $F^o_{T+\varepsilon+t}/F^o_{t+\varepsilon}$ - measurable, whence by composition $Z^h_t \cdot \theta_T$ is $(H \times F^o_{T+t+\varepsilon})/H$ - measurable. Since also Z^h_T is F^o_{T+}/H - measurable, by composing again it follows that

$$Z_t^{Z_T^h} \cdot \theta_T$$

is $F^o_{T+t+\varepsilon}$ - measurable, and thus F^o_{T+t+} - measurable. Since $F^h_{T+t_k^+} = F^h_{T+t_k}$, the proof of Theorem 1.11 is thus reduced to showing

$$E^h(Y\, Z^h_{T+t_k} f) = E^h(Y(Z_{t_k}^{Z_T^h} f) \cdot \theta_T) \tag{1.10}$$

for each $Y \in b(F^o_{T+t_k})$ and all $f \in \{f_n\}$ used above.

To prove (1.10) we need two simple lemmas.

LEMMA 1.12. $F^o_{T+t_k}$ is contained in the σ-field F_Y generated by all Y of the form

$$Y = (b \cdot \theta_T)g \ , \quad g \in b(F^o_{T+}) \ , \quad b \in b(F^o_{t_k}) \ .$$

PROOF. It is easily seen from Galmarino's Test [4, IV, (100)] that $F^o_{T+t_k}$ is generated by the stopped process $X_{(T+t_k) \wedge s}$, $0 \le s$. Hence we need only show that for each s this is in the σ-field F_Y. Clearly we have $F^o_{T+} \subset F_Y$, $X_{s \wedge T} \in F^o_{T+}$, and $T \in F^o_{T+}$. Hence $X_{s\wedge T}\, I_{\{s \le T\}} \in F_Y$. Now we have

$$(T+t_k) \wedge s = \begin{cases} s \wedge T & \text{on } \{s \le T\} \\ T + ((s-T) \wedge t_k) & \text{on } \{s > T\} \end{cases}$$

hence it remains only to consider the case $s > T$. Letting $T_n = j\, 2^{-n}$ on $\{j2^{-n} \le T < (j+1)2^{-n}\}$, $1 \le j$, we have

$$I_{\{s>T\}}\, X_{T+(s-T) \wedge t_k}$$

$$= I_{\{s>T\}}\, (\lim_{n\to\infty} X_{T+(s-T_n)\wedge t_k}) \ .$$

For each n, we can write

$$X_{T+((s-T_n)\wedge t_k)} = X_{(s-j\,2^{-n})\wedge t_k} \circ \theta_T$$

on $\{T_n = j\,2^{-n} < s\}$, which is in F_Y. Then it is easy to see that the above limit on $\{s > T\}$ in F_Y, completing the argument of the Lemma.

REMARK. In fact we have $F^o_{T+t_k} = F_Y$, as is easily checked, but not needed below.

The class of finite linear combinations of Y having the form stated in Lemma 1.12 is closed under multiplication, hence it suffices to prove (1.10) only for Y of this form. Then assuming $Y = (b \circ \theta_T)\, g$, and writing t for t_k, by Theorem 1.7b) we have

$$
\begin{aligned}
(1.11)\qquad & E^h(Y\, Z^h_{T+t}\, f) \\
&= E^h(Y\, E^h(f \circ \theta_{T+t} | F^o_{(T+t)+})) \\
&= E^h(gE^h((b \circ \theta_T)(f \circ \theta_{T+t}) | F^o_T)) \\
&= E^h(gE^{Z^h_T}(b(f \circ \theta_t))) \\
&= E^h(gE^{Z^h_T}(b\, E^{Z^h_T}(f \circ \theta_t | F^o_{t+}))) \\
&= E^h(gE^{Z^h_T}(b\, Z_t^{Z^h_T} f)) \;.
\end{aligned}
$$

To go from here to the right side of (1.10) we need to reintroduce a conditioning by F^o_{T+} on certain occurrences of w. To justify this, we have

LEMMA 1.13. Let $K(w_1, w_2)$ be a bounded, $F^o_{T+} \times F^o$ - measurable function. Then

$$E^{Z^h_T(w)}\, K(w, w_2) = E^h(K(w, \theta_T w(|F^o_{T+})$$

for P^h-a.e. w, where the expectation on the left is with respect to w_2 over Ω.

PROOF. By linearity and an obvious monotone class argument, it suffices to prove this for K of the form $K_1(w_1)\, K_2(w_2)$, $K_1 \in b(F^o_{T+})$, $K_2 \in b(F^o)$. Then $K_1(w)$ factors out on both sides, and the result follows by Theorem 1.7b).

We now apply Lemma 1.13 to the last expression in (1.11) with

$$K(w_1,w_2) = g(w_1)\ b(w_2)(Z_t^{Z_T^h(w_1)} f)(w_2) .$$

This is justified by composition of $Z_t^z f(w)$, which is $\mathcal{H} \times \mathcal{F}^o$ - measurable, with Z_T^h, which is $\mathcal{F}_{T+}^o/\mathcal{H}$ - measurable. We obtain

$$(1.12) \qquad E^h(YZ_{T+t}^h f)$$

$$= E^h(E^{Z_T^h(w)}(g(w)\ b(w_2)(Z_t^{Z_T^h(w)} f)(w_2)))$$

$$= E^h[E^h((b \cdot \theta_t)g\ (Z_t^{Z_T^h} f) \cdot \theta_T) | \mathcal{F}_{T+}^o)]$$

$$= E^h(Y(Z_t^{Z_T^h} f) \cdot \theta_T)$$

proving (1.10), and hence Theorem 1.11.

It is now easy to deduce that the processes Z_t^h are all strong Markov processes with the same Borel transition function. One may describe this as the "<u>intrinsic</u> Markov property." When considering Z_t^h as h varies we frequently write z in place of h.

DEFINITION 1.14. We set

$$q(t,z,A) = P^z(Z_t^z \in A) , \quad z \in H , \quad 0 \le t , \quad A \in \mathcal{H} .$$

Since $Z_t^z(w)$ is $\mathcal{H} \times \mathcal{F}^o$ - measurable, we have by Fubini's Theorem $q(t,z,A) \in \mathcal{F}$ for fixed (t,A). Further, by right-continuity in t, $q(t,z,A) \in \mathcal{B}^+ \times \mathcal{H}$ for fixed $A \in \mathcal{H}$.

THEOREM 1.15. As processes with state space $(H,\mathcal{H})$ and σ-fields $\mathcal{F}_{t+}^o$, the Z_t^h, $h \in H$, are strong-Markov with transition function q.

PROOF. Let $T < \infty$ be an $\mathcal{F}_{t+}^o$ - stopping time. Then by Theorem 1.11 and Lemma 1.13,

$$P^h(Z_{T+t}^h \in A | \mathcal{F}_{T+}^o) = P^h(Z_t^{Z_T^h} \cdot \theta_T \in A | \mathcal{F}_{T+}^o)$$

$$= P^{Z_T^h}(Z_t^{Z_T^h} \in A)$$

$$= q(t,Z_T^h,A) , \quad A \in \mathcal{H} ,$$

as asserted.

Using Theorem 1.15, it is surprisingly easy to obtain a remarkable result of P.-A. Meyer [12] concerning the set H_0 of "non-branching points".

DEFINITION 1.16. Let $H_0 = \{z \in H : P^z\{Z_o^z = z\} = 1\}$ and let $\mathcal{H}_0$ denote the intersections of sets in $\mathcal{H}$ with H_0.

Clearly $H_0 \in \mathcal{H}$, and $\mathcal{H}_0$ is its topological Borel field. We will see that H_0 is an alternative state space for Z_t^z.

THEOREM 1.17. For $h \in H$, $P^h\{Z_t^h \in H_0$ for all $t \geq 0\} = 1$, and for $h \in H_0$ the processes Z_t^h comprise a Borel right-process on H_0, in the sense of Meyer.

PROOF. The second assertion follows immediately from the first and the definition of a right-process (see Getoor [8, (9.7) and (9.4)]) since the Z_t^h are right-continuous, $P^h\{Z_0^h = h\} = 1$ on H_0, and q is Borel.

The first assertion is really a familiar consequence of the strong Markov property. To prove it, we introduce momentarily for fixed $h \in H$ a canonical version (P, Z_t) of Z_t^h on the space of r.c.l.l. paths with values in $(H, \mathcal{H})$, and let θ_t^Z denote the usual translation operators, and $\mathcal{F}_t^P$ the usual right-continuous, P-augmented σ-fields on this space. Of course by Theorem 1.15 this makes sense, and Z_t remains a strong-Markov process on this space, with transition function q. Also, Z_t is $\mathcal{F}_t^P$ - optional, hence by the section theorem it suffices to show that for optional $T < \infty$, $P\{Z_T \in H_0\} = 1$. But since $Z_0 \cdot \theta_T^Z = Z_T$ and Z_T is $\mathcal{F}_{T+}^P/\mathcal{H}$ - measurable, the strong-Markov property implies

$$
\begin{aligned}
P\{Z_T \in H_0\} &= P\{q(0, Z_T, \{Z_T\}) = 1\} \\
&= P\{P(Z_0 \cdot \theta_T^Z = Z_T | \mathcal{F}_{T+}^P) = 1\} \\
&= 1 ,
\end{aligned}
$$

which completes the proof.

REMARK. Usage of this sample space is introduced systematically in the following section. Here it was used only for notational convenience, because we do not have $Z_0^h \cdot \theta_T = Z_T^h$.

We turn finally to the "moderate Markov property" of the left-limit processes Z_{t-}^h, $t > 0$, in the terminology of Chung and Walsh [2]. This was anticipated by Theorem 1.7 c), and provides a "practical" form of the Markov property in the sense that it can be applied without knowledge of the future (unlike Theorem 1.15).

THEOREM 1.18. For $h \in H$, let T be an $\mathcal{F}_{t+}^h$ - previsible stopping time with $0 < T < \infty$. Then for $t \geq 0$ and $A \in \mathcal{H}$,

$$P^h(Z^h_{T+t} \in A \mid F^h_{T-}) = q(t, Z^h_{T-}, A) \ , \quad P^h\text{-a.s.}$$

PROOF. By [4, IV, Theorem 78] there is an F^o_{t+} - previsible stopping time T^o equal P^h - a.s. to T, and it suffices to prove the assertion for T^o. Further, by [4, IV, Theorem 71] we can as well assume there is an increasing sequence $T^o_n < T^o$, $P^h\{\lim_{n\to\infty} T^o_n = T^o\} = 1$. We may replace T^o by $\lim_{n\to\infty} T^o_n$ in proving the assertion, and then $F^o_{T^o-} = \bigvee_n F^o_{T^o_n}$, where the T^o_n are F^o_{t+} - stopping times [4, IV, Theorem 56]. Again by [4, IV, Theorem 78], $F^h_{T^o-}$ and $F^o_{T^o-}$ differ only by P^h-null sets, so it is enough to prove the result conditional on $F^o_{T^o-}$. In short, we have shown that the entire assertion is equivalent to that obtained by replacing T by a strict limit of an increasing sequence of F^o_{t+} - stopping times, and might as well have been so formulated (except for the fact that F^h_{t+} - stopping times are needed in applications).

