Statistics, Probability and Game Theory
IMS Lecture Notes – Monograph Series (1996) Volume 30

DE FINETTI'S THEOREM IN CONTINUOUS TIME

D. A. FREEDMAN
University of California, Berkeley

Abstract.

This paper gives a simpler proof of theorems characterizing mixtures of processes with stationary, independent increments, or mixtures of continuous-time Markov chains.

1. Introduction. This paper gives a simpler proof of two theorems. The first, due to Bühlmann (1963), characterizes mixtures of processes with stationary, independent increments. The second theorem, due to Freedman (1963), characterizes mixtures of Markov chains in continuous time with recurrent, stable states; stationarity conditions are eliminated, as are conditions on the sample paths.

To state the first result, let I be the real line with the Borel σ-field, Ω the set of functions from $[0,\infty)$ to I, and $\{X_t\}$ the coordinate process on Ω. Let $\mathcal{F}$ be the product σ-field in Ω and let $\mathcal{P}$ be a probability on $\mathcal{F}$. Let $P \in \Pi$ iff P is the law of a process with stationary, independent increments, which starts from 0, and which is continuous in probability. In particular, P is a probability on $\mathcal{F}$. As is easily verified, Π is a standard Borel space.

THEOREM 1. $\mathcal{P} = \int_\Pi P\mu(dP)$ for some probability μ on Π iff:
(i) $\mathcal{P}(X_0 = 0) = 1$, and
(ii) $\{X_t\}$ is continuous in $\mathcal{P}$-probability, and
(iii) for each $h > 0$, the $\mathcal{P}$-law of $\{X_{nh} - X_{(n-1)h} : n = 1, 2, \ldots\}$ is exchangeable.
The mixing measure μ is unique.

For the second result, let I be a countable set, which will be the state space. Let Ω be the set of functions from $[0,\infty)$ to I. As before, $\{X_t\}$ is the coordinate process on Ω, $\mathcal{F}$ is the product σ-field in Ω, and $\mathcal{P}$ is a probability on $\mathcal{F}$. Fix $i_0 \in I$, which will be the starting state. Let $P \in \Pi$ iff P is a standard stochastic semigroup on a subset I_P of I, with I_P a single recurrent class of stable states and $i_0 \in I_P$. Again, Π is a standard Borel space. Let P_{i_0} make $\{X_t\}$ a Markov chain with stationary transitions P, starting from i_0. Thus, P_{i_0} is a probability on $\mathcal{F}$.

The theorem characterizes $\mathcal{P}$ which are mixtures of $P \in \Pi$, using a certain kind of symmetry (Freedman 1962, 1963; Diaconis and Freedman, 1980a). To state the condition, let σ and τ be finite sequences of states in I. Write σ_i for the ith element of σ; suppose $\sigma_0 = \tau_0 = i_0$. Write $\sigma \sim \tau$ iff σ

and τ exhibit the same number of transitions from i to j, for every pair of states $i, j \in I$. For instance,

$$1\ 2\ 1\ 3\ 2 \sim 1\ 3\ 2\ 1\ 2 \quad \text{but} \quad 1\ 2\ 3 \not\sim 1\ 3\ 2.$$

As is easily seen, if $\sigma \sim \tau$, the two sequences have the same length and end at the same state. The symmetry condition is the following: say

"the $\mathcal{P}$-law of $\{X_{nh} : n = 0, 1, \ldots\}$ depends only on the transition counts"

iff $\sigma \sim \tau$ entails

$$\mathcal{P}\{X_{nh} = \sigma_n \text{ for } n = 0, 1, \ldots, N\} = \mathcal{P}\{X_{nh} = \tau_n \text{ for } n = 0, 1, \ldots, N\},$$

where N is the length of σ.

THEOREM 2. $\mathcal{P} = \int_\Pi P_{i_0} \mu(dP)$ for some probability μ on Π iff:
(i) $\mathcal{P}(X_0 = i_0) = 1$, and
(ii) $\{X_t\}$ has no fixed points of discontinuity, and
(iii) $\mathcal{P}\{X_n = i_0$ for infinitely many integers $n\} = 1$, and
(iv) for each $h > 0$, the $\mathcal{P}$-law of $\{X_{nh} : n = 1, 2, \ldots\}$ depends only on the transition counts.
The mixing measure μ is unique.

Neither theorem requires smoothness conditions on sample paths. In both theorems, necessity is obvious, and the uniqueness of μ follows from corresponding results in discrete time. Sufficiency is proved by approximation through the binary rationals R, and only h of the form $1/2^k$, $k = 1, 2, \ldots$ are needed in Theorem 1(iii) or Theorem 2(iv).

