A THEOREM OF LÉVY AND A PECULIAR SEMIGROUP[1]

By David A. Freedman

University of California, Berkeley

1. Introduction. Let I be a finite or countably infinite set. For each $t \geqq 0$ let $P(t)$ be a stochastic matrix on I, such that $P(t+s) = P(t)P(s)$, $P(0)$ is the identity matrix, and $P(t) \to P(0)$ coordinatewise as $t \to 0$. Then P is called a standard stochastic semigroup on I.

The result of Lévy (1958) referred to in the title is:

(1) Theorem. *For each pair i, j with $i \neq j$, there are only two possibilities: either $P(t, i, j) = 0$ for all $t \geqq 0$, or $P(t, i, j) > 0$ for all $t > 0$.*

One object of this note is to sketch an alternative proof of this fact. For historical discussion and some of the known proofs, see (Chung, 1960).

As is well known, P has a coordinatewise derivative at 0, called the infinitesimal generator Q. Another object of this note is to sketch the construction which proves

(2) Theorem. *There is a standard stochastic semigroup P on $I = \{1, 2, \cdots\}$ whose infinitesimal generator Q is given by:*

$$Q(i, i) = -\infty \quad \text{for all} \quad i \quad \text{in} \quad I \tag{3}$$

$$Q(i, j) = 0 \quad \text{for all} \quad i \neq j \quad \text{in} \quad I. \tag{4}$$

These results are discussed together because they involve the same technique, restricting a Markov chain to a subset of its state space. For simplicity, suppose all states are recurrent.

2. Restricting a Markov chain. Give I the discrete topology, and let $\bar{I} = I$ when I is finite, $\bar{I}$ = one point compactification of I when I is infinite. Let $\{X(t):0 \leqq t < \infty\}$ be an $\bar{I}$-valued stochastic process on a probability triple $(\Omega, \mathfrak{F}, \mu)$, which is a Markov chain with stationary standard transitions P. For technical safety, suppose the sample functions of X are quasiregular (the definition is in Section 5). Let J be a finite subset of I, and let X_J be the *restriction* of X to J, that is, X watched only when in J. More formally, let $\tau_J(t)$ be the greatest s such that the Lebesgue measure of $\{u:0 \leqq u \leqq s, X(u) \in J\}$ is t. Then $X_J(t) = X[\tau_J(t)]$. From the strong Markov property, X_J is a Markov chain with stationary transitions, call them P_J. Plainly, P_J is a standard stochastic semigroup on J. Call its infinitesimal generator Q_J, and say that P_J (respectively, Q_J) is P (respectively, Q) restricted to J. Plainly, for $K \subset J$, $(P_J)_K = P_K$ and $(Q_J)_K = Q_K$. It is not hard to check that

$$Q \leqq Q_J. \tag{5}$$

Received 14 July 1966.

[1] Prepared with the partial support of the National Science Foundation, Grant GP-5059; and a Sloan Foundation Grant.

Indeed, let $i \neq j$ be in J, and suppose $X(0) = i$ a.e. Then $X(t) = j$ implies X_J hits j before time t; up to an event of probability $o(t)$ as $t \to 0$, this implies $X_J(t) = j$. It is tempting to conjecture that $\lim_{J \uparrow I} Q_J = Q$.

It is possible to show that $X_J(t) \to X(t)$ in probability (this result will not be used). It would be very nice if $\{P_J\}$ were equicontinuous or equidifferentiable, but I have no results on this point.

For convenience, let $I = \{1, 2, \cdots\}$, $I_n = \{1, \cdots, n\}$. Write X_n, P_n, Q_n for X_{I_n}, P_{I_n}, Q_{I_n} respectively. Let $\Gamma_n(i, j) = Q_n(i, j) / -Q_n(i, i)$ for $i \neq j$ in I_n and $Q_n(i, i) \neq 0$; let $\Gamma_n(i, j) = 0$ elsewhere. Let $\pi_{n+1}(i) = 1 - \Gamma_{n+1}(i, n+1) \cdot \Gamma_{n+1}(n+1, i)$. It is not hard to check that for $i \in I_n$,

$$Q_n(i, i) = \pi_{n+1}(i) Q_{n+1}(i, i) \tag{6}$$

and for $i \neq j$ in I_n,

$$\pi_{n+1}(i)\Gamma_n(i, j) = \Gamma_{n+1}(i, j) + \Gamma_{n+1}(i, n+1)\Gamma_{n+1}(n+1, j). \tag{7}$$

The balance of this section is concerned with the conditional distribution of X_{n+1} given X_n. Of course, both have sample functions which are right continuous step functions.

