NOTES

ON THE CHAPMAN-KOLMOGOROV EQUATION[1]

By Jack Karush

University of California, Berkeley

A partial answer is given to the question of whether every Markov random function comes from a system of transition probabilities satisfying the Chapman-Kolmogorov equation. A given Markov random function determines the transition probabilities up to sets of probability zero and for any choice of the transition probabilities the Chapman-Kolmogorov equation holds up to sets of probability zero. The problem then is one of selecting appropriate versions of the transition probabilities so that the Chapman-Kolmogorov equation holds everywhere. It is shown that such selections exist whenever the time parameter set is countable or whenever the joint distribution of any two of the random variables is absolutely continuous with respect to the product of the marginal distributions. Although the latter condition is always satisfied when the state space is countable, or more generally, when each random variable assumes a countable number of values with probability one, this case, being especially simple, is treated separately. The results are based on exploiting the device of using the marginal distribution when in doubt about what the conditional probability distribution should be.

Let $(X_t, t \varepsilon T)$ be a Markov random function, where T is a set of real numbers with elements denoted by r, s, t, u, v. Let $\mathcal{S}$ be the σ-field of linear Borel sets, and for every t define $P_t(S) = P[X_t \varepsilon S]$, $S \varepsilon \mathcal{S}$. For every $s, t, s < t$, consider the joint probability distribution of X_s, X_t. There exists what we shall call a version of the conditional probability distribution of X_t given X_s or, more concisely, a version of $P(X_t \mid X_s)$, that is, a function P_{st} of x, S, x real, $S \varepsilon \mathcal{S}$, such that $P_{st}(\cdot, S)$ is Borel for every $S \varepsilon \mathcal{S}$, $P_{st}(x, \cdot)$ is a probability distribution on $\mathcal{S}$ for every x, and

$$\int_S P_s(dx) P_{st}(x, S') = P[X_s \varepsilon S, X_t \varepsilon S'], \qquad S, S' \varepsilon \mathcal{S}.$$

The Markov property implies that for $r < s < t$, $P_{rs} * P_{st}$ is a version of $P(X_t \mid X_r)$, where by definition

$$(P_{rs} * P_{st})(x, S) = \int P_{rs}(x, dy) P_{st}(y, S), \qquad \text{all } x, S \varepsilon \mathcal{S},$$

Received September 26, 1960; revised March 27, 1961.

[1] This paper was prepared with the partial support of the Office of Ordnance Research, U. S. Army under Contract No. DA-04-200-ORD-171, Task Order 3.

so that the *Chapman-Kolmogorov* $(C - K)$ *equation*

$$P_{rt}(x, S) = (P_{rs}*P_{st})(x, S)$$

holds for $x \not\in N \in \mathcal{S}$, $S \in \mathcal{S}$, where $P_r(N) = 0$ and N depends on r, s, t, S.

On the other hand the usual approach ([1] pp. 89, 255–6) is to start out with $(P_{st}, s, t \in T, s < t)$ satisfying the $C - K$ equation identically, together with an arbitrary initial probability distribution P_{t_0}, T being assumed to have a minimum value t_0, and to construct the probability distribution of the corresponding random functions. A natural question is whether the probability distributions of all Markov random functions with T having a minimum value are obtained in this manner, or slightly more generally, whether, or under what conditions, one may select versions P_{st} of $P(X_t \mid X_s)$, $s < t$, satisfying the $C - K$ equation identically.

Each of the conditions 1–4 below ensures such a selection; 1 and 2 are special cases of 4 and 3, respectively, but are isolated because of their simplicity.

1. T **= integers.** In this case an obvious selection is available. For every n take any version of $P(X_{n+1} \mid X_n)$ and define for $m > 0$, all n,

$$P_{n,n+m}(x, S) = \int P_{n,n+1}(x, dy_1) \int P_{n+1,n+2}(y_1, dy_2) \cdots \int P_{n+m-2,n+m-1}(y_{m-2}, dy_{m-1}) P_{n+m-1,n+m}(y_{m-1}, S), \quad \text{all } x, S \in \mathcal{S}.$$

It is easily verified that $P_{n,n+m}$ is a version of $P(X_{n+m} \mid X_n)$ and the $C - K$ equation is satisfied identically. This amounts to verifying that the operation "$*$" is associative.