The two proofs follow the program of Diaconis and Freedman (1981), and are very similar. Theorem 1 is proved in section 2: it is shown that, conditional on a certain remote σ-field, the process has stationary, independent increments. Characterizations are also given for mixtures of Brownian motions or Poisson processes; connections are made with the Laplace transform—analogous to the connection between de Finetti's theorem for coin tossing and the Hausdorff moment problem. Extensions could be made to processes taking values in Euclidean space, or second-countable, locally compact abelian groups; that will not be done here. The proof of Theorem 2 is somewhat more technical, and it is given in section 3. Possible generalizations, and the connection with David Blackwell's work, are discussed in section 4.

Markov chains are discussed in Chung (1967), also see Freedman (1983). For surveys on de Finetti's theorem, see Aldous (1985) or Diaconis and Freedman (1981); section 10 in the first reference has a nice discussion of the present Theorem 1. Kallenberg (1973) has a weak convergence argument; Kallenberg (1982) gives a martingale argument which is connected to ideas of Ryll-Nardzewski (1957). For a proof via Choquet theory, see Accardi and Lu (1993). These papers also have interesting extensions. Other work on de Finetti's theorem will be referenced, below.

2. The Proof of Theorem 1. The first lemma is easy, and the proof is omitted.

LEMMA 2.1. Let μ_n, μ be probabilities in j-dimensional Euclidean space. Suppose $\mu_n \to \mu$ in the weak-star topology. Define the real-valued function f as the sum, $f(x_1, \ldots, x_j) = x_1 + \cdots + x_j$. Then $\mu_n f^{-1} \to \mu f^{-1}$.

Next, a brief review of de Finetti's theorem for real-valued random variables. Suppose that $\{\xi_n\}$ are real-valued and exchangeable, on some probability triple $(\Omega, \mathcal{F}, \mathcal{P})$. Let $\mathcal{T}$ be the tail σ-field of the ξ's, and let $\phi_n(\omega)$ be the empirical distribution of the $\xi_1(\omega), \ldots, \xi_n(\omega)$, assigning mass $1/n$ to each $\xi_j(\omega)$. Let $F_\omega = \lim_{n\to\infty} \phi_n(\omega)$, on the set $G(\xi)$ where the weak-star limit exists.

A fairly standard version of de Finetti's theorem asserts that $G(\xi)$ is in $\mathcal{T}$, with $\mathcal{P}\{G(\xi)\} = 1$; and the ξ's are conditionally independent given $\mathcal{T}$, with common distribution F_ω. The next lemma states this more precisely, and a little bit more. Indeed, let $\phi_{j,n}(\omega)$ be the empirical distribution of the first n j-tuples of ξ, assigning mass $1/n$ to $\left(\xi_{(\nu-1)j+1}(\omega), \ldots, \xi_{\nu j}(\omega)\right)$ for each $\nu = 1, \ldots, n$. Thus, $\phi_{j,n}(\omega)$ is a (discrete) probability in Euclidean j-space. Let $G_j(\xi)$ be the set of $\omega \in G(\xi)$ with $\phi_{j,n}(\omega) \to F_\omega^{\times j}$ as $n \to \infty$; again, it is weak-star convergence that is at issue. Here, $\times$ is the cartesian product; if F is a distribution function, $F^{\times n}$ is the n-fold product of F with itself—a distribution in Euclidean n-space. Likewise, $\star$ denotes convolution, so $F^{\star n}$ is the n-fold convolution of F with itself—a distribution on the line.

LEMMA 2.2.
(i) $G_j(\xi) \in \mathcal{T}$ and $\mathcal{P}\{G_j(\xi)\} = 1$ for $j = 1, 2, \ldots$.
(ii) Given the tail σ-field $\mathcal{T}$, a regular conditional $\mathcal{P}$-distribution for the ξ's makes them independent with common distribution F_ω.

REMARKS.
(i) For all $\omega \in G_j(\xi)$, as $n \to \infty$, the empirical distribution of the n sums

$$\xi_1(\omega) + \cdots + \xi_j(\omega), \ldots, \xi_{nj-1}(\omega) + \cdots + \xi_{nj}(\omega)$$

converges weak-star to $F_\omega^{\star j}$.
(ii) Let $G_\infty(\xi) = \bigcap_j G_j(\xi)$. Then $G_\infty(\xi) \in \mathcal{T}$ and $\mathcal{P}\{G_\infty(\xi)\} = 1$.

PROOF. The case $j = 1$ in the Lemma may be found in, e.g., (Diaconis and Freedman, 1980b). The case $j > 1$ follows, or may be proved by similar arguments. Remark (i) follows from Lemma 2.1, and (ii) is clear. ◊

Recall that R is the binary rationals, while the probability $\mathcal{P}$ on $\mathcal{F}$ makes the coordinate process $\{X_t\}$ have exchangeable increments, in the sense of Theorem 1. Let $h = 1/2^k$ for $k = 0, 1, \ldots$.