We need to use the analogue of Theorem 1.11 for previsible stopping times.

LEMMA 1.19. For $h \in H$, and F^o_{t+} - previsible T with $P^h\{0 < T < \infty\} = 1$, we have

$$Z^h_{T+t} = Z_t^{Z^h_{T-}} \circ \theta_T \quad \text{for all } t \geq 0, \quad P^h \text{ - a.s.}$$

PROOF. As for Theorem 1.11, the problem reduces to showing the analogue of (1.10) with Y, f, and t_k as before:

$$(1.13) \qquad E^h(Y\, Z^h_{T+t_k} f) = E^h(Y (Z_{t_k}^{Z^h_{T-}} f) \circ \theta_T) \ .$$

To do this, we need only apply the analogues of Lemmas 1.12 and 1.13. The latter is proved just as before, and we have

$$E^{Z^h_{t-}(w)} K(w, w_2) = E^h(K(w, \theta_T w) \mid F^o_{T-}) \ ,$$

for $F^o_{T-} \times F^o$ - measurable $K(w_1, w_2)$. As to the former, where $g \in b(F^o_{T+})$ is replaced by $g \in b(F^o_{T-})$, the same proof applies except that we must use the familiar fact that $T \in F^o_{T-}$ for previsible T. Then we write

$$(T+t_k) \wedge s = \begin{cases} s \wedge T & \text{on } \{s < T\} \\ T + ((s-T) \wedge t_k) & \text{on } \{s \geq T\} \end{cases}$$

where $X_{s \wedge T} I_{\{s < T\}} = X_s I_{\{s < T\}} \in F^o_{T-}$ (by definition of F^o_{T-}) . For the second case, there is no change on $\{s > T\}$. Finally, on $\{s = T\}$ we have $X_s I_{\{s=T\}} = (X_0 \cdot \theta_T) I_{\{s=T\}}$, which also has the required form.

We now apply the two lemmas, along with Theorem 1.7 c), replacing F^o_{T+} and Z^h_{T+} by F^o_{T-} and Z^h_{T-}, to obtain first the analogue of (1.11) and next the analogue of (1.12). This then completes the proof of Lemma 1.19.

To complete the proof of Theorem 1.18, one need only apply Lemma 1.19 and the analogue of Lemma 1.13 to obtain

$$\begin{aligned} P^h(Z^h_{T+t} \in A \mid F^o_{T-}) &= P^h(Z_t^{Z^h_{T-}} \cdot \theta_T \in A \mid F^o_{T-}) \\ &= P^{Z^h_{T-}}(Z_t^{Z^h_{T-}} \in A) \\ &= q(t, Z^h_{T-}, A) \ , \end{aligned}$$

completing the proof.

2. PREDICTION SPACES AND RAY TOPOLOGIES.

As already became apparent (in the proof of Theorem 1.17 for example) it is a technical obstacle to have to define Z^z_t separately for each z . Furthermore, in view of Theorems 1.9 and 1.15 it is an unnecessary obstacle. We have $X_t = \varphi(Z^z_t)$ for all t, except on a fixed set $N \in F^o$ with $P^z(N) = 0$ for $z \in H$. Thus we are free to transfer the Z^h_t to a more convenient sample space, and study X_t in terms of Z instead of conversely. This leads to the following concepts and definitions.

DEFINITION 2.1 1) The <u>prediction space</u> of $(\Omega, G^o, F^o_{t+}, G^o_{t+}, \theta_t, X_t)$ consists of $(\Omega_Z, Z^o, Z^o_t, \theta^Z_t, Z_t)$ where

i) Ω_Z is the set of all right-continuous H_0-valued paths $w_z(t)$, $t \geq 0$, with left limits $w_Z(t-)$ in H for $t > 0$.

ii) Z^o_t is the σ-field generated by $\{w_Z(s),\ s \leq t\}$; $Z^o = \bigvee_t Z^o_t$.

iii) $\theta^Z_t : \Omega_Z \to \Omega_Z$ is defined by $\theta^Z_t w_Z(s) = w_Z(s+t)$; $0 \leq s,t$.

iv) $Z_t(w_Z) = w_Z(t)$, $0 \leq t$.

2) The <u>prediction process</u> (without specification of a fixed probability) is the canonical process Z_t on prediction space with transition function $q(t,z,A)$, as justified by Theorems 1.15, 1.17, and 1.18. Thus, Z_t is a strong Markov process and Z_{t-} is a moderate Markov process on H ("process" being meant in the sense of E. B. Dynkin [6]),

and it is a right-process on H_0 when considered only for initial distributions concentrated on H_0. In both cases, we have the same augmented right-continuous σ-fields $\mathcal{Z}_t$ containing all P^μ - null sets in the P^μ completion of $\mathcal{Z}^o$ for all permissible μ, since by right-continuity of path at $t = 0$ every μ on H indices a μ_o on H_0 with $P^\mu = P^{\mu_o}$. Finally, in view of Definition 1.8 and Theorem 1.9, for each $h \in H$ the processes $(\varphi(Z_t), Z_t)$ are jointly P^h-equivalent in distribution to (X_t, Z_t^h), and both components are P^h - a.s. right-continuous with left limits in their respective topologies. It is to be noted that we use the same notation P^h for a probability on either $(\Omega, \mathcal{G}^o)$ or $(\Omega_Z, \mathcal{Z}^o)$, the distinction being clear from the context.

3) By a <u>packet</u> of the prediction process we mean a non-void universally measurable subset U of H such that, for all $h \in U$, $P^h\{Z_t \in U$ for all $t \geq 0\} = 1$ in the sense of outer measure. If $U \in \mathcal{H}$, then U is a "Borel packet", while if $U \subset H_0$ it is an "H_0 packet". We say that a packet U is "complete" if $P^h\{Z_{t-} \in U$ for all $t > 0\} = 1$ for $h \in U$.

REMARKS. Given a packet U, it is clear that $U \cap H_0$ is an H_0 - packet, and on an H_0 - packet Z_t is a right process in the sense of Getoor [8]. But completeness may be lost in this operation, and on a complete packet one has the moderate Markov property of Z_{t-}. In anticipation of things to follow, we point out that starting with a process Z_t, or collection of such (i.e., of P's on $(\Omega, \mathcal{G}^o)$), it is often possible to find a packet which contains the given process (or processes), but little or nothing superfluous. This is beneficial in applying the prediction process.

As a first step in the construction of packets, we prove

THEOREM 2.1. a) Given any non-void subset $A \subset H$, let R_A be a Borel subset of H (i.e., $R_A \in \mathcal{H}$) with $P^z\{Z_t \in R_A$ for all $t > 0\} = 1$ for $z \in A$. Then the set $H_A = \{h \in H : P^h\{Z_t \in R_A, \ t > 0\} = 1$ is a packet, with $A \subset H_A$.

b) For each $h \in H_A$, there is a Borel packet H_h with $h \in H_h \subset H_A$, and further $H_h \supset \{z \in H : P^z\{Z_t \in H_h$ for all $t \geq 0\} = 1\}$.

c) The packet H_A of a) is complete.

PROOF. Let $T = \inf\{t > 0 : Z_t \in H_0 - R_A\}$ be the hitting time of R_A^c on Ω_Z. Then T is $\mathcal{Z}(=\vee_t \mathcal{Z}_t)$ - measurable, and for $\alpha > 0$ the function $E^h(\exp - \alpha T)$ is α-excessive for the right-process on H_0 (as in [1, I, (2.8)]). Further, we have for any $\alpha > 0$ $H_A \cap H_0 = \{h \in H_0 : E^h(\exp -\alpha T) = 0\}$, which (since the right-process has a Borel transition function) is a nearly-Borel set [8, (9.4) (i)]. Since we have $H_A = \{h : q(0,h,H_Z \cap H_0) = 1\}$, it follows that H_A is nearly Borel in H. Hence it is universally measurable. Also, for $h \in H$ the process $E^{Z_t}(\exp -\alpha T)$ is h - a.s. right-continuous. For $h \in H_A$, $e^{-\alpha t} E^{Z_t}(\exp - \alpha T)$ is thus a positive right-continuous supermartingale starting at 0. Hence it is 0 for all t, and H_A is a packet.

Turning to the proof of b), we use a familiar reasoning due to P.-A. Meyer. Since H_A is nearly Borel, for $h \in H_A$ there is a Borel set H_h^1 with $H_1^h \subset (H_A^1 \cap H_0)$ and $P^h\{Z_t \in H_h^1$ for all $t\} = 1$. Then by the same reasoning as for part a) the set $H_h^2 = \{z \in H_h^1 : P^z\{Z_t \in H_h^1,\ t > 0\} = 1\}$ is a packet with $P^h\{Z_0 \in H_h^2\} = 1$. Similarly, we define by induction a sequence $H_A \supset H_h^1 \supset H_h^2 \supset \ldots$ such that for all n, $H_h^{2n-1} \in \mathcal{H}$ and H_h^{2n} is a packet with $P^h\{Z_0 \in H_h^{2n}\} = 1$. Then plainly $H_h^\infty = \bigcap_n H_h^{2n} = \bigcap_n H_h^{2n-1}$ defines a Borel packet and $P^h\{Z_0 \in H_h^\infty\} = 1$. Finally, we set $H_h = \{z : P^z\{Z_0 \in H_h^\infty\} = 1\}$. Then H_h is a Borel packet, $h \in H_h$, and if $P^z\{Z_t \in H_h$ for $t \geq 0\} = 1$ then obviously $z \in H_h$.

Before proving c), we mention two simple Corollaries.

COROLLARY 2.2. For any probability μ on H_A, there is a Borel packet $H_\mu \subset H_A$ with $P^\mu\{Z_0 \in H_\mu\} = 1$, and further $H_\mu \supset \{z \in H : P^z\{Z_t \in H_\mu$ for all $t \geq 0\} = 1\}$.

PROOF. By definition of nearly - Borel set, there is an $H_\mu^1 \subset H_A \cap H_0$, $H_\mu^1 \in \mathcal{H}$, with $P^\mu\{Z_t \in H_\mu^1$ for all $t\} = 1$. Then as in part a) the set $H^2 = \{z \in H_\mu^1 : P^z\{Z_t \in H^1,\ t \geq 0\} = 1\}$ is a packet, and $P^\mu\{Z_0 \in H_\mu^2\} = 1$. Proceeding by induction as in b), we obtain a decreasing sequence $H_\mu^n \subset H_A \cap H_0$ with $H_\mu^{2n-1} \in \mathcal{H}$, and H^{2n} a packet such that $P^\mu\{Z_0 \in H_\mu^{2n}\} = 1$. Now let $H_\mu^\infty = \bigcap_n H^n$, and $H_\mu = \{z : P^z\{Z_0 \in H_\mu^\infty\} = 1\}$.