2. For every t, P_t is discrete, that is, there exists a countable set C_t such that $P_t(C_t) = 1$. For every s, if $P_s(\{x\}) > 0$ then necessarily

$$P_{st}(x, S) = \frac{P[X_s = x, X_t \in S]}{P[X_s = x]}, \qquad t > s, S \in \mathcal{S},$$

and if $P_s(\{x\}) = 0$ define

$$P_{st}(x, S) = P_t(S), \qquad t > s, S \in \mathcal{S}.$$

Since $P_s[x: P_s(\{x\}) = 0] = 0$, P_{st} is a version of $P(X_t \mid X_s)$. If $r < s < t$ and $P_r(\{x\}) > 0$ then $P_{rt}(x, S) = (P_{rs}*P_{st})(x, S)$, $S \in \mathcal{S}$. If $P_r(\{x\}) = 0$ then

$$P_{rt}(x, S) = P_t(S) = \int P_s(dy) P_{st}(y, S) = \int P_{rs}(x, dy) P_{st}(y, S) = (P_{rs}*P_{st})(x, S), \quad S \in \mathcal{S}.$$

3. For every s, t, $s < t$, there exists a version P_{st} such that $P_{st}(x, \cdot)$ is absolutely continuous with respect to $P_t(P_{st}(x, \cdot) \ll P_t)$ for all x, or equivalently, the joint probability distribution of X_s, $X_t \ll$ the product measure $P_s \times P_t$,

or equivalently, $\ll$ some product measure $\lambda \times \mu$, λ, μ σ-finite. We first establish the equivalences. Assume $P_{st}(x, \cdot) \ll P_t$ for all x and suppose $(P_s \times P_t)(B) = 0$, where B is a two-dimensional Borel set; then there exists $N \varepsilon \mathcal{S}$ such that $P_s(N) = 0$ and $P_t(B_x) = 0$ for $x \not\varepsilon N$, where $B_x = [y\colon (x, y) \varepsilon B]$, and we have

$$P[(X_s, X_t) \varepsilon B] = \int P_s(dx)P_{st}(x, B_x) = 0.$$

Conversely, if the joint probability distribution of X_s, $X_t \ll \lambda \times \mu$, λ, μ σ-finite, then $P_s \ll \lambda$ and $P_t \ll \mu$, so that there exist densities $dP_s/d\lambda$, $dP_t/d\mu$. Let $S = [x\colon (dP_s/d\lambda)(x) > 0]$, $S' = [x\colon (dP_t/d\mu)(x) > 0]$. Then $P_s(S) = 1$, $P_t(S') = 1$, $\lambda \ll P_s$ on S, and $\mu \ll P_t$ on S', so that $\lambda \times \mu \ll P_s \times P_t$ on $S \times S'$ and $P[(X_s, X_t) \not\varepsilon S \times S'] = 0$. It follows that the joint probability distribution of X_s, $X_t \ll P_s \times P_t$ and therefore has a density which can be taken to be of the form $p_s(x)p_{st}(x, y)$ where

$$\int p_{st}(x, y)P_t(dy) = 1 \qquad \text{for all } x.$$

Then

$$P_{st}(x, S) = \int_S p_{st}(x, y)P_t(dy), \qquad \text{all } x, S \varepsilon \mathcal{S},$$

defines a version of $P(X_t \mid X_s)$ and $P_{st}(x, \cdot) \ll P_t$ for all x.