LEMMA 2.3. Suppose $\{X_r : r \in R\}$ has stationary, independent increments. For each fixed real $t \geq 0$, as $r \to t$, X_r converges a.e.

PROOF. This is well known; the restriction of the time domain to a countable set is critical. A relatively simple direct argument can be made by considering for each $s \in R$ the martingale

$$r \to e^{isX_r}/E\left\{e^{isX_r}\right\}. \qquad \diamondsuit$$

Let $\mathcal{F}_h$ be the tail σ-field of $\{X_{nh} - X_{(n-1)h} : n = 1, 2, \ldots\}$. Given $\mathcal{F}_h$, the differences $X_{nh} - X_{(n-1)h}$ are independent with common distribution $F_{h,\omega}$. Lemma 2.4 states this more carefully.

LEMMA 2.4. There is a set $G_h \in \mathcal{F}_h$ with $\mathcal{P}(G_h) = 1$, and for each $\omega \in G_h$ a distribution function $F_{h,\omega}$, such that
(i) $\omega \to F_{h,\omega}$ is $\mathcal{F}_h$-measurable, and
(ii) a regular conditional $\mathcal{P}$-distribution for the differences $X_{nh} - X_{(n-1)h}$ given $\mathcal{F}_h$ makes these differences independent with common distribution $F_{h,\omega}$, and
(iii) $G_h \subset G_{2h}$ for $h \leq 1/2$, and
(iv) $\omega \in G_h$ entails $F_{h,\omega} \star F_{h,\omega} = F_{2h,\omega}$ for $h \leq 1/2$.

PROOF. This is immediate from Lemma 2.2: let $\xi_n = X_{nh} - X_{(n-1)h}$, and let G_h be the $G_\infty(\xi)$ in Remark (ii) after Lemma 2.2; the only j's of interest are the powers of 2. $\diamondsuit$

Clearly, $\mathcal{F}_h$ increases as $h \downarrow 0$; let $\mathcal{F}_0$ be the σ-field spanned by $\bigcup_h \mathcal{F}_h$. Let $G_0 = \bigcap_h G_h$. Then $G_0 \in \mathcal{F}_0$ and $\mathcal{P}(G_0) = 1$. In view of Lemma 2.4(iv), we can define $F_{j/2^k,\omega} = \left(F_{1/2^k,\omega}\right)^{\star j}$, with no ambiguity. Then $S_\omega = \{F_{r,\omega} : r \in R\}$ is a convolution semigroup for each $\omega \in G_0$:

$$F_{(j+k)/2^n,\omega} = F_{1/2^n,\omega}^{\star(j+k)} = F_{1/2^n,\omega}^{\star j} \star F_{1/2^n,\omega}^{\star k} = F_{j/2^n,\omega} \star F_{k/2^n,\omega}.$$

The martingale convergence theorem and Lemma 2.4 yield the following. Given $\mathcal{F}_0$, a regular conditional $\mathcal{P}$-distribution for $\{X_r : r \in R\}$ makes $\{X_r : r \in R\}$ have stationary, independent increments governed by S_ω. Let $Q(\omega, A)$ denote this regular conditional distribution. In particular, $Q(\omega, \cdot)$ can be viewed for each $\omega \in G_0$ as a probability on the σ-field in Ω spanned by $\{X_r : r \in R\}$. Recall that $\mathcal{F}$ is the σ-field Ω spanned by $\{X_t : t \geq 0\}$.

LEMMA 2.5. For each $\omega \in G_0$, the probability $Q(\omega, \cdot)$ can be extended to all of $\mathcal{F}$. Call this extension $\tilde{Q}(\omega, \cdot)$. Then $\tilde{Q}$ is an rcd for $\{X_t\}$ given $\mathcal{F}_0$, relative to $\mathcal{P}$.

PROOF. Fix non-negative real times $t_1 < t_2 < \cdots < t_k$ and bounded continuous functions $f_1, f_2, \ldots, f_k$. Let $r_j \to t_j$ through R. The $\tilde{Q}(\omega, \cdot)$-integral of

$$\prod_{j=1}^{k} f_j(X_{t_j})$$

can by Lemma 2.3 be safely defined as

$$\lim_{r_j \to t_j} \int \prod_{j=1}^{k} f_j(X_{r_j})\, dQ(\omega, \cdot).$$

The rest of the extension argument is omitted. That $\tilde{Q}$ is an rcd for $\{X_t\}$ given $\mathcal{F}_0$ follows from condition (ii) of the theorem. $\diamondsuit$

Theorem 1 is more or less immediate from Lemma 2.5. Two special cases may be worth noting.