The X_{n+1}-sample function is obtained by cutting the X_n-sample function, and inserting those time intervals on which the X_{n+1}-sample function takes the value $n + 1$. See the figure. Given X_n and the locations of the cuts, the lengths of these inserted intervals are independent and exponentially distributed, with common parameter $-Q_{n+1}(n + 1, n + 1)$. It remains to specify the conditional distribution of the cuts, given X_n. There are two kinds of cuts: the first kind appears at a jump of X_n, and the second kind appears interior to an interval of constancy for X_n. At a jump from i to j, the probability of a cut appearing is the ratio of $\Gamma_{n+1}(i, n+1)\Gamma_{n+1}(n+1, j)$ to $\pi_{n+1}(i)\,\Gamma_n(i, j)$. Cuts of the first kind appear independently from jump to jump, and independently of the location of cuts of the second kind. Locations of cuts of the second kind within each interval of constancy for X_n are independent from interval to interval. Within a particular k-interval, the location of cuts has a Poisson distribution, with parameter $-Q_{n+1}(k, k)[1 - \pi_{n+1}(k)]$.

Only the last claim will be argued. It is equivalent to this proposition about a Poisson process $\mathcal{P}$ of points on $[0, \infty)$, with parameter 1. Let $0 < \pi \leqq 1$. Given $\mathcal{P}$, toss a π-coin independently at each point of $\mathcal{P}$, from left to right, until a head is first obtained. Let T be the abscissa of the first head.

(8) PROPOSITION. *T is exponentially distributed with parameter π. Given T, the restriction of $\mathcal{P}$ to $[0, T)$ is distributed like a Poisson process with parameter $1 - \pi$ on $[0, T)$.*

This proposition is equivalent to (9). To state (9), let $\mathcal{R}$ be a Poisson process of points on $[0, \infty)$, with parameter $1 - \pi$. Let S be independent of $\mathcal{R}$, and exponentially distributed with parameter π.

(9) PROPOSITION. *The restriction of $\mathcal{R}$ to $[0, S)$ is distributed like the restriction of $\mathcal{P}$ to $[0, T)$.*

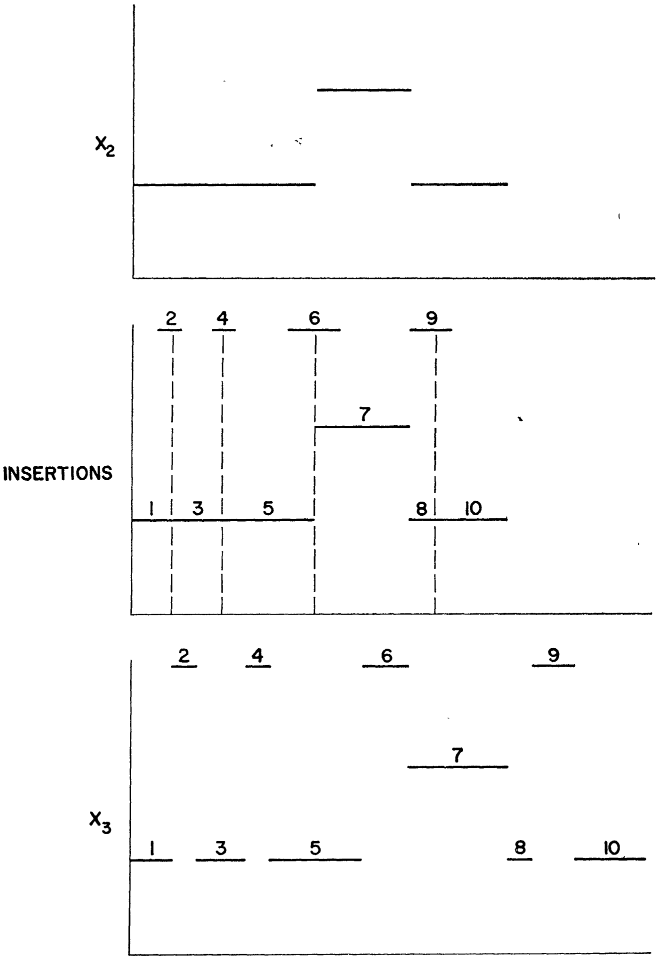

FIG. Intervals with the same number have the same length.