Let U be the union of a countable dense subset of T and the countable set of points of T which are not two-sided limit points of T. For every t let

$$N_t = \bigcup_{\substack{t<u<v\\ u,v\varepsilon U\\ y \text{ rational}}} [x\colon P_{tv}(x, S_y) \neq (P_{tu}*P_{uv})(x, S_y)],$$

where $S_y = (-\infty, y)$, and define, for $t < u \varepsilon U$,

$$\begin{aligned}\hat{P}_{tu}(x, \cdot) &= P_{tu}(x, \cdot) && \text{if } x \not\varepsilon N_t\\ &= P_u && \text{if } x \varepsilon N_t.\end{aligned}$$

Since $P_t(N_t) = 0$, $\hat{P}_{tu}$ is a version of $P(X_u \mid X_t)$ and since a probability distribution on $\mathcal{S}$ is determined by its values on S_y, y rational, we have, for $x \not\varepsilon N_t$,

$$P_{tv}(x, \cdot) = (P_{tu}*P_{uv})(x, \cdot), \qquad t < u < v, u, v \varepsilon U,$$

and hence

$$\hat{P}_{tv}(x, \cdot) = (\hat{P}_{tu}*P_{uv})(x, \cdot), \qquad t < u < v, u, v \varepsilon U.$$

For $x \varepsilon N_t$

$$\hat{P}_{tv}(x, \cdot) = (\hat{P}_{tu}*P_{uv})(x, \cdot) = P_v, \quad t < u < v, u, v \varepsilon U,$$

by the same reasoning used in 2. Therefore

$$\hat{P}_{tv} = \hat{P}_{tu}*P_{uv}, \qquad t < u < v, u, v \varepsilon U.$$

Now $P_{tu}(x, \cdot) \ll P_u$ for every x; consequently $\hat{P}_{tu}(x, \cdot) \ll P_u$ for every x. It follows that $\hat{P}_{tu} * P_{us}$ is independent of the version P_{us} of $P(X_s \mid X_u)$ for any $s > u$. Let P'_{us} be another version; then for every $S \varepsilon \mathcal{S}$,

$$P_u[y\colon P_{us}(y, S) \neq P'_{us}(y, S)] = 0$$

so that for every x, $\hat{P}_{tu}(x, [y\colon P_{us}(y, S) \neq P'_{us}(y, S)]) = 0$ and hence

$$\int \hat{P}_{tu}(x, dy) P_{us}(y, S) = \int \hat{P}_{tu}(x, dy) P'_{us}(y, S)$$

or $\hat{P}_{tu} * P_{us} = \hat{P}_{tu} * P'_{us}$. In particular

$$\hat{P}_{tv} = \hat{P}_{tu} * P_{uv} = \hat{P}_{tu} * \hat{P}_{uv}, \qquad t < u < v, u, v \varepsilon U.$$

If $s < t \not\varepsilon U$ there exists a $u \varepsilon U$ such that $s < u < t$ and we define $\hat{P}_{st} = \hat{P}_{su} * P_{ut}$. Then $\hat{P}_{st}$ is a version of $P(X_t \mid X_s)$, is independent of the version of $P(X_t \mid X_u)$ selected, and is well defined, for if $s < u < v < t, u, v \varepsilon U$, we have

$$\hat{P}_{sv} * P_{vt} = (\hat{P}_{su} * \hat{P}_{uv}) * P_{vt} = \hat{P}_{su} * (\hat{P}_{uv} * P_{vt}) = \hat{P}_{su} * P_{ut}$$

since $\hat{P}_{uv} * P_{vt}$ is a version of $P(X_t \mid X_u)$. Finally, the $\hat{P}_{st}$'s satisfy the $C - K$ equation identically. Suppose $r < s < t$. If $s \varepsilon U$ then $\hat{P}_{rt} = \hat{P}_{rs} * \hat{P}_{st}$ by definition. If $s \not\varepsilon U$ there exists $u \varepsilon U$ such that $r < u < s < t$ and, since $\hat{P}_{us} * \hat{P}_{st}$ is a version of $P(X_t \mid X_u)$,

$$\hat{P}_{rt} = \hat{P}_{ru} * (\hat{P}_{us} * \hat{P}_{st}) = (\hat{P}_{ru} * \hat{P}_{us}) * \hat{P}_{st} = \hat{P}_{rs} * \hat{P}_{st}.$$