1) Suppose that a version of $\{X_t\}$ has continuous sample paths, under $\mathcal{P}$. Then Ω can be taken as the set of continuous functions, $Q(\omega, \cdot)$ can be restricted to the set of functions on R with continuous extensions to $[0, \infty)$, and $\tilde{Q}(\omega, \cdot)$ can be restricted to the set of continuous functions on $[0, \infty)$. Of course, a process with continuous sample paths and stationary, independent increments is Brownian motion. In consequence, $\mathcal{P}$ must be a mixture of Brownian motions with different scale and drift parameters.

2) Suppose that a version of $\{X_t\}$ has sample paths which are step functions, under $\mathcal{P}$. By similar reasoning, $\mathcal{P}$ is a mixture of Poisson processes. Implications are discussed in section 4.

3. The Proof of Theorem 2.

Let $\{\xi_n : n = 0, 1, \ldots\}$ be a discrete-time process (not necessarily Markov), whose state space is a subset of the countable set I. Let ν_{in} be the number of visits to $i \in I$ up to time n, i.e., the number of indices $m < n$ with $\xi_m = i$. Similarly, ν_{ijn} is the number of doublets ij up to time n, i.e., the number of indices $m < n$ with $\xi_m = i$ and $\xi_{m+1} = j$.

CONDITION 3.1. Let $\{\xi_n\}$ be a discrete-time Markov chain with stationary transitions P, starting from i_0; suppose that ξ_n returns to i_0 infinitely often with probability 1.

LEMMA 3.1. Condition 3.1 is in force. For each $i \in I$, either
(i) $\nu_{in} = 0$ a.e. for all n, or
(ii) $\nu_{in} \to \infty$ a.e. as $n \to \infty$, and $\nu_{ijn}/\nu_{in} \to P(i, j)$ a.e. for all $j \in I$.
Let $I_0 = \{i : i \in I \text{ and } \nu_{in} \to \infty \text{ a.e.}\}$. Then P is a stochastic matrix on $I_0 \ni i_0$, and I_0 forms a single, recurrent class of states relative to P.

LEMMA 3.2. Let P be a stochastic matrix on I_0; suppose I_0 consists of one recurrent class of states; and suppose that $P(i_0, i_0) > 0$ for some $i_0 \in I_0$. Then I_0 consists of one recurrent class of states for P^2.

Proofs are omitted; Lemma 3.1 is well known, and 3.2 is routine. Next, a review of de Finetti's theorem for discrete-time chains. As before, let $I_\omega = \{i : i \in I \text{ and } \nu_{in}(\omega) \to \infty\}$. Let $P_\omega(i,j) = \lim_{n\to\infty} \nu_{ijn}(\omega)/\nu_{in}(\omega)$ when this limit exists. Let $\mathcal{T}$ be the tail σ-field of ξ_n. let G be the set of ω such that, for all $i \in I$, either

(i) $\nu_{in}(\omega) = 0$ for all n, or

(ii) $\nu_{in}(\omega) \to \infty$, and $\nu_{ijn}(\omega)/\nu_{in}(\omega) \to P_\omega(i,j)$ as $n \to \infty$ for all $j \in I$, and $\sum_{j\in I} P_\omega(i,j) = 1$.

Plainly, $G \in \mathcal{T}$, $P_\bullet$ is $\mathcal{T}$-measurable, and $\{\omega : j \in I_\omega\} \in \mathcal{T}$.

CONDITION 3.2. Let $\{\xi_n\}$ be a discrete-time process on the probability triple $(\Omega, \mathcal{F}, \mathcal{P})$, starting from i_0; suppose the law of $\{\xi_n\}$ depends only on the transition counts, as defined in section 1; and suppose that ξ_n returns to i_0 infinitely often with probability 1.

LEMMA 3.3. Condition 3.2 entails $\mathcal{P}(G) = 1$; furthermore, a regular conditional $\mathcal{P}$-distribution for $\{\xi_n\}$ given $\mathcal{T}$ makes this process Markov with stationary transitions P_ω; I_ω is one recurrent class of states relative to P_ω; and $i_0 \in I_\omega$.

PROOF. The Lemma follows from Diaconis and Freedman (1980a) via Lemma 3.1. ◇

The next lemma demonstrates a kind of consistency among the regular conditional distributions given the tail σ-fields for $\xi_n : n = 0, 1, 2, \ldots$ and for $\xi_{2n} : n = 0, 1, 2, \ldots$. In discrete time, the result may seem a bit artificial; its utility will be apparent later. Let $\mathcal{T}_2$ be the tail σ-field of ξ_{2n}; let $\nu_{2,in}$ be ν_{in} applied to ξ_{2n}, i.e., the number of indices $m = 0, 1, \ldots, n-1$ with $\xi_{2m} = i$. Define $\nu_{2,ijn}$, G_2, and $P_{2,\omega}$ in the analogous way. It is an irritating feature of the situation that $\nu_{2,in} \neq \nu_{i(2n)}$. The condition for the next lemma is a bit stronger than Condition 3.2, because recurrence at even times is required.