COROLLARY 2.3. For any packet K such that $K \cap H_0 = H_A \cap H_0$, we have $K \subset H_A$. Thus H_A is the largest packet having the given non-branching points of H_A.

PROOF. For any packet K, one has $q(0,z,H_A \cap H_0) = 1$ for $z \in K$. But it follows by the definition of H_A, using the Markov property again, that H_A contains all z with $q(0,z,H_A \cap H_0) = 1$. Thus the Corollary is proved.

REMARK. We observe that for any initial probability μ on H_A, an element h of H_A is defined by

$$h(S) = \int_{H_A} \int_{H_0} q(0,h,dy)y(S)\mu(dh)$$

$$= \int_{H_0} \left(\int_{H_A} q(0,h,dy)\mu(dh)\right) y(S) \ , \quad S \in \mathcal{G}_0 \ ,$$

where the probability in parentheses is concentrated on $H_0 \cap H_A$.

Returning now to the proof of Theorem 2.1 c), for $h \in H_A$ let $H_h \subset H_A$ be a Borel packet as in b). We wish to show that $P^h\{Z_{t-} \in H_A$ for all $t > 0\} = 1$, and we know that $P^h\{Z_t \in H_h$ for all $t \geq 0\} = 1$. Now $1_{H_h}(Z_{t-})$ is a $\mathcal{Z}_t$ - previsible process, and for each previsible stopping time, T, $0 < T < \infty$, we have by the moderate Markov property

$$P^h(Z_t \in H_h \text{ for all } t \geq T | \mathcal{Z}_{T-})$$

$$= P^{Z_{T-}}(Z_t \in H_h \text{ for all } t \geq 0)$$

$$= 1 .$$

Consequently, by b) we have $P^h\{Z_{T-} \in H_h\} = 1$. By the previsible section theorem it follows that $P^h\{I_{H_h}(Z_{t-}) = 1$ for all $t > 0\} = 1$, and so $P^h\{Z_{t-} \in H_A$ for all $t > 0\} = 1$ as required.

A natural question is whether, given a set $A \in \mathcal{H}$, there is a smallest packet containing it. The example of a Brownian motion $B^2(t)$ in R^2, with $A = \{(0,0)\}$, shows however that no smallest packet need exist. Here the points $(x,y) \in R^2$ correspond to points of R_A via the usual $P^{(x,y)}$, and clearly any polar set may be subtracted from R^2 (but no non-polar set may be subtracted) to leave a packet. It can be shown that in this example H_A is the set of all Brownian probabilities corresponding to initial distributions μ on $(R^2, \mathcal{B}^2)$, but the proof probably requires Ray compactifications (see Discussion 3) and 4) of Conjecture 2.10 below).

It also should be noted that the definition of packet depends only on the transition measures $q(t,h,dz)$ of the prediction process, and these do not depend on the exact choice of Z_t^h (which is not unique since it involves the W_r^h of Definition 1.6). In short, a packet is just a continuous time analogue of "conservative set" for a Markov chain. In the case that the elements of A are themselves Markovian probabilities on Ω (as in the Brownian example above) the measures $q(t,h,\cdot)$, $h \in A$, are usually easy to identify, and the appropriate packet becomes evident.

This leads to a method of finding a "nice" transition function for a Markov process, which is the subject of the third essay. Here we can illustrate it in a more classical case by continuing our example of $B^2(t)$. Let B_0 be a Borel, non-polar set in R^2, and consider the usual killed process $B^2_\Delta(t) = B^2(t)$ for $t < T_{B_0}$, and $B^2_\Delta(t) = \Delta$ for $t \geq T_{B_0}$

where Δ is adjoined as an isolated point. Classically, the probabilities $P^{(x,y)}\{B_\Delta^2(t) \in C\}$ $C \in \mathcal{B}^2$, are only known to be universally measurable in (x,y). Thus one obtains for B_Δ^2 a universally measurable transition function. However, using the prediction process it is easy to get a transition function on a countably generated subfield of universally measurable sets which is the restriction of a Borel transition function on a larger space. The natural state space of B_Δ^2 is Δ together with the (finely open) set $(B_0^r)^c = \{(x,y) : P^{(x,y)}\{T_{B_0} > 0\} = 1\}$, i.e., the complement of the set of regular points for B_0. Since α-excessive functions for B^2 are Borel measurable, and $E^{(x,y)}(\exp -\alpha T_{B_0})$ is α-excessive, it is not hard to show that $(B_0^r)^c$ is a Borel set, but we need only its universal measurability. Identifying $B_\Delta^2(t)$ with $(w_2(t), w_4(t))$, where $\Delta = (\infty,\infty)$ and all other coordinates are set identically 0 for $P_\Delta^{(x,y)}$, we obtain a one-to-one mapping of $(B_0^r)^c \cup \Delta$ into H defined by $(x,y) \to P_\Delta^{(x,y)}$. Let R_Δ denote the image in H. We have

$$R_\Delta = \{z \in H : \varphi(z) \in \{(B_0^r)^c \cup (\infty,\infty)\} \times X_{n=3}^{\infty}(0,0)$$

$$\text{and} \quad z(S) = P_\Delta^{\varphi_2(z)}(S) \;, \quad S \in G^0\},$$

where $\varphi(z)$ is the Borel mapping of Theorem 1.9, and $\varphi_2(z)$ denotes its first two coordinates. Since G^o is countably generated, and $P_\Delta^{(x,y)}$ is universally measurable from $(B_0^r)^c \cup \Delta$ into H, it follows by using a generating sequence S_n in place of S that R_Δ is universally measurable in H. Then the trace of $\mathcal{H}$ on R_Δ is mapped by φ_2 onto a countably generated σ-field of universally measurable sets in $(B_0^r)^c \cup \Delta$, and q on the trace maps by φ_2 into the transition function of B_Δ^2 on the image. In the present case, it can be shown that the image σ-field is really the Borel field, but this seems to require in general Meyer's hypothesis of "absolute continuity".

The theory of Ray processes (and Ray semigroups) is rather well understood, and will not be developed here. We refer instead to Getoor [8] for all of the facts we shall need. By means of the familiar compactification procedure (to be described below) this theory may be brought to bear on any parcel of the prediction process. Thus, it leads to a more satisfactory form of Theorem 2.1 (Corollary 2.12), and also to an interesting open problem (Conjecture 2.10) which is discussed in some detail. It also makes possible a transcription of much of the "comparison of processes" from [8] to the prediction process setting, but some of this we leave to the reader. Part of the material which we do cover is needed again for the fourth essay.

We start with any prediction packet which we denote by H_A for convenience although A alone is unspecified and H_A has no reference in general to Theorem 2.1. It is clear from Theorem 1.17 that Z_t becomes a right process on $H_A \cap H_0$, with the Borel transition function q (even if $H_A \cap H_0$ is not Borel, we have for $z \in H_A \cap H_0$, $q(t,z,B) = q(t,z,B \cap H_A \cap H_0)$ for $B \in \mathcal{H}$, where the right side is the extension to a universally measurable set). Consequently, we may consider $H_A \cap H_0$ as a subset of the compact metric space $\bar{H}$ of Proposition 1.3, and form its Ray compactification (as in Chapter 10 of [8]) relative to $\bar{H}$, which will be denoted by $(H_A \cap H_0)^+$.

The definition of $(H_A \cap H_0)^+$ is as follows. Let C^+ denote the restriction to $H_A \cap H_0$ of non-negative continuous functions on $\bar{H}$. Then C^+ has a countable subset which is dense in C^+ in the uniform norm. Letting $R_\lambda g(z)$ denote the resolvent of the right-process Z_t on $H_A \cap H_0$, we form the minimal set of functions containing $\{R_\lambda g : \lambda > 0,\ g \in C^+\}$ and closed under the two operations:

a) application of R_λ for $\lambda > 0$,

b) formation of minima $f \wedge g$.

Since we have $(f \wedge g) + (h \wedge k) = (f+h) \wedge (g+h) \wedge (f+k) \wedge (g+k)$, it is easy to see by simple induction that the set is closed under formation of linear combinations with non-negative coefficients. Hence, it is the minimal convex cone closed under operations a) and b). A crucial lemma ([8, (10.1)]) now asserts that this cone contains a countable uniformly dense subset. Furthermore, the cone separates points in $H_A \cap H_0$ since R_λ does so.

We now define $(H_A \cap H_0)^+$ to be the compact metrizable space obtained by completing $H_A \cap H_0$ in a metric $\Sigma_{n=1}^\infty \alpha_n |f_n(z_1) - f_n(z_2)|$, where (f_n) is uniformly dense in the cone, $\alpha_n > 0$, and

$$\Sigma_{n=1}^\infty \alpha(\max f_n) < \infty .$$

Clearly the topology of $(H_A \cap H_0)^+$ does not depend on the particular choice of f_n or α_n. It is homeomorphic to the closure of the image of $H_A \cap H_0$ in $\times_{n=1}^\infty [0,\infty)$ by the function $f(z) = (f_1(z), f_2(z), \ldots)$. If H_A is Borel, then its one-to-one image in $(H_A \cap H_0)^+$ is also Borel, while in general its image is universally measurable [8, (11.3)].

It is now easy to see by the Stone-Weierstrass Theorem that the space $C(H_A \cap H_0)^+$ of continuous functions on $(H_A \cap H_0)^+$ is the uniform closure of the differences $g_1 - g_2$ of elements of the cone, extended to $(H_A \cap H_0)^+$ by continuity. Letting f denote a uniform limit of such differences on $H_A \cap H_0$, and $\bar{f}$ its extension by continuity to $(H_A \cap H_0)^+$, we now define a resolvent on $C(H_A \cap H_0)^+$, by

$$\bar{R}_\lambda \bar{f} = \overline{R_\lambda f} \;, \quad \bar{f} \in C(H_A \cap H_0)^+ \;, \quad \lambda > 0 \;. \tag{2.2}$$

The resolvent $\bar{R}_\lambda$ has the special property that it carries $C(H_A \cap H_0)^+$ into itself. Finally, one shows [8, (10.2)] that every element of the cone is λ-excessive for some $\lambda > 0$, hence $\bar{R}_\lambda$ separates points and so $\bar{R}_\lambda$ is a Ray resolvent on $(H_A \cap H_0)^+$.

It follows by a Theorem of D. Ray that there is a unique right-continuous Markov semigroup $\bar{P}_t$ on $C(H_A \cap H_0)^+$ with resolvent $\bar{R}_\lambda$, whose transition measures we denote by $\bar{p}(t,h,dz)$. We also introduce the Ray Space (of Z_t on $H_A \cap H_0$) as in [8, Chapter 15].