4. T is countable. Here we impose no condition on the X_t's, but, guided by 3, we enlarge the exceptional set N_s to obtain absolute continuity to the extent needed. For every s define N_s as above with $U = T$ and set

$$M_s = N_s \cup \bigcup_{t>s} [x\colon P_{st}(x, N_t) > 0].$$

Then $P_s(M_s) = 0$ since $P_s(N_s) = 0$ and for $t > s$

$$0 = P_t(N_t) = \int P_s(dx) P_{st}(x, N_t)$$

which implies $P_s[x\colon P_{st}(x, N_t) > 0] = 0$. Suppose $r < s$ and $x \not\varepsilon M_r$. Then $P_{rs}(x, M_s) = 0$; for $P_{rs}(x, N_s) = 0$ and if $t > s$,

$$0 = P_{rt}(x, N_t) = \int P_{rs}(x, dy) P_{st}(y, N_t)$$

which implies $P_{rs}(x, [y\colon P_{st}(y, N_t) > 0]) = 0$. For $s < t$ define

$$\begin{aligned} \hat{P}_{st}(x, \cdot) &= P_{st}(x, \cdot) && \text{if } x \not\varepsilon M_s \\ &= P_t && \text{if } x \varepsilon M_s. \end{aligned}$$

Since $P_s(M_s) = 0$, $\hat{P}_{st}$ is a version of $P(X_t \mid X_s)$. Suppose $r < s < t$. Then arguing separately for $x \not\varepsilon M_r$ and $x \varepsilon M_r$, we obtain for all x, $\hat{P}_{rt}(x, S_y) = (\hat{P}_{rs} * P_{st})(x, S_y)$, y rational, so that $\hat{P}_{rt}(x, \cdot) = (\hat{P}_{rs} * P_{st})(x, \cdot)$. Since

$P_{st}(y, \cdot) = \hat{P}_{st}(y, \cdot)$ if $y \varepsilon M_s$ and $\hat{P}_{rs}(x, M_s) = 0$ for all x it follows that $\hat{P}_{rt} = \hat{P}_{rs} * \hat{P}_{st}$.

REFERENCE

[1] J. L. DOOB, *Stochastic Processes*, John Wiley and Sons, New York, 1953.

A GENERALIZATION OF A THEOREM OF BALAKRISHNAN[1]

BY N. DONALD YLVISAKER[2]

New York University

1. Introduction. Given a stochastic process $\{X(t), t \varepsilon T\}$ on some probability space with second moment kernel

$$\mathcal{E}[X(s)\overline{X(t)}] = K(s, t),$$

a characterization is given of the function

$$m(t) = \mathcal{E}X(t).$$

This characterization includes the result of Balakrishnan [2] for the case of second order stationary, discrete or continuous parameter processes.

2. The characterization. Let T be an abstract set and let K be a positive definite kernel on $T \times T$. A function m on T is said to be an admissible mean value function for the kernel K if there exists a stochastic process $\{X(t), t \varepsilon T\}$ on some probability space with

$$\mathcal{E}[X(s)\overline{X(t)}] = K(s, t) \qquad \text{and} \qquad \mathcal{E}X(t) = m(t).$$

LEMMA 1. *m is an admissible mean value function for the kernel K if and only if $K(s, t) - m(s)\overline{m(t)}$ is positive definite.*

PROOF. if $K(s, t) - m(s)\overline{m(t)}$ is a positive definite kernel on $T \times T$, let $\{X(t), t \varepsilon T\}$ be a Gaussian process with mean function m and covariance kernel $K(s, t) - m(s)\overline{m(t)}$, ([3], p. 72). Then

$$\begin{aligned}\mathcal{E}[X(s)\overline{X(t)}] &= \mathcal{E}[X(s) - m(s)][\overline{X(t) - m(t)}] + m(s)\overline{m(t)} \\ &= K(s, t).\end{aligned}$$

Conversely, if m is admissible,

$$\mathcal{E}[X(s) - m(s)][\overline{X(t) - m(t)}] = K(s, t) - m(s)\overline{m(t)}$$

is positive definite.

Received January 4, 1961; revised May 29, 1961.

[1] This research was sponsored by the Office of Naval Research under Contract Number Nonr 266(33), Project Number NR 042-034, while the author was at Columbia University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] Now at the University of Washington.