CONDITION 3.3. Let $\{\xi_n\}$ be a discrete-time process on the probability triple $(\Omega, \mathcal{F}, \mathcal{P})$, starting from i_0; suppose the law of $\{\xi_n\}$ depends only on the transition counts, as defined in section 1; and suppose

$$\mathcal{P}\{\xi_{2n} = i_0 \text{ infinitely often }\} = 1.$$

LEMMA 3.4. Condition 3.3 implies that $\mathcal{P}(G_2) = 1$, and a regular conditional $\mathcal{P}$-distribution for $\{\xi_{2n}\}$ given $\mathcal{T}_2$ makes this process Markov with stationary transitions $P_{2,\omega}$; there is one recurrent class of states $I_{2,\omega}$ and $i_0 \in I_{2,\omega}$. Furthermore, $I_{2,\omega} = I_\omega$ and $P_{2,\omega} = P_\omega^2$ for almost all ω with $P_\omega(i_0, i_0) > 0$.

PROOF. The first assertion is just Lemma 3.3 applied to ξ_{2n}. Plainly, $I_{2,\omega} \subset I_\omega$; if i_0 has period 2, the inclusion may be strict—but that is precluded when $P_\omega(i_0, i_0) > 0$: see Lemma 3.2. Now, given $\mathcal{T}$, ξ_{2n} is Markov with transitions P_ω^2; but, given $\mathcal{T}_2$, this process is Markov with transitions $P_{2,\omega}$. Lemma 3.1 completes the argument. $\diamondsuit$

Turn now to continuous time. Let $\mathcal{P}$ satisfy the conditions of Theorem 2. Let $h = 1/2^k$ for some $k = 0, 1, \ldots$. Let $\mathcal{F}_h$ be the tail σ-field of $\{X_{nh} : n = 0, 1, \ldots\}$. The proof of Lemma 3.5 is omitted as routine.

LEMMA 3.5. $\mathcal{F}_h$ increases as $h \downarrow 0$.

Let $\mathcal{F}_0$ be the σ-field spanned by $\bigcup_h \mathcal{F}_h$. The program is to construct a set $G_0 \in \mathcal{F}_0$ with $\mathcal{P}(G_0) = 1$, and for each $\omega \in G_0$ a standard stochastic semigroup $P_{t,\omega}$ with state space I_ω such that

(i) $i_0 \in I_\omega$, and
(ii) I_ω consists of one recurrent class of states relative to $P_{t,\omega}$, and
(iii) given $\mathcal{F}_0$, a regular conditional $\mathcal{P}$-distribution for $\{X_t\}$ makes this process Markov with stationary transitions $P_{t,\omega}$ starting from i_0.

The next lemma proves the analogous—but easier—result for $\mathcal{F}_h$ and $\{X_{nh} : n = 0, 1, \ldots\}$. Use the notation of Lemmas 3.1–3. Fix $h = 1/2^k$. Let ν_{hin} be the number of indices $m < n$ with $X_{mh} = i$; similarly, ν_{hijn} is the number of $m < n$ with $X_{mh} = i$ and $X_{(m+1)h} = j$. Let $I_{h,\omega}$ be the set of $i \in I$ with $\nu_{hin}(\omega) \to \infty$ as $n \to \infty$. Let $P_{h,\omega}(i,j) = \lim_{n\to\infty} \nu_{hijn}(\omega)/\nu_{hin}(\omega)$ when this limit exists. Let G_h be the set of ω such that, for all $i \in I$, either

(i) $\nu_{hin}(\omega) = 0$ for all n, or
(ii) $\nu_{hin}(\omega) \to \infty$ as $n \to \infty$, and $\nu_{hijn}(\omega)/\nu_{hin}(\omega) \to P_{h,\omega}(i,j)$ for all $j \in I$, and $\sum_{j\in I} P_{h,\omega}(i,j) = 1$.

We also require of $\omega \in G_h$ that $I_{h,\omega}$ consists of one recurrent class of states relative to $P_{h,\omega}$, and $i_0 \in I_{h,\omega}$.

LEMMA 3.6. Fix $h = 1/2^k$.

(i) $\omega \to P_{h,\omega}$ is $\mathcal{F}_h$-measurable.
(ii) $G_h \in \mathcal{F}_h$ and $\mathcal{P}(G_h) = 1$.
(iii) Given $\mathcal{F}_h$, a regular conditional $\mathcal{P}$-distribution for $\{X_{nh} : n = 0, 1, \ldots\}$ makes this process Markov with stationary transitions $P_{h,\omega}$ starting from i_0.