DEFINITION 2.4. The Ray Space is the set

$$U_A = \{ \bar{z} \in (H_A \cap H_0)^+ : \lambda \bar{R}_\lambda \; I_{H_A \cap H_0} (\bar{z}) = 1 \} \;.$$

REMARKS. More properly, one should write $U_{H_A \cap H_0}$, but no confusion will arise. It is clear that U_A does not depend on $\lambda > 0$, and that it is universally measurable in $(H_A \cap H_0)^+$. If $H_A \in \mathcal{H}$ then U_A is also Borel.

Three basic facts about $\bar{P}_t$ from [8, Chapter 15] which serve to connect $\bar{P}_t$ with the prediction process may be summarized as follows.

PROPOSITION 2.5.

1. For $z \in H_A \cap H_0$ and $\bar{f} \in C(H_A \cap H_0)^+$ we have $\bar{P}_t \bar{f}(z) = Q_t f(z)$. Thus $\bar{p}$ and q may be identified on $H_A \cap H_0$.

2. For $z \in U_A$ we have $\bar{P}_t(I_{H_A \cap H_0}(z)) = 1$ for $t > 0$ (where $\bar{P}_t$ is defined for universally measurable functions by the usual extension procedure).

3. For the canonical Ray process $(\bar{X}_t, \bar{P}^z)$ on the probability space of r.c.l.l. paths with values in $(H_A \cap H_0)^+$, we have for $z \in U_A$

$$\bar{P}^z\{\bar{X}_t \in H_A \cap H_0 \text{ for all } t > 0\} = 1 \;,$$

and

$$\bar{P}^z\{\bar{X}_{t-} \in U_A \text{ for all } t > 0\} = 1 \;.$$

Recalling again the space $\bar{H}$ of probabilities on the compact metrizable space $\bar{\Omega}$ of equivalence classes of measurable functions, we will show that the Ray topology is stronger on $H_A \cap H_0$ than the $\bar{H}$-topology. Hence $(H_A \cap H_0)^+$ is "saturated" by the equivalence classes of elements corresponding to the same element in $\bar{H}$, and these classes reduce to single elements on $H_A \cap H_0$. Furthermore, on U_A the corresponding elements of $\bar{H}$ have a special form: they assign probability one to paths which are r.c.l.l. for $t > 0$. Only the right-limits at $t = 0$ are not known to exist, hence the mapping does not quite have its range in H. Nevertheless, it is sufficient to permit properties of the Ray process to be applied to the process Z^h_{t-} for $h \in H_A \cap H_0$.

Turning to the details, we first characterize convergence in $\bar{H}$ by

LEMMA 2.6. A sequence $h_k \in H$ is Cauchy in $\bar{H}$ if and only if, for the dense sequence f_n of Notation 1.4,

$$E^{h_k} \int_0^\infty \exp(-\beta t)\, f_n \cdot \theta_t \, dt$$

is a real Cauchy sequence in k for each n and $\beta > 0$.

PROOF. By Theorem 1.2 b) the integrals are uniformly continuous on $\bar{\Omega}$. Hence our condition is clearly necessary. To prove sufficiency, we observe by the same result that

$$E^{h_k} f_n \cdot \theta_t$$

are uniformly continuous and bounded in t, uniformly in k, for each n. Then by inversion of the Laplace transforms (as in Lemma 1.6)

$$\int_0^\infty \exp(-\beta t)\, E^{h_k} f_n \cdot \theta_t \, dt$$

we have convergence in k of

$$E^{h_k} f_n \cdot \theta_t$$

for each $t \geq 0$ and n. For $t = 0$ this reduces to convergence of (h_k) in $\bar{H}$, as required.

Using the Lemma, we may compare the Ray and $\bar{H}$-topologies.

THEOREM 2.7. If we have $h_k \in H_A \cap H_0$, $1 \leq k$, and $\lim_{k\to\infty} h_k = z$ exists in the topology of $(H_A \cap H_0)^+$, then $\lim_{k\to\infty} h_k = h$ exists in the topology of $\bar{H}$. Furthermore, let $h(z)$ denote the induced mapping: $h(z) = z$ on $H_A \cap H_0$, $h(z) = h$ if $z \notin H_A \cap H_0$ and (z,h) correspond as above. Then $h(z)$ is continuous on $(H_A \cap H_0)^+$. Finally, for $z \in U_A$ we have $P^{h(z)}\{\text{paths r.c.l.l. for } t > 0\} = 1$.

PROOF. Let $h_k \in H_A \cap H_0$ be a convergent sequence in the Ray topology, with limit $z \in (H_A \cap H_0)^+$. This requires convergence of $R_\lambda g(h_k)$ for $g \in C^+$. Still more particularly, let $g(z) = E^z f$ (= zf in Notation 1.10) for $0 \le f \in C(\bar{\Omega})$. Then we have

$$R_\lambda g(h_k) = E^{h_k} \int_0^\infty e^{-\lambda t} E^{Z_t} f \, dt \tag{2.3}$$

$$= E^{h_k} \int_0^\infty e^{-\lambda t} f \cdot \theta_t \, dt, \quad \lambda > 0 .$$

Thus convergence in the Ray topology implies convergence in the topology of $\bar{H}$ by Lemma 2.6. Accordingly, there is a unique $h(z) \in \bar{H}$ such that $h_k \to h(z)$. Since $H_A \cap H_0$ is dense in $(H_A \cap H_0)^+$, the mapping $h(z)$ is well-defined and continuous: $(H_A \cap H_0)^+ \to \bar{H}$, and reduces to the identity on $H_A \cap H_0$.

We will examine more closely the case $z \in U_A$. Passing to the limit in (2.3) yields

$$\overline{R_\lambda g}(z) = E^{h(z)} \int_0^\infty e^{-\lambda t} f \cdot \theta_t \, dt , \tag{2.4}$$

but the middle term in (2.3) is no longer well-defined in the limit if $h(z) \notin H$ (in the context of [9], Z_t becomes the prediction process on $\bar{H}$) . However, the same limit may be expressed in terms of the Ray process $\bar{X}_t$ of Proposition 2.5, since $\bar{X}_t = h(\bar{X}_t)$ on $H_A \cap H_0$. To this end, we need to establish

LEMMA 2.7. For $g(z) = E^z f$, f continuous on $\bar{\Omega}$, and $z \in U_A$, we have

$$\overline{R_\lambda g}(z) = \bar{E}^z \int_0 e^{-\lambda t} E^{\bar{X}_t} f \, dt .$$

REMARK. This was also used for [10, Theorem 2.4 d)] with incomplete proof.

PROOF. For $\beta > 0$, the function $\bar{R}_\beta \overline{R_\lambda g}$ is β-excessive for $\bar{X}_t$, hence it is known [8, (5.8)] that

$$\lim_{t \to 0} \bar{R}_\beta \overline{R_\lambda g} (\bar{X}_t) = \bar{R}_\beta \overline{R_\lambda g}(z) , \quad \bar{P}^z\text{-a.s.}$$

Also, by (2.2) and the resolvent equation,

$$\lim_{\beta \to \infty} \beta \bar{R}_\beta \overline{R_\lambda g} = \lim_{\beta \to \infty} \beta \overline{R_\beta R_\lambda g}$$

$$= \lim_{\beta \to \infty} (\beta/(\beta - \lambda)) (\overline{R_\lambda g} - \overline{R_\beta g})$$

$$= \overline{R_\lambda g} ,$$

and the limit is uniform on $(H_A \cap H_0)^+$. It follows that limits can be interchanged to obtain

$$\overline{R_\lambda g}(z) = \lim_{\beta\to\infty} \lim_{t\to 0} \beta \bar{E}^z \bar{R}_\beta \overline{R_\lambda g}(\bar{X}_t)$$

$$= \lim_{t\to 0} \lim_{\beta\to\infty} \beta \bar{E}^z \bar{R}_\beta \overline{R_\lambda g}(\bar{X}_t)$$

$$= \lim_{t\to 0} \bar{E}^z \overline{R_\lambda g}(\bar{X}_t) .$$

But since for $t > 0$ we have $\bar{X}_t \in H_A \cap H_0$ $\bar{P}^z$-a.s., the last expression becomes

$$= \lim_{t\to 0} \bar{E}^z R_\lambda g(\bar{X}_t)$$

$$= \lim_{t\to 0} \bar{E}^z E^{\bar{X}_t} \int_0^\infty e^{-\lambda s} E^{\bar{X}_s} f \, ds$$

$$= \lim_{t\to 0} \bar{E}^z \int_0^\infty e^{-\lambda s} E^{\bar{X}_{t+s}} f \, ds$$

$$= \lim_{t\to 0} \bar{E}^z e^{\lambda t} \int_t^\infty e^{-\lambda s} E^{\bar{X}_s} f \, ds$$

$$= \bar{E}^z \int_0^\infty e^{-\lambda t} E^{\bar{X}_t} f \, dt ,$$

completing the proof.

Combining this with (2.4) yields

$$\bar{E}^z \int_0^\infty e^{-\lambda t} E^{\bar{X}_t} f \, dt = E^{h(z)} \int_0^\infty e^{-\lambda t} f \cdot \theta_t \, dt . \tag{2.5}$$

Since $\bar{X}_t$ is right-continuous in the Ray topology, which we have seen is stronger than the $\bar{H}$-topology on H, $\bar{E}^z E^{\bar{X}_t} f$ is right-continuous in $t > 0$. By Theorem 1.2 b), $E^{h(z)} f \cdot \theta_t$ is also right-continuous. Thus by inversion of the transforms in (2.5) we obtain $\bar{E}^z E^{\bar{X}_t} f = E^{h(z)} f \cdot \theta_t$, $t > 0$, for $0 \le f$ continuous on $\bar{\Omega}$. By Proposition 2.5.1., the left side is

$$\bar{E}^z E^{\bar{X}_t}(f I_\Omega) .$$

By monotone class argument the equality extends to bounded Borel f, hence it follows that the right side is $E^{h(z)}((f I_\Omega) \cdot \theta_t)$. This implies that for $t > 0$, $P^{h(z)}\{\text{paths which are r.c.l.l. in } [t,\infty)\} = 1$. Letting $t \to 0$, the last assertion of Theorem 2.7 is proved.

It is thus plausible that for $z \in U_A$ the Ray process may be expressed as a prediction process on a slightly larger space than H, but smaller than $\bar{H}$. We introduce

NOTATION 2.8. Let $\Omega_1 = \{\text{elements in } \bar{\Omega} \text{ which are r.c.l.l. for } t > 0\}$, and let $H_1 = \{h \in \bar{H} : h(\Omega_1) = 1\}$, $\mathcal{H}_1 = \{A \cap H_1 : A \in \bar{\mathcal{H}}\}$.