PROOF. Let τ be the waiting time from 0 to the first point of $\mathcal{R}$, an exponential random variable with parameter $1 - \pi$. Now, $\tau > S$ has probability π, and this is the probability that no points of $\mathcal{R}$ lie in $[0, S)$. Given $\tau < S$, τ is exponential with parameter 1, so if a point of $\mathcal{R}$ lies in $[0, S)$, the waiting time for the first such point is exponential with parameter 1. Given $\tau < S$ and given τ, the restriction of $\mathcal{R}$ to $[\tau, \infty)$, translated to the left by τ, is again Poisson with parameter $1 - \pi$, independent of $S - \tau$, which is again exponential with parameter π. This proves (9).

3. Proof of (1). Since $P(t, k, k) \geqq P(t/n, k, k)^n$, $P(t, k, k) > 0$ for all t. Fix $i \neq j$. Suppose $P(t, i, j) > 0$. Then

$$P(t + s, i, j) \geqq P(t, i, j)P(s, j, j) > 0.$$

Consequently, $\{t:P(t, i, j) > 0\}$ is either empty or an open half-line. Suppose by way of contradiction that $P(t, i, j) > 0$ for $t > 1$, and $P(t, i, j) = 0$ for $0 \leqq t \leqq 1$. Let X be a Markov chain on the probability triple $(\Omega, \mathfrak{F}, \mu)$, with stationary transitions P, starting from i, having quasiregular sample functions. Let σ be the least t if any with $X(t) = j$, $\sigma = \infty$ if none. Plainly, $\mu[\sigma \leqq 1] = 0$, $\mu[\sigma < 2] > 0$. A contradiction will be obtained by proving $\mu[\sigma < \frac{2}{3}] > 0$. Namely, let $\alpha_n =$ Lebesgue $\{t:0 \leqq t \leqq \sigma, X(t) > n\}$. Plainly, on $\sigma < \infty$, $\alpha_n \downarrow 0$ as $n \uparrow \infty$. Fix n so large that $\mu[\alpha_n < \frac{1}{3}, \sigma < 2] > 0$. Consider X_n. Let β_n be the least t if any with $X_n(t) = j$, $\beta_n = \infty$ if none. Plainly, $\beta_n \leqq \sigma$, so $\mu[\alpha_n < \frac{1}{3}, \beta_n < 2] > 0$. Let Z_1, Z_2, $\cdots$ be the states X_n passes through, with holding times U_1, U_2, $\cdots$ respectively. There must be a positive integer m, and states $i_2, \cdots, i_m \varepsilon I_n$, such that $\mu[\alpha_n < \frac{1}{3}, Z_1 = i, Z_2 = i_2, \cdots, Z_m = i_m, Z_{m+1} = j, U_1 + \cdots + U_m < 2] > 0$. Let μ_A be μ conditioned on the event $A = [Z_1 = i, Z_2 = i_2, \cdots, Z_m = i_m, Z_{m+1} = j]$. The finite random variables $U_1, \cdots, U_m$ are μ_A-independent and exponentially distributed. Let G be the set of m-vectors $u_1, \cdots, u_m$ with positive coordinates having sum less than 2, such that, given $U_1 = u_1, \cdots, U_m = u_m$, the event $[\alpha_n < \frac{1}{3}]$ has positive conditional μ_A-probability. Plainly, G has positive Lebesgue measure. Thus, $\frac{1}{6}G$ has positive Lebesgue measure, and $\mu_A[(U_1, \cdots, U_m) \varepsilon \frac{1}{6}G] > 0$. If $(v_1, \cdots, v_m) = \frac{1}{6}(u_1, \cdots, u_m)$, where $(u_1, \cdots, u_m) \varepsilon G$, the conditional μ_A-distribution of α_n given $U_1 = v_1, \cdots, U_m = v_m$ is stochastically smaller (the definition is below) than the conditional μ_A-distribution of α_n given $U_1 = u_1, \cdots, U_m = u_m$. This claim follows from Section 2, by an argument sketched below. Thus, given $U_1 = v_1, \cdots, U_m = v_m$, the conditional μ_A-probability that $\alpha_n < \frac{1}{3}$ is positive. But $v_1 + \cdots + v_m < \frac{1}{3}$, so, $\mu_A[\alpha_n < \frac{1}{3}$ and $U_1 + \cdots + U_m < \frac{1}{3}] > 0$. But on A, $\sigma \leqq \alpha_n + U_1 + \cdots + U_m$, and $\mu(A) > 0$, so $\mu[\sigma < \frac{2}{3}] > 0$. This completes the proof of Lévy's Theorem (1), provided the claim is argued.