PROOF. This is immediate from Lemma 3.3. $\diamondsuit$

The next step is to make sure that the P_h fit together properly as $h = 1/2^k$ varies; Lemma 3.7 is a preliminary.

LEMMA 3.7.
(i) $P_{h,\bullet}(i_0, i_0) \to 1$ in $\mathcal{P}$-probability as $h \to 0$.
(ii) $P_{h,\omega}(i_0, i_0) \to 1$ for $\mathcal{P}$-almost all ω, as $h \to 0$ rapidly.
(iii) $P_{h,\omega}(i_0, i_0) > 0$ for all $h = 1/2^k$, for $\mathcal{P}$-almost all ω.

PROOF. Of course,

$$\int P_{h,\omega}(i_0, i_0)\, \mathcal{P}(d\omega) = \mathcal{P}(X_h = i_0) \to 1$$

as $h \to 0$ by conditions (i) and (ii) of the theorem. This proves (i); claim (ii) is immediate, and then (iii) follows. $\diamondsuit$

Let $\bar{G}_h$ be the set of $\omega \in G_h \cap G_{2h}$ such that $P_{h,\omega}(i_0, i_0) > 0$, $I_{2h,\omega} = I_{h,\omega}$, and $P^2_{h,\omega} = P_{2h,\omega}$.

LEMMA 3.8. Fix $h = 1/2^k$. Then $\bar{G}_h \in \mathcal{F}_h$ and $\mathcal{P}(\bar{G}_h) = 1$.

PROOF. Measurability follows from Lemmas 3.5–6; and the probability assertion, from Lemmas 3.4 and 3.7(iii). $\diamondsuit$

Let $G = \bigcap_h \bar{G}_h$. Recall $\mathcal{F}_0$ is the σ-field spanned by $\bigcup_h \mathcal{F}_h$. Plainly, $G \in \mathcal{F}_0$ and $\mathcal{P}(G) = 1$.

LEMMA 3.9.
(i) For all $\omega \in G$ and $h = 1/2^k$, $I_{h,\omega} = I_{1,\omega}$; abbreviate $I_\omega = I_{1,\omega}$.
(ii) Suppose n_1 and n_2 are positive integers, and $n_1 h_1 = n_2 h_2$. For all $\omega \in G$, $P^{n_1}_{h_1,\omega} = P^{n_2}_{h_2,\omega}$.

PROOF. $h_1 = 1/2^{k_1}$ and $h_2 = 1/2^{k_2}$; without loss of generality, suppose $k_1 > k_2$. Then $n_1 = n_2 2^{k_1 - k_2}$ and

$$P^{n_1}_{h_1,\omega} = P^{n_1/2}_{2h_1,\omega} = P^{n_1/4}_{4h_1,\omega} = \cdots = P^{n_2}_{h_2,\omega}. \qquad \diamondsuit$$

In particular, for all $\omega \in G$, $P_{h,\omega}$ extends to a stochastic semigroup $S_\omega = \{P_{r,\omega} : r \in R\}$, where R consists of the non-negative binary rationals; the matrices are all defined on the state space $I_\omega \ni i_0$. Indeed, $P_{j/2^k,\omega}$ is well-defined as $P^j_{1/2^k,\omega}$ by Lemma 3.9; and the semi-group property follows:

$$P_{(j+k)/2^n,\omega} = P^{j+k}_{1/2^n,\omega} = P^j_{1/2^n,\omega} P^k_{1/2^n,\omega} = P_{j/2^n,\omega} P_{k/2^n,\omega}.$$

LEMMA 3.10. Given $\mathcal{F}_0$, relative to $\mathcal{P}$, an rcd for $\{X_r : r \in R\}$ makes this process Markov with stationary transitions S_ω starting from i_0.

PROOF. Use the martingale convergence theorem and Lemma 3.6. $\diamondsuit$

The next objective is extending $P_{\bullet,\omega}$ from R to $[0,\infty)$. Recall that $\{X_t\}$ is the coordinate process on Ω.

LEMMA 3.11. Fix $t \geq 0$; $\lim_{r\to t} X_r(\omega) = X_t(\omega)$ for $\mathcal{P}$-almost all ω: as usual, r is restricted to R.