THEOREM 2.9. Let $\mathcal{F}^o_{t+} = \bigcap_{\varepsilon>0} \sigma(X_s, 0 < s < t+\varepsilon)$ on Ω_1. Then for $h \in H_1$, one can define the prediction process Z^h_t, and extend the transition function q to $(H_1, \mathcal{H}_1)$, in such a way that Theorems 1.7 and 1.15 remain true. Setting $H_{01} = \{h \in H_1 : P^h\{Z^h_0 = h\} = 1\}$, Theorem 1.17 also applies for $h \in H_1$, if H_0 is replaced by H_{01}.

REMARK. Note that X_0 is not $\mathcal{F}^o_{0+}$-measurable. This conforms to the fact that as a "coordinate in $\bar{\Omega}$", X_0 is not even well-defined. The meaning of X_t for $t > 0$ is really in the sense of an essential right limit, which happens to coincide with X_t since $X_{t+} = X_t$.

PROOF. We will not elaborate all details, Z^h_t is just a special case of the prediction process on $\bar{H}$ of [9]. The point is that, since $P^h(\theta^{-1}_t \Omega | \mathcal{F}^o_{t+}) = 1$ for $t > 0$, we can use exactly the same σ-fields $\mathcal{G}^o$ and the same construction as before to define Z^h_t for $t > 0$, to show that $P^h\{Z^h_t \in H_0 \text{ for } t > 0\} = 1$, and to show that the same transition function q continues to apply for Z^h_t, $t > 0$. On the other hand, for f continuous on $\bar{\Omega}$ it follows by Hunt's Lemma that for rationals $r > 0$,

$$\lim_{r\to 0+} Z^h_r f = \lim_{r\to 0+} E^h(f \cdot \theta_r | \mathcal{F}^o_{r+})$$

$$= E^h(f | \mathcal{F}^o_{0+}) , \quad P^h\text{-a.s.}$$

Since Z^h_t is right-continuous for $t > 0$, we see that $\lim_{t\to 0} Z^h_t = Z^h_0$ exists in the topology of $\bar{H}$, P^h-a.s., and

$$E^h(S | \mathcal{F}^o_{0+}) = Z^h_0(S) , \quad S \in \mathcal{G}^o(\bar{\Omega}) .$$

Now if we define $q(t,h,A) = P^h\{Z^h_t \in A\}$ for $h \in H_1 - H$, $A \in \mathcal{H}_1$, and $q(t,h,A) = q(t,h,A\cap H_0)$ for $h \in H$, the Markov property of Z^h_s for $s > 0$ implies that $q(s+t,h,A) = \int q(s,h,dz)\, q(t,z,A)$ for all $s > 0$ and $t \geq 0$. On the other hand, for $s = 0$ we have for $t \geq 0$

$$q(t,h,A) = P^h\{Z^h_t \in A\}$$

$$= E^h(P^h(Z^h_t \in A | \mathcal{F}^o_{0+}))$$

$$= E^h P^{Z_0^h}(Z_t^h \in A)$$

$$= \int q(0,h,dz)\; q(t,z,A) \;,$$

completing the verification of the Chapman-Kolmogorov property of q. Since $H_0 \subset H_{01}$, it only remains to verify that for $h \in H_1$, $P^h\{Z_0^h \in H_{01}\} = 1$. Since, by construction, Z_0^h is $\mathcal{F}_{0+}^o/\mathcal{H}_1$-measurable, this last is a consequence of the strong Markov property with $T = 0$. Formally, it follows because

$$1 = P^h\{Z_{0+0}^h = Z_0^h\}$$

$$= E^h P^h(Z_{0+0}^h = Z_0^h \mid \mathcal{F}_{0+}^0)$$

$$= E^h(P^{Z_0^h}\{Z_0^{Z_0^h} = Z_0^h\}) \;,$$

implying that the expression in the last parentheses equals 1, P^h-a.s.

In view of Theorem 2.9, we define the prediction space and prediction process of $(\Omega_1, \mathcal{G}^o, \mathcal{F}_{t+}^o, \theta_t, X_t)$ in complete analogy with Definition 2.1, and it has the same Markov properties noted there. We are now in a position to state an interesting conjecture concerning the relation of this prediction process to the Ray processes (see also Theorem 2.7).

CONJECTURE 2.10. For any packet H_A, $h \in U_A$, and $dy \in (\bar{\mathcal{H}}|H_1)$ let $\mu_h(dy) = \bar{P}^h(h(\bar{X}_0) \in dy)$. Then $\bar{X}_t$ is $\bar{P}^h$-equivalent in distribution for $t > 0$ to the prediction process of $(\Omega_1, \mathcal{G}^o, \mathcal{F}_{t+}^o, \theta_t, X_t)$ on H_1, with initial distribution $\mu_h(dy)$.

DISCUSSION. 1) Since $\bar{X}_t$ has right limits in H_{01} at $t = 0$ in the $\bar{H}$-topology, $\bar{P}^h$-a.s., the conjecture follows if it is shown that the mapping $h(z)$ is one-to-one on the non-Ray-branching points of U_A. The converse implication is also clear.

2) We do not conjecture that $\bar{X}_t$ is $\bar{P}^h$-equivalent to the prediction process of a fixed element of H_1. This is false in general. For example, consider the sequence h_n, $1 \le n$, where h_n is the probability of the process X_t which with probability $1/2$ chooses one of the two paths

$$w_1(t) = n^{-1} + (t-1)^+ \quad\text{or}\quad w_2(t) = -(t-1)^+ \;,$$

where $(t-1)^+ = \max(0,t-1)$. Then in the Ray topology $\lim_{n\to\infty} h_n = h$, where h is the Ray branch-point which with probability $1/2$ gives the prediction process of either of the deterministic processes $X_t = (t-1)^+$ or $X_t = -(t-1)^+$. It is not hard to see that this initial distribution for the prediction process cannot be expressed as $P^z\{Z_0 \in (\cdot)\}$ for any $z \in H_1$. The necessary and sufficient condition for such a representation is contained in Theorem 1.2 of [10]. On the other hand, in the $\bar{H}$-topology $\lim_{n\to\infty} h_n = z$, where $z \in H_0$ is the obvious probability concentrated on 2 points of Ω.

3) The importance of the conjecture, at least from the standpoint of theory, lies in the fact that all entrance laws for the transition function q on $H_A \cap H_0$ (having mass 1) are expressed by initial distributions on $(H_A \cap H_0)^+$ for the Ray process. This fact seems to have first been noted by H. Kunita and T. Watanabe [11, Theorem 1]. Hence, our conjecture is equivalent to the assertion that every (finite) entrance law for the prediction process on $H_A \cap H_0$ is realized by an initial distribution of the prediction process of $(\Omega_1, G^o, F^o_{t+}, \theta_t, X_t)$. Of course, it suffices here to consider the case $H_A \cap H_0 = H_0$. The analogous conjecture for the prediction process of [9] on $\bar{H}$ (or equivalently, on the set $\bar{H}_0 = \{h \in \bar{H} : P^h\{Z_0 = h\} = 1\}$) would be that it is already closed under formation of entrance laws. Hence the Ray space of $\bar{H}_0$ would correspond to a subset of initial distributions over $\bar{H}_0$. It is easily shown that this Ray space does define a process corresponding to each P^h, $h \in \bar{H}$, and by Discussion 1) above it is then strictly larger than $\bar{H}$. The class of processes for $t > 0$ obtained from initial distributions on the Ray space is then the same class as those obtained from all initial distributions on $\bar{H}_0$ (or equivalently, on $\bar{H}$), if this extended conjecture holds.

4) For the packet of an autonomous germ - Markov process, the conjecture holds and $\bar{X}_t$ is even represented by a single element of H_1 (see [10, Theorem 2.4] for a more general setting).

As far as concerns the left-limit process Z_{t-}, it will be seen that the result of Conjecture 2.10 does hold, at least for Borel packets. A still more satisfactory result will be shown subsequently.

THEOREM 2.11. For any Borel H_0-packet $H_A \cap H_0$, let $C_A = \{z \in (H_A \cap H_0)^+ :$ for $\bar{f} \in C(H_A \cap H_0)^+$ with corresponding f as in (2.2), $\bar{R}_\lambda \bar{f}(z) = \int \bar{p}(0,z,dy)\, R_\lambda f(h(y))\}$, where the integral is over $\{y : h(y) \in H_A \cap H_0\}$. Then C_A is Borel in $(H_A \cap H_0)^+$, and for any $z \in U_A$, $P^z\{\bar{X}_{t-} \in C_A$ for all $t > 0\} = 1$.

PROOF. Since $h(y)$ is continuous and $\bar{p}(t,z,dy)$ is a Borel transition function, while $q(t,h(y),A)$, $A \in \mathcal{H}$, is also Borel in the Ray-topology, it is clear by letting $\bar{f}$ range through a countable dense set that C_A is Borel in $(H_A \cap H_0)^+$. Therefore, $I_{C_A}(\bar{X}_{t-})\cdot$ is a previsible process for the Ray σ-fields. To prove the second assertion, it suffices to assume $t \geq \varepsilon$ for some $\varepsilon > 0$, and since $\bar{X}_t$ and Z_t are identified for $t > 0$ we may as well assume $z \in H_A \cap H_0$. Then $\bar{X}_t$ and Z_t are identified for $t \geq 0$, and since the Ray and $\bar{H}$-topologies induce the same σ-fields on $H_A \cap H_0$, we see that $I_{C_A}(\bar{X}_{t-})$ is previsible for the P^z-augmented σ-fields $\mathcal{Z}_t^z$ generated by Z_s, $s \leq t$. By the previsible section theorem, it now is enough to show that for previsible T with $0 < T < \infty$, $P^z\{I_{C_A}(\bar{X}_{T-}) = 1\} = 1$. Since $\bar{X}_{T+t} \in H_A \cap H_0$ for $t \geq 0$, and the Ray processes have the moderate Markov property, it follows that

$$\bar{R}_\lambda \bar{f}(\bar{X}_{T-}) = E^{\bar{X}_{T-}} \int_0^\infty e^{-t\lambda} f(\bar{X}_t)\, dt \tag{2.6}$$

$$= \int \bar{p}(0,\bar{X}_{T-},dy)\, R_\lambda f(y),\ P^z\text{-a.s.}$$

Since $h(y) = y$ on $H_A \cap H_0$, this is the asserted result.