DEFINITION. Let F and H be distribution functions on the real line. Then F is *stochistically smaller than* H iff $F(x) \geqq H(x)$ for all x.

Suppose it is possible to construct random variables ξ and ζ on a common probability triple, ξ having distribution F, ζ having distribution H, and $\xi \leqq \zeta$ everywhere. Then F is stochastically smaller than H.

LEMMA. *Let F_i be stochastically smaller than H_i. Then the convolution of F_1 and F_2 is stochastically smaller than the convolution of H_1 and H_2.*

ARGUMENT FOR CLAIM. For simplicity, suppose $m = 2$. One immediate problem is to specify the version of the conditional distribution for which the claim is true. To specify it, consider the regular conditional distribution obtained in Section 2 for $X_{n+1}, X_{n+2}, \cdots$ given the X_n-sample function, with $Z_1 = i$, $Z_2 = i_2$, $Z_3 = j$, $U_1 = u_1$, $U_2 = u_2$. Let $\hat{\alpha}_n$ be the sum of the lengths of the intervals inserted at or before the second discontinuity in X_n. Plainly, the distribution of $\hat{\alpha}_n$ coincides with the conditional μ_A-distribution of α_n given $U_1 = u_1$, $U_2 = u_2$. Now $\hat{\alpha}_n$ is the sum of four random variables: γ_i, the sum of the lengths of the intervals inserted interior to the ith interval of constancy for X_n, with $i = 1$ or 2; and δ_i, the sum of the lengths of the intervals inserted at the ith discontinuity of X_n, with $i = 1$ or 2. These four variables are independent, and the distribution of

δ_i does not depend on u_1, u_2. In view of the lemma, it is enough to show that if u_1 and u_2 are decreased the distribution of γ_i becomes stochastically smaller. Only γ_1 will be argued.

Construct inductively a sequence $\{Y_n(u):0 \leqq u < \infty\}$, $\{Y_{n+1}(u):0 \leqq u < \infty\}$, $\cdots$ of I_n, I_{n+1}, $\cdots$-valued stochastic processes, whose sample functions are right continuous step functions. Namely, $Y_n(u) = i$ for all u. Suppose Y_{n+m} is defined. The Y_{n+m+1}-sample function is obtained by cutting the Y_{n+m}-sample function, and inserting those time intervals on which the Y_{n+m+1}-sample function takes the value $n + m + 1$. Given Y_{n+m} and the locations of the cuts, the lengths of these inserted intervals are independent and exponentially distributed, with common parameter $-Q_{n+m+1}(n + m + 1, n + m + 1)$. It remains to specify the conditional distribution of the cuts, given Y_{n+m}. There are two kinds of cuts: the first kind appears at a jump of Y_{n+m}, and the second kind appears interior to an interval of constancy for Y_{n+m}. At a jump from i to j, the probability of a cut appearing is the ratio of

$$\Gamma_{n+m+1}(i, n + m + 1)\Gamma_{n+m+1}(n + m + 1, j) \quad \text{to} \quad \pi_{n+m+1}(i)\,\Gamma_{n+m}(i, j).$$

Cuts of the first kind appear independently from jump to jump, and independently of the location of cuts of the second kind. Locations of cuts of the second kind within each interval of constancy for Y_{n+m} are independent from interval to interval. Within a particular k-interval, the location of cuts has a Poisson distribution, with parameter $-Q_{n+m+1}(k, k)[1 - \pi_{n+m+1}(k)]$.

Let $\tau_{n+m}(u)$ be the least t such that Lebesgue

$$\{s:0 \leqq s \leqq t,\ Y_{n+m}(s) = i\} = u.$$

The restriction of Y_{n+m} to the time interval $[0, \tau_{n+m}(u))$ is distributed like X_{n+m} up to the first discontinuity of X_n, given $U_1 = u$. Let $\zeta(u) = \lim_{m\to\infty}$ Lebesgue $\{s:0 \leqq s \leqq \tau_{n+m}(u), Y_{n+m}(s) \geqq n + 1\}$. On the one hand, $\zeta(u)$ is nondecreasing with u. On the other, the distribution of $\zeta(u)$ coincides with the distribution of γ_1 given $U_1 = u$. Incidentally, ζ has stationary and independent increments.