PROOF. This follows from condition (ii) of the Theorem 2, by a standard countable-additivity argument. Fix $j \in I$ and let $A = \{\omega : X_t(\omega) = j\}$. Let

$$A_n = \{\omega \in A : \exists r \in R \text{ with } |t-r| < 1/n \text{ and } X_r(\omega) \neq j\}.$$

In principle, the r in the definition of A_n may depend on ω. Of course, A_n is monotone decreasing as n increases; suppose by way of contradiction that $\mathcal{P}(A_n) > \epsilon > 0$ for all n. There would then be a rapidly growing—but deterministic—sequence $g(n)$ of positive integers such that $\mathcal{P}(\bigcap_n B_n) > \epsilon/2$, where

$$B_n = \{\omega \in A : \exists\ r \in R \text{ with } |t-r| < 1/n \\ \text{and } \operatorname{order}(r) < g(n) \text{ and } X_r(\omega) \neq j\}.$$

(The "order" of r is the least integer k such that $2^k r$ is an integer.) Let now s_m be the deterministic sequence which enumerates the binary rationals r with $\operatorname{order}(r) < g(n)$ and $|t-r| < 1/n$; here, r_1 precedes r_2 if $\operatorname{order}(r_1) < \operatorname{order}(r_2)$, or $\operatorname{order}(r_1) = \operatorname{order}(r_2)$ and $r_1 < r_2$. Of course, $s_m \to t$ while $\mathcal{P}\{X_{s_m} \not\to X_t\} > \epsilon/2$. This violates condition (ii) of the Theorem, a contradiction which proves the Lemma; that R is countable is a critical ingredient. ◇

Lemma 3.12. For $\mathcal{P}$-almost all $\omega \in G$,

$$P_{r,\omega}(j,j) \to 1 \text{ as } r \to 0 \text{ through } R, \text{ for each } j \in I_\omega. \tag{3.1}$$

PROOF. Let $H_j = \{\omega \in G : j \in I_\omega\}$ and $H_{j,n} = \{\omega \in G : P_{n,\omega}(i_0, j) > 0\}$, so that $H_j = \bigcup_n H_{j,n}$. If $\mathcal{P}(H_j) = 0$, there is nothing to prove; otherwise, find an n with $\mathcal{P}(H_{j,n}) > 0$. By Lemma 3.11, for $\mathcal{P}$-almost all ω, given $X_n(\omega) = j$, $X_r(\omega)$ must equal j for all $r \in \big[n, n+\epsilon(\omega)\big)$. This remains true conditional on $\mathcal{F}_0$, and the Lemma follows. ◇

We may assume that (3.1) holds for all $\omega \in G$.

LEMMA 3.13. Fix $\omega \in G$, restrict r to R and i,j,k to I_ω. Claim (iii) holds for all $t \geq 0$; claim (iv) holds for all $t, s \geq 0$.

(i) $P_{r,\omega}(j,k)$ is uniformly continuous in r.
(ii) $P_{\bullet,\omega}(j,k)$ extends to a continuous function on $[0,\infty)$.
(iii) $P_{t,\omega}$ is a substochastic matrix on I_ω.
(iv) $P_{t+s,\omega}(i,k) \geq \sum_{j\in I_\omega} P_{t,\omega}(i,j) P_{s,\omega}(j,k)$.

PROOF. The argument is straightforward. To begin with, for $r, s \in R$ and $j, k \in I_\omega$,

$$P_{r+s,\omega}(j,k) = \sum_{i \in I_\omega} P_{s,\omega}(j,i) P_{r,\omega}(i,k).$$

Therefore,

$$P_{r+s,\omega}(j,k) - P_{r,\omega}(j,k) = [P_{s,\omega}(j,j) - 1]\, P_{r,\omega}(j,k) + \sum_{i \neq j} P_{s,\omega}(j,i) P_{r,\omega}(i,k).$$

The first term on the right is negative; the second is positive; each is in absolute value bounded by $1 - P_{s,\omega}(j,j)$. Thus

$$|P_{r+s,\omega}(j,k) - P_{r,\omega}(j,k)| \leq 1 - P_{s,\omega}(j,j).$$

Lemma 3.12 completes the proof of (i); the continuity is even uniform in k, although that will not matter here. Claim (ii) immediate. Claims (iii) and (iv) now follow via Fatou's lemma. $\diamondsuit$

Fix $\omega \in G$. If $i \in I_\omega$ and $j \notin I_\omega$, then $P_{r,\omega}(i,j) = 0$ for all $r \in R$. (See Lemma 3.6.) Setting $P_{t,\omega}(i,j) = 0$ gives the continuous extension to $t \geq 0$. If $i \notin I_\omega$, set $P_{t,\omega}(i,j) = 0$ for all $t \geq 0$ and $j \in I$. Again, this gives continuity.

LEMMA 3.14. Fix a sequence of times $0 = t_0 < t_1 < t_2 < \cdots < t_k$ and states $i_0, i_1, i_2, \ldots, i_k \in I$. Let

$$A = \{X_{t_m} = i_m \text{ for } m = 0, \ldots, n\}.$$

Let $B \in \mathcal{F}_0$. Then

$$\mathcal{P}(A \cap B) = \int_B \left[\prod_{m=0}^{n-1} P_{t_{m+1}-t_m,\omega}(i_m, i_{m+1})\right] \mathcal{P}(d\omega). \tag{3.2}$$

The integrand in (3.2) is 0 unless $i_m \in I_\omega$ for all m.