Irrespective of Conjecture 2.10, we can regard C_A as a complete Borel packet in the Ray space, each of whose elements corresponds to an initial distribution on $H_A \cap H_0$. However, a stronger result is evident by comparison of (2.6) with the moderate Markov property of Z_t (Theorem 1.18). Thus the expression in (2.6) must also equal

$$\int_{0+}^\infty \int_{H_A \cap H_0} e^{-\lambda t}\, q(t,Z_{T-},dz) f(z)\ ,$$

since both determine the probabilities of Z_{T+t} given $\mathcal{Z}_{T-}$. Denoting this expression by $R_\lambda f(Z_{T-})$, it follows by the previsible section theorem that for $z \in U_A$,

$$P^z\{\bar{R}_\lambda \bar{f}(\bar{X}_{t-}) = R_\lambda f(Z_{t-}) \text{ for all } t > 0\} = 1\ .$$

But by continuity of $h(z)$ we have

$$Z_{t-} = \lim_{s \to t-} Z_s = \lim_{s \to t-} h(\bar{X}_s) = h(\bar{X}_{t-})\ .$$

Substituting for Z_{t-} in the above, we have shown

COROLLARY 2.12. For any Borel H_0-packet $H_A \cap H_0$, let

$D_A = \{z \in U_A :$ for $\bar{f} \in C(H_A \cap H_0)^+$ with f as in (2.2), $\bar{R}_\lambda \bar{f}(z) = R_\lambda f(h(z))\}$. Then D_A is Borel in $(H_A \cap H_0)^+$, and for $z \in U_A$, $P^z\{\bar{X}_{t-} \in D_A$ for all $t > 0\} = 1$. Finally, the image $h(D_A) \cap H$ is a complete Borel packet in H containing $H_A \cap H_0$.

PROOF. Only the final assertion remains to be shown, since obviously D_A is Borel in $(H_A \cap H_0)^+$. But since z is determined uniquely in $(H_A \cap H_0)^+$ by $\{\bar{R}_\lambda \bar{f}(z)\}$, we see that $h(z)$ is one-to-one on D_A . Hence $h(D_A)$ is Borel in H_1, and $h(D_A) \cap H$ is Borel in H . Since for $z \in h(D_A) \cap H$ we have

$$P^z\{Z_{t-} \in H \text{ and } h(\bar{X}_{t-}) = Z_{t-} \text{ for all } t > 0\} = 1 ,$$

the result is proved.

According to Corollary 2.12, starting from any Borel H_0-packet $H_A \cap H_0$, we can form the complete Borel packet $h(D_A) \cap H$ containing it, all of whose elements determine the same processes as corresponding initial distributions on $h(D_A) \cap H_0$, and have the property that the process Z_t remains in $H_A \cap H_0$ for all $t > 0$. Thus it is quite natural to replace the process on $H_A \cap H_0$ by the right process on $h(D_A) \cap H_0$ with left-limits in $h(D_A) \cap H$. Since $h(z)$ is one-to-one on D_A, we can regard this process equivalently in either the Ray or the H-topology in so far as concerns its times of discontinuity. Thus, there is no need to make an elaborate "comparison of processes," as in [8, Chapter 13] for example. Instead, we can transcribe results for the Ray process directly into results for the H-process. To conclude the present section, let us illustrate this by transcribing Theorem (7.6) of [8, Chapter 7].

THEOREM 2.13. For a Borel H_0-packet $H_A \cap H_0$, let μ be a fixed initial distribution on $H_A \cap H_0$ (or more generally on $h(D_A) \cap H_0$), and let T be a Z_t^μ-stopping time of Z_t (Z_t^μ are the usual augmented σ-fields for P^μ) .

(i) If $Z_T = Z_{T-}$ on $\{0 < T < \infty\}$, P^μ-a.s. then T is Z_t^μ-previsible and $Z_T^\mu = Z_{T-}^\mu$.

(ii) Let B denote the set of Ray branch-points in $(H_A \cap H_0)^+$. Then the totally inaccessible part of T is

$$T_A = \begin{cases} T & \text{on } A \\ \infty & \text{on } \Omega_Z - A \end{cases} ,$$

where

$$A = \{0 < T < \infty, \bar{X}_{T-} \in (H_A \cap H_0)^+ - B, \bar{X}_{T-} \neq \bar{X}_T\}$$

$$= \{0 < T < \infty, Z_{T-} \in H_0, Z_{T-} \neq Z_T\}, \quad P\text{-a.s.}$$

PROOF. Both (i) and the first expression for A in (ii) are taken directly from [8]. It remains only to verify the second expression for A. Clearly if $z \in D_A$ and $h(z) \in H_0$, then $z = h(z)$ and $P^z\{\bar{X}_0 = z\} = 1$, hence $z \notin B$. Conversely, if $z \in D_A$ and $h(z) \in H - H_0$, then $P^{h(z)}\{Z_0 = h(z)\} = 0$. Hence $\bar{P}^z\{h(\bar{X}_0) = h(z)\} = 0$, and so $\bar{P}^z\{\bar{X}_0 = z\} = 0$. Then $z \in B$, completing the proof.

3. A VIEW TOWARD APPLICATIONS.

Since the object of the present work is not to study the prediction process per se but to develop it for applications to other processes, we conclude this essay with some general observations and partly heuristic discussion of the simplest types of examples. It may appear at present that by choosing different packets H_A one can obtain in the form Z_t practically any kind of r.c.l.l. strong-Markov process, but this is not quite true. A special feature of Z_t that is important in applications is the absence of "degenerate branch points." Here a degenerate branch point is one from which the left limit process jumps to a fixed point of the state space. But since we have a Borel transition function $q(t,z,A)$ for the moderate Markov property, and $q(0,h,\{z\}) = 1$ if and only if $z = h \in H_0$, such deterministic jumps do not occur. This is again an expression of the fact that, by Corollary 2.12, Z_t is practically just the Ray process of a right-process.

The same fact permits us to give criteria for Z_t to be a Hunt process, or for it to be a $\mathcal{Z}_t$-previsible process.

THEOREM 3.1. Let H_A be a complete Borel packet for Z_t (Definition 2.1, 3)). Then a) Z_t is a Hunt process on H_A relative to the usual σ-fields $\mathcal{Z}_t$ if and only if $H_A \subset H_0$ (i.e., H_A is an H_0-packet). This implies the quasi-left-continuity of the σ-fields $\mathcal{Z}_t^\mu$ for each initial distribution μ on H_A; b) Z_t is $\mathcal{Z}_t$-previsible if and only if it is continuous (P^μ a.s. for all μ).

PROOF. If $H_A \subset H_0$ then clearly Z_t is a right-process. The requirement that it be a Hunt process is then quasi-left-continuity. By decomposing any $\mathcal{Z}_t$-stopping time T into accessible and totally inaccessible parts for P^μ ([4, IV, 81]) one sees that for quasi-left-continuity it is necessary and sufficient that for any increasing sequence T_n of $\mathcal{Z}_{t+}^o$-stopping times with $\lim_{n\to\infty} T_n = T$ and $P^\mu\{T_n < T\} = 1$, one have

$P^{\mu}\{Z_T = Z_{T-}\} = P^{\mu}\{T < \infty\}$. But by the moderate Markov property,

$$(3.1) \qquad P^{\mu}(Z_T = Z_{T-} \ ; \quad T < \infty)$$

$$= E^{\mu}(P^{\mu}(Z_T = Z_{T-} | \mathcal{Z}_{T-}) \ ; \quad T < \infty)$$

$$= E^{\mu}(q(0, Z_{T-}, \{Z_{T-}\}) \ ; \quad T < \infty)$$

$$= P^{\mu}\{T < \infty\} \ ,$$

since $q(0,z,\{z\}) = 1$ on H_0 and H_A is complete. Finally, $H_A \subset H_0$ is necessary even for a right-process, so the converse is obvious. The last statement of a) is proved for Ray topology in [8, (13.2), (i) and (iv)]. Thus it is another way of ensuring that $\bar{X}$ is a Hunt process in the Ray topology, as remarked in [ibid, (13.3)]. However, by considering $H_A (\subset H_0)$ as a subset of $h(D_A) \cap H_0$ from Corollary 2.12, we see that for μ concentrated on H_A the process Z_t is quasi-left-continuous if and only if it is quasi-left-continuous in the Ray topology. Hence the result carries over.

Turning to b), continuity implies previsibility so we need only prove the converse. Then if Z_t is $\mathcal{Z}_t$-previsible, both Z_{t-} and Z_t are $\mathcal{Z}_t$-previsible processes, and to prove that Z_t is continuous we need only prove them indistinguishable. By the previsible section theorem, it is enough to show that for $\mathcal{Z}_t^{\mu}$-previsible T, $0 < T < \infty$, $P^{\mu}\{Z_{T-} = Z_T\} = 1$ (as usual, we may replace general previsible T by $T_N = N$ on $\{T = 0 \text{ or } T \geq N\}$, and let $N \to \infty$) . By the moderate Markov property we have $P^{\mu}\{Z_{T-} = Z_T\} = E^{\mu} q(0, Z_{T-}, \{Z_{T-}\})$, hence we must show that $P^{\mu}\{Z_{T-} \in H - H_0\} = 0$. However, since Z_t is previsible it is known [4, IV, 57] that Z_T is $\mathcal{Z}_{T-}^{\mu}$-measurable. Since

$$P^{\mu}(Z_T \in A | \mathcal{Z}_{T-}^{\mu}) = q(0, Z_{T-}, A) \ ,$$

and there are no degenerate branching points, we must have $Z_{T-} \in H_0$ as required.

To give a feeling for the applications, we will consider briefly three situations:

a) X_t is a Markov process,

b) Z_t is a Markov chain,

c) $(X_t, (w_{2n-1}(t)))$ is a Markov additive process.

It is to be noted that b) is a condition on Z_t, while a) and c) are conditions on X_t. Thus our examples illustrate the point that in the combined study of X_t and Z_t neither is necessarily the first to be considered. One may start either with a known process or a known prediction process. To be sure, one does not ordinarily make assumptions on both X_t and Z_t, since each determines the other uniquely.

To study the case of Markovian X_t, if we are not interested in any "hidden information" we can assume for convenience that $P\{w_{2n-1}(t) = 0,\ 1 \leq n,\ t \geq 0\} = 1$, and drop the coordinates w_{2n-1} from our notation. To relate the Markov properties of X_t and Z_t, since Z_t has the role of a conditional distribution relative to $\mathcal{F}^o_{t+1}$ we must assume that X_t is Markov relative to $\mathcal{F}^o_{t+}$. Alternatively, we could equivalently use X_{t-}, Z_{t-}, and $\mathcal{F}^o_{t-}$, but we cannot use ^o_t in general since Z_t might not be $\mathcal{F}^o_t$-measurable.* Let $k \in H$ denote the probability of X_t, and let H_k be a Borel prediction packet for X_t as, for example, in Theorem 2.1 and the discussion following its proof. As noted in Definition 2.1 2), $(\varphi(Z_t), Z_t)$ is P^k-equivalent to (X_t, Z^k_t), and it suffices to look at the former pair. It is not hard to see how the Markov property of X_t translates into an instantaneous property of Z_t. In the first place, in view of Theorem 1.9 and the Remark following, $\mathcal{Z}^o_t$ is P^k-equivalent to the σ-field χ_{t+} where $\chi_t = \sigma\{\varphi(Z_s),\ s \leq t\}$. Hence, the Markov property of X_t is equivalent to the conditional independence (for P^k) of $\mathcal{Z}^o_t$ and $\sigma\{\varphi(Z_{t+s}),\ 0 \leq s\}$ given $\varphi(Z_t)$. But Z_t is also defined as a conditional probability over the latter σ-field given $\mathcal{Z}^o_t$, namely

$$Z_t(S) = P^k((\theta^Z_t)^{-1}\{\varphi(Z_{(\cdot)}) \in S\} \mid \mathcal{Z}^o_t)\ , \quad S \in \mathcal{F}^o\ .$$

Since $\mathcal{F}^o$ is countably generated, it follows that Z_t is determined P^k-a.s. by $\varphi(Z_t)$ (the details of this transparent reasoning are given in [10, Theorem 2.2] and fortunately need not be repeated here). It follows that there is a P^k-null set N_t and a $\bar{\mathcal{B}}_\infty/\mathcal{H}$-measurable ψ_t such that $Z_t = \psi_t(\varphi(Z_t))$ for $w_Z \notin N_t$. Conversely, if such N_t and ψ_t exist, then plainly X_t was Markov at time t relative to $\mathcal{F}^o_{t+}$. The function ψ_t plays the role of transition function for X_t, by assigning to it the conditional future $\psi_t(X_t)$. If ψ_t may be chosen free of t, then by definition X_t is homogeneous in time.