4. Outline of construction for (2). The method is inverse to that of Section 2. Namely, choose a sequence Q_2, Q_3, $\cdots$ of matrices which are the infinitesimal generators of standard stochastic semigroups P_2, P_3, $\cdots$ on I_2, I_3, $\cdots$, where $I_n = \{1, \cdots, n\}$. Make the choice so that:

(10) Q_n is the restriction of Q_{n+1} to I_n;

(11) For any pair $i \neq j$ in I, $Q_n(i, j)$ is 0 for all large n;

(12) For any i in I, $Q_n(i, i) \to -\infty$;

(13) $Q_{n+1}(n + 1, n + 1) \to -\infty$ very quickly.

In view of Section 2, these conditions are compatible, as the next two paragraphs show.

For any n, Q_n, i in I_n, there exists a Q_{n+1} satisfying (10) with $Q_{n+1}(i, j) = 0$ for all $j \neq i$ in I_n and $Q_{n+1}(n + 1, n + 1)$ arbitrarily negative. To see this, suppose $Q_n(i, i) \neq 0$. For $k \neq i$ in I_n, let $\Gamma_{n+1}(k, n + 1) = 0$, and $\Gamma_{n+1}(k, j) = \Gamma_n(k, j)$ for $j \varepsilon I_n$. Let $\Gamma_{n+1}(i, n + 1) = 1$, $\Gamma_{n+1}(i, j) = 0$ for $j \varepsilon I_n$, and $\Gamma_{n+1}(n + 1, k) = \Gamma_n(i, k)$ for $k \varepsilon I_n$. Let $Q_{n+1}(j, j) = Q_n(j, j)$ for $j \varepsilon I_n$ and $Q_{n+1}(j, k) = -Q_{n+1}(j, j)\Gamma_{n+1}(j, k)$ for $j \neq k$ in I_{n+1}.

For any n, Q_n, i in I_n, there exists a Q_{n+1} satisfying (10) with $Q_{n+1}(i, i)$ and $Q_{n+1}(n + 1, n + 1)$ arbitrarily negative.

In view of (13), P_n will converge coordinatewise to a limiting standard stochastic semigroup P on I, with infinitesimal generator Q, and Q_n will be the restricton of Q to I_n. Then (3) and (4) follow from (5).

To study convergence, it may be helpful to construct, on a common probability triple, a sequence $X_2, X_3, \cdots$ of stochastic processes, whose sample functions are right continuous step functions, such that X_n is a Markov chain with stationary standard transitions P_n, and X_n is the restriction of X_{n+1} to I_n. In view of Section 2, (13) can be interpreted so stringently that the sum of lengths of the time intervals inserted to the left of each point on the X_2-time scale is finite. Then, there is a unique $\bar{I}$-valued process X, with quasiregular sample functions, such that X_n is X restricted to I_n. If $Q_{n+1}(n + 1, n + 1)$ is very negative, $X_{n+1}(t) = X_n(t)$ with high probability, uniformly on compact t-sets, by Section 2. By interpreting (13) even more stringently, X_n must converge a.e. The limit is necessarily X. This completes the proof of (2).

In general, I do not know which sequence Q_n have the property that Q_n is the restriction of some Q to I_n. I also do not know when a matrix is the infinitesimal generator of a standard stochastic semigroup.

5. Definition of quasiregularity. Let $i_\alpha : \alpha \varepsilon A$ be a net with values in $\bar{I}$. If $j \varepsilon \bar{I}$ but $j \not\varepsilon I$, q-lim $i_\alpha = j$ means lim $i_\alpha = j$ in the usual sense. If $j \varepsilon I$, q-lim $i_\alpha = j$ means: for any finite subset D of $I - \{j\}$, there is an $\alpha(D) \varepsilon A$ such that for all $\alpha > \alpha(D)$, $i_\alpha \not\varepsilon D$; and for any $\alpha \varepsilon A$ there is an $\alpha' > \alpha$ such that $i_{\alpha'} = j$. Let f be a function from $[0, \infty)$ to $\bar{I}$. Say f is *quasiregular* iff for all $t \geqq 0$, $f(t) = q$-lim $f(r)$ as binary rational r decreases to t, and for all $t > 0$, q-lim $f(r)$ exists as binary rational r increases to t. If P is a standard stochastic semigroup on I, there is a Markov chain with stationary transitions P starting from any $i \varepsilon I$, all of whose sample functions are of quasiregular (Chung (1960), II.7).

REFERENCES

CHUNG, KAI LAI. (1960). *Markov Chains with Stationary Transition Probabilities*. Springer, Berlin.

LÉVY, P. (1958). Processus markoviens et stationnaires. Cas dénombrable. *Ann. Inst. H. Poincaré*. **16** 7–25.