---

*A variety of analytic conditions making X Markov relative to $\mathcal{F}^o_{t+}$ is given in H. J. Englebert [7]. If X_t is only Markov relative to $\mathcal{F}^o_t$, then it is still germ - Markov relative to $\mathcal{F}^o_{t+}$ in the sense of [10], and may be approached by the method developed there under suitable conditions. From the standpoint of $\bar{\Omega}$ (as in [9]) $\mathcal{F}_t$ coincides with $\mathcal{F}_{t-}$, and the distinction becomes meaningless.

Perhaps the most noteworthy fact here is that even if X_t is neither homogeneous nor strong-Markov, the process Z_t^k (i.e., a standard modification of $\psi_t(X_t)$) with transition function q has both properties. Thus any such irregularities of X_t are due to ψ_t and N_t, not to Z_t. This provides a ready method of investigating transition functions of Markov processes which, as mentioned already, is the subject of the third essay.

At present, one may gain further insight by comparing this method to another one: that of the "space-time" process. It is a familiar fact that any Markov process (P^x, X_t) becomes homogeneous in time if we replace X_t by (t, X_t), so that an initial value (s,x) means that one considers the process X_{s+t}, $t \geq 0$, conditional upon $X_s = x$, but with the added coordinate $s + t$ so that no value of the pair can recur. While this device is very useful in particular cases, such as in studying the heat generator $(\frac{\Delta}{2} - \frac{\partial}{\partial t})$, it has also been used occasionally in a general role (E. B. Dynkin, [6, 4.6]). Contrary to first impressions, the method of the prediction process apparently is quite unrelated to this as a method of "making a Markov process homogeneous". Not only are the respective topologies quite different (assuming the product topology for the space-time process), but more importantly the prediction process can repeat values, and hence may be simpler. For example, a particle confined to the unit circle $0 \leq \theta < 2\pi$ and moving with velocity $v(t) = t - [t]$ (a saw-tooth function) has prediction process with states corresponding to pairs (v, θ), $0 \leq v < 1$, while its space-time process has states (t, θ), $0 \leq t < \infty$. In general, if X_t happens to be a time-homogeneous Markov process then it is usually equivalent to its prediction process, while (t, X_t) may be somewhat artificial and intractable.

Taking up our second illustration, since Z_t is always a homogeneous Markov process it is natural to ask under what conditions it is a process of some special type. For instance, if Z_t is a pure jump process, i.e., a sum of finitely many jumps with exponentially distributed waiting times for the next jumps given the past, then $X_t = \varphi(Z_t)$ obviously has the same property. But unlike Z_t, X_t need not be a Markov process.

To indicate the possibilities for X_t, we again take $w_{2n-1}(t) \equiv 0$, $1 \leq n$, and suppose also that $w_{2n}(t) \equiv 0$ for $2 \leq n$, so that X_t may be regarded as the real-valued process $w_2(t)$. To construct a process X_t having a pure-jump prediction process (apart from the case of Markovian X_t) one can begin with any family $K_n(x_1, \ldots, x_n; t_1, \ldots, t_n; (dx_{n+1} \times d\lambda_{n+1}))$, $1 \leq n$ of probability kernels over $\bar{R} \times [\varepsilon, \infty)$, for fixed $\varepsilon > 0$, and

$x_k \in \bar{R}$, $t_k > 0$, $1 \le k \le n$. Letting (x_1, λ_1) have any initial distribution on $\bar{R} \times [\varepsilon, \infty)$, define $X_t = x_1$ for $0 \le t < e_1$ where e_1 is a random variable with $P\{e_1 > t\} = \exp(-\lambda_1 t)$, independent of x_1 given λ_1. Proceeding by induction, suppose that $x_1, \ldots, x_n$ and $e_1, \ldots, e_n$ have been determined, and that X_t has been defined for $0 \le t < \Sigma_{k=1}^{n} e_k$. Then we select a pair (x_{n+1}, λ_{n+1}) distributed according to the kernel K_n with $t_0 = 0$, $t_k = \Sigma_{j=1}^{k} e_j$, and $x_k = X_{t_{k-1}}$, $1 \le k \le n$. The inductive definition is completed by setting $X_t = x_{n+1}$ for $\Sigma_{k=1}^{n} e_k \le t < \Sigma_{k+1}^{n+1} e_k$, where e_{n+1} is a random variable conditionally independent of $\{x_1, \ldots, x_{n+1}, e_1, \ldots, e_n\}$ given λ_{n+1}, and $P\{e_{n+1} > t\} = \exp(-\lambda_{n+1} t)$. On the P-null set where $\Sigma_{n=1}^{\infty} e_n < \infty$ we define $X_t = 0$ for $\Sigma_{n=1}^{\infty} e_n \le t$. It is evident that such X_t has a pure-jump prediction process, and it is plausible that any pure-jump prediction process Z_t all of whose expected waiting times exceed ε with probability 1 is obtained in this way (if $\varphi(Z_t)$ is a.s. 0 except for the first coordinate).

In this construction, even if X_t can assume only a finite number of distinct values, Z_t may have an uncountable state space since it "predicts" the whole future sequence of X_t-values. On the other hand, it is easy to give sufficient conditions on the K_n which imply that Z_t is even a finite Markov chain (other than X_t being itself one). Thus if, for some fixed N and all $n \ge N$, $K_n = K_N$ depends only on $(x_{n-N+1}, x_{n-N+2}, \ldots, x_n)$ while X_t has a finite state space, and if moreover λ_{n+1} is a fixed function $\lambda_{n+1}(x_{n-N+1}, x_{n-N+2}, \ldots, x_n, x_{n+1})$ depending only on the x_k's shown, then it is clear that the finitely many possibilities for these x_k's imply that Z_t will be a finite Markov chain. In particular, if the λ_n's reduce to a single constant λ, then X_t is a "generalized Poisson process based on an N-dependent Markov chain," in the evident sense of dependence on the past only through the last N states visited. Obviously, then, the possibilities for X_t such that Z_t is a pure jump process are quite great, and we do not pursue them farther here.

For a type of example which involves a non-Markovian X_t, and in which the unobserved data $(w_{2n-1}(t))$ are of basic importance, we consider briefly the "Markov additive processes" (in the sense of E. Cinlar; see [3] and [15] for a vivid introduction and further references). Roughly speaking, a standard Markov additive process is a pair (X_t^1, X_t^2)

where X_t^1 is a standard process (in the sense of Blumenthal and Getoor) and X_t^2 is a real-valued process with conditionally independent increments given X_t^1. In the applications X_t^2 is observed, and one would like to make inferences about the underlying process X_t^1. For simplicity of notation we assume that $w_n(t) \equiv 0$ for $n > 2$, and that X_t^1 is real-valued, so that we may identify the trap state Δ as ∞, and let $X_t^1 = w_1(t)$, $X_t^2 = w_2(t)$ on Ω. Since X_t^1 is Markovian, and given X_t^1 the future increments of X_t^2 are independent of F_{t+}^o, it is to be expected that the prediction process Z_t^P of (X_t^1, X_t^2) is determined by the value of X_t^2 and the conditional distribution of X_t^1 given F_{t+}^o.

If one is concerned only with X_t^1, it is simpler to treat $X_t^2 - X_0^2$ as an additive functional, and consider Z_t^P restricted to sets of the form $S \cap \{X_0^2 = 0\}$, $S \in G^o$. Then the value X_t^2 becomes irrelevant in determining Z_t^P if the values of $Z_t^P\{X_0^1 \in B\}$, $B \in \bar{\mathcal{B}}$, are known. We can incorporate this change of view by redefining our translation operators appropriately. We turn now to the necessary notation and hypotheses.

DEFINITION 3.2. Let $\Omega^* = \{(w_1(t), w_2(t)) : w_2(0) = 0$ and $w_2(t) \neq \pm\infty$ for all $t\}$. Further, let $G_{t+}^* = \{S \cap \Omega^* : S \in G_{t+}^o\}$ and $F_{t+}^* = \{S \cap \Omega^* : S \in F_{t+}^o\}$. Finally, let $\theta_0^*((w_1,w_2)(s)) = (w_1(s), w_2(s) - w_2(0))$ and $\theta_t^*(w_1,w_2) = \theta_0^*\theta_t(w_1,w_2)$ on Ω^*.

HYPOTHESIS 3.3. A standard Markov additive process $(w_1(t), w_2(t))$ on Ω^* is a collection of probabilities P^x on G^* $(= \vee_t G_{t+}^*)$, $x \in \bar{R}$, such that $w_1(t)$ is a standard Markov process (we take $\Delta = +\infty$ as the terminal point), and

(i) $P^x\{(w_1(t), w_2(t)) \in B_2\}$ is $\bar{\mathcal{B}}$-measurable for $B_2 \in \bar{\mathcal{B}}_2$,

(ii) For G_{t+}^*-optional $T < \infty$, one has

$$P^x(\theta_T^*(w_1,w_2)(s) \in B_2 | G_{T+}^*)$$

$$= P^{w_1(T)}((w_1,w_2)(s) \in B_2)\ , \quad B_2 \in \bar{\mathcal{B}}_2\ .$$

We now introduce a notation for the process of conditional probabilities of $w_1(t)$ given F_{t+}^o, which is our main concern.

DEFINITION 3.3. The filtering process of $w_1(t)$ by $w_2(t)$ for initial distribution μ is the process $F_t^\mu(\cdot)$: $F_t^\mu(B) = Z_t^\mu\{w_1(0) \in B\}$, $B \in \bar{\mathcal{B}}$, where for each initial distribution μ on $\bar{R}$ we let Z_t^μ denote the prediction process for

$$P^\mu = \int P^x \mu(dx)\ , \quad \text{with } P^\mu(\Omega - \Omega^*) = 0\ .$$

We remark that, of course, we have $F_t^\mu(B) = P^\mu(w_1(t) \in B | F_{t+}^o)$. A remarkable result of M. Yor [15, Theorem 4] asserts that the $F_t^\mu(\cdot)$ are themselves r.c.l.l. strong-Markov processes with a single Borel transition function. Here we will deduce this from the corresponding fact for the Z_t^μ . However, this does not quite give as nice a topology as [15] (see the remarks following the proof). For our proof, we need a further notation and lemma.

LEMMA 3.4. For each initial μ on $\bar{R}$ and $y \in R$, we define a measure P_y^μ on (Ω, G^o) by first $P_y^x(S) = P^x\theta_0^*(S_y)$, where $S_y = S \cap \{w_2(0) = y\}$, $S \in G^o$, and then $P_y^\mu = \int P_y^x \mu(dx)$. Let $H^* = \{P_y^\mu, y \in R, \text{ all } \mu\}$. Then H^* is a Borel prediction packet, and for each μ we have

$$(3.2) \qquad P^\mu\{Z_t^\mu = P_{w_2(t)}^{F_t^\mu} \text{ for all } t \geq 0\} = 1 .$$

PROOF. For S of the form

$$S = \{a < w_2(0) < b, S_{w_2(0)} = S^*\} \text{ for } S^* \in G^* ,$$

we have $P_y^x(S) = I_{(a,b)}(y)\, P^x(S^*)$. Let S_n be a countable sequence of such sets which generates G^o . Since by (i) and (ii) P^x is a one-to-one Borel kernel of probabilities on G^* with $P^x\{w_1(0) = x\} = 1$, we see that P^μ is also one-to-one and Borel with respect to the measure μ . Then it follows that the sequential range $\{(P_y^\mu(S_n)), y \in R, \mu \text{ a probability on } \bar{R}\}$ is a Borel set in $X_{k=1}^\infty [0,1]$, implying that $H^* \in H$. To prove that H^* is a packet it suffices to show (3.2), since clearly $P^\mu = P_0^\mu$ and if (3.2) is true then

$$(3.3) \qquad P_y^\mu\{Z_t = P_{y+w_2(t)}^{F_t^\mu} \text{ for all } t \geq 0\} = 1 , \quad y \in R$$

by translation (we omit the superscript P_y^μ on Z_t) . Since P_y^x is Borel in (x,y), and F_t^μ is F_{t+}^μ -optional, it is clear that both sides of (3.2) are F_{t+}^μ-optional, hence it is enough to prove

$$P^\mu\{Z_T = P_{w_2(T)}^{F_T^\mu}\} = 1$$

for F_{t+}^μ-optional $T < \infty$. Now by (ii) and the definition of Z_t^μ, we have

$$
\begin{aligned}
Z_T^\mu(S) &= P^\mu(\theta_T^{-1}(S)\,|\,\mathcal{F}_{T+}^\mu) \\
&= E^\mu(P^\mu(\theta_T^{-1}(S)\,|\,\mathcal{G}_{T+}^\mu)\,|\,\mathcal{F}_{T+}^\mu) \\
&= E^\mu[P^\mu(\theta_0^*\theta_T(w_1,w_2) \in \theta_0^*(S_{w_2(T)})\,|\,\mathcal{G}_{T+}^\mu)\,|\,\mathcal{F}_{T+}^\mu] \\
&= E^\mu(P_{w_2(T)}^{w_1(T)}(S)\,|\,\mathcal{F}_{T+}^\mu) \\
&= P_{w_2(T)}^{F_T^\mu}(S) \quad ,
\end{aligned}
$$

as asserted.

By this lemma, we can introduce the filtering process as a function of the prediction process with state space H^*, and derive its properties from the latter.

THEOREM 3.4. The probability-valued process $F_t(B) = Z_t\{w_1(0) \in B\}$, $B \in \bar{\mathcal{B}}$, as a function of the prediction process Z_t on H^*, is a right-continuous, strong-Markov process for a suitable topology such that the space $(\bar{M},\bar{\mathcal{M}})$ of probabilities on $\bar{\mathcal{B}}$ with its generated σ-field is a metrizable Lusin space. Accordingly, the same results are true for the processes F_t^μ .

PROOF. For $h = P_y^\mu \in H^*$, set $F^h(B)=\mu(B), B\in\bar{\mathcal{B}}$ (this is not to be mistaken for F_t^μ, which has a subscript). Then for $M \in \bar{\mathcal{M}}$, we let $A_M = \{h \in H^* : F^h \in M\}$. Clearly $A_M \in \mathcal{H}$, and writing now P_y^μ for the probability of the canonical prediction process on H^* with $h = P_y^\mu$ as initial measure we have

$$
\begin{aligned}
(3.4) \qquad & P_y^\mu(F_{T+t} \in M\,|\,\mathcal{Z}_T^o) \\
&= P_y^\mu\,(Z_{T+y} \in A_M\,|\,\mathcal{Z}_T^o) \\
&= q(t,Z_T,A_M) \quad .
\end{aligned}
$$

On the other hand, recalling the σ-fields χ_t generated by $\varphi(Z_s)$, $s \le t$, where $\varphi(Z_s)$ is P_y^μ-equivalent to $w_2(s)$, we can transfer (3.3) to the canonical space and rewrite (3.4) in the form

$$(3.5)\qquad P_y^{\mu}(F_{T+t} \in M \mid \chi_{T+}) = q(t, P_{y+\varphi(Z_T)}^{F_T}, A_M)$$

$$= P_{y+\varphi(Z_T)}^{F_T}\{Z_t \in A_M\}$$

$$= P_0^{F_T}\{F_t \in M\}$$

$$= q(t, P_0^{F_T}, A_M) \ ,$$

where we used (3.3) with F_T in place of μ, along with the fact that in distribution F_t does not depend on y for initial probabilities of the form $P_y^{\mu} \in H^*$. Accordingly, we may define a transition function q^* for F by $q^*(t,\mu,M) = q(t,P_0^{\mu},A_M)$, and (3.5) becomes

$$(3.6)\qquad P_y^{\mu}(F_{T+t} \in M \mid Z_T^{o}) = q^*(t,F_T,M) \ .$$

Since P^x was assumed to be Borel in x and P_0^{μ} is one-to-one in μ, it is not hard to see that q^* is a Borel transition function on $(\bar{M},\bar{\mathcal{M}})$. Finally, the topology on $\bar{M}$ referred to in the theorem is just that induced by the mapping $\mu \to P_0^{\mu}$ and the topology of H^*, since it is easily seen that right-continuity of

$$Z_t = P_{y+w_2(t)}^{F_t^{\mu}}$$

in (3.3) implies right-continuity of

$$P_0^{F_t^{\mu}}$$

(from the right-continuity of $w_2(t)$). Thus Theorem 3.4 is proved.

DISCUSSION. It follows directly from the (known) fact that the optional projections of the r.c.l.l. processes $f(w_1(t)))$, $f \in C(\bar{R})$, are again r.c.l.l. P_y^{μ}-a.s. ([5, Chapter 2, Theorem 20]), that F_t^{μ} is even r.c.l.l. in the usual weak -* topology. This, together with further applications, is found in [15]. From an applied viewpoint, it is only the processes $F_{t-}^{\mu}(b) = Z_{t-}^{\mu}\{w_1(0) \in B\}$, $B \in \bar{\mathcal{B}}$, which are realistic, since only they do not depend on the future element of $\mathcal{F}_{t+}^{o}$. Further, with the usual convention that $\mathcal{F}_{0-}^{o}$ is degenerate, one has $P_y^{\mu}\{F_{0-}^{\mu} = \mu\} = 1$, unlike F_{0+}^{μ}. Using the fact that $P^{\mu}\{w_1(T-) = w_1(T)\} = 1$ at previsible $T < \xi$, however, it is clear that F_t^μ has no previsible discontinuities except perhaps at the lifetime ξ of $w_1(t)$. Hence, the moderate Markov property of F_{t-}^μ follows from the Markov property of F_t^μ .

A final remark seems merited concerning the Definition 2.1 of the prediction space Ω_Z . According to [4, IV, 19], Ω_Z is a coanalytic subset of the space of all r.c.l.l. paths with values in $\bar{H}$, and this space is a measurable Lusin space. The question naturally arises of whether, by restricting this space to the r.c.l.l. paths in some stronger topology, one might preserve its function of representing the processes Z_t^h and yet improve some other properties. A natural candidate is then the Skorokhod topology of measures on Ω . However, as shown by D. Aldous (unpublished) one does not have $P^h\{Z_t^h$ is r.c.l.l. in the Skorokhod topology$\} = 1$. The difficulty is that the Skorokhod left-limits do not exist unless X_t is P^h-quasi-left-continuous. Hence the topology of $\bar{H}$ seems to be the most reasonable alternative.

REFERENCES

1. Blumenthal, R. M. and Getoor, R. K. Markov Processes and Potential Theory. Academic Press, New York, 1968.

2. Chung, K. L. and Walsh, J. "To reverse a Markov process," Acta Math. 123, 1970, 225-251.

3. Cinlar, E. Markov additive processus and semi-regeneration. Proc. Fifth Conf. on Probability Theory (Bresov), Acad. R.S.R., Bucharest.

4. Dellacherie, C. and Meyer, P.-A. Probabilités et Potentiel, Chap. I a IV. Hermann, Paris. Chap. V - VII (to appear), 1975.

5. Dellacherie, C. Capacités et Processus Stochastiques. Springer-Verlag, Berlin, 1972.

6. Dynkin, E. B. Theory of Markov Processes. Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1961.

7. Engelbert, H. J. "Markov processes in general state spaces" (Part II), Math. Nachr. 82, 1978, 191-203.

8. Getoor, R. K. Markov Processes: Ray Processes and Right Processes. Lecture Notes in Math, No. 440. Springer-Verlag, New York, 1975.

9. Knight, F. B. "A predictive view of continuous time processes," The Annals of Probability, 3, 1975, 573-596.

10. Knight, F. B. "Prediction processes and an autonomous germ-Markov property," The Annals of Probability, 7, 1979, 385-405.

11. Kunita, H. and Watanabe, T. Some theorems concerning resolvents over locally compact spaces. Proceedings of the Fifth Berkeley Symposium on Math. Stat. and Prob. Vol. II, Part 2. University of Cal. Press, 1966, 131-163.

12. Meyer, P. A. La theorie de la prédiction de F. Knight. Seminaire de Prob. X, Université de Strasbourg, 1976, 86-104.

13. Meyer, P. A. and Yor, M. Sur la theorie de la prédiction, et le probleme de decomposition des tribus F^0_{t+} . Seminaire de Prob. X, Université de Strasbourg, 1976, 104-117.

14. Seminaire de Probabilités I-XII. Université de Strasbourg. Lecture Notes in Math # 39, 51, 88, 124, 191, 258, 321, 381, 465, 511, 581, 649.

15. Yor, M. Sur les théories du filtrage et de la prédiction. Seminaire de Prob. XI, Université de Strasbourg, 257-297.