PROOF. Equation (3.2) holds for binary rational t by Lemma 3.10. Now approximate real t_m by binary rationals. The left side of (3.2) converges to the correct limit by condition (ii) of the Theorem. The right side can be handled by Lemma 3.13(ii) and dominated convergence. $\diamondsuit$

We do not yet know that $P_{t,\omega}$ is a standard stochastic semigroup, so Lemma 3.14 is not the end of the road—but it is close. Recall from Lemma 3.12 that $H_j = \{\omega \in G : j \in I_\omega\}$ and $H_{j,n} = \{\omega \in G : P_{n,\omega}(i_0, j) > 0\}$.

LEMMA 3.15. Fix $t \geq 0$ and $j \in I$. For $\mathcal{P}$-almost all $\omega \in H_j$,

$$\sum_{k \in I_\omega} P_{t,\omega}(j,k) = 1.$$

PROOF. By Lemma 3.14,

$$\mathcal{P}(X_n = j) = \int_{H_{j,n}} P_{n,\omega}(i_0, j)\, \mathcal{P}(d\omega) \tag{3.3}$$

$$\mathcal{P}(X_n = j,\ X_{n+t} = k) = \int_{H_{j,n}} P_{n,\omega}(i_0, j) P_{t,\omega}(j,k)\, \mathcal{P}(d\omega). \tag{3.4}$$

We may suppose that $\mathcal{P}(H_j) > 0$; now fix n so the left side of equation (3.3) is positive. Sum equation (3.4) over k. Since $\sum_k \mathcal{P}(X_n = j,\ X_{n+t} = k) = \mathcal{P}(X_n = j)$, and $\sum_k P_{t,\omega}(j,k) \leq 1$ with $\mathcal{P}$-probability 1 by Lemma 3.13(iii), the Lemma follows. $\diamondsuit$

In principle, the exceptional null set in Lemma 3.15 could depend on t; that difficulty is eliminated by the next result.

LEMMA 3.16. For $\mathcal{P}$-almost all $\omega \in G$, $\{P_{\bullet,\omega}\}$ is a standard stochastic semigroup.

PROOF. Let $\mathcal{B}$ be the Borel σ-field in $[0,\infty)$. Plainly, $(t,\omega) \to P_{t,\omega}(i,j)$ is $\mathcal{B} \times \mathcal{F}_0$-measurable. Restrict ω to G. Let

$$L_\omega = \{t : \sum_{j \in I} P_{t,\omega}(i,j) = 1 \text{ for all } i \in I_\omega\}.$$

Let $\bar{L}_\omega$ be the complement of L_ω with respect to $[0,\infty)$. Of course, L_ω and $\bar{L}_\omega$ are Borel subsets of the line. By Lemma 3.15 and Fubini's theorem, for $\mathcal{P}$-almost all ω, $\bar{L}_\omega$ has Lebesgue measure 0. On the other hand, L_ω is closed under addition, by Lemma 3.13. Thus, $L_\omega = [0,\infty)$ for $\mathcal{P}$-almost all ω. For such ω, $P_{t,\omega}$ is a stochastic matrix for all $t \geq 0$; then the semigroup property also follows from Lemma 3.13. $\diamondsuit$

REMARK. Let $G_0 = \{\omega : \omega \in G \text{ and Lebesgue}(\bar{L}_\omega) = 0\}$. Then $G_0 \in \mathcal{F}_0$, $\mathcal{P}(G_0) = 1$, and $P_{\bullet,\omega}$ is a standard stochastic semigroup for each $\omega \in G_0$.

As noted before, $I_\omega \ni i_0$ and is a single recurrent class of stable states relative to $P_{\bullet,\omega}$. (See Lemma 3.6.) Lemmas 3.14–16 prove the next result, which in turn gives the theorem.

Proposition 3.1. Relative to $\mathcal{P}$, given $\mathcal{F}_0$, the process $\{X_t\}$ is conditionally Markov with stationary transitions $P_{t,\omega}$.

4. Discussion. We discuss some features of Theorem 2, then turn back to Theorem 1. Loosely speaking, a state in a Markov chain is "instantaneous" if the process stays there for no proper interval of time. Thus, if j is instantaneous and $X_t = j$, there must be binary rationals r converging to t from the right with $X_r \to \infty$; of course, there will also be r—and in some sense many more of them—with $X_r = j$. David Blackwell (1958) gave a beautiful example of a chain whose states were all instantaneous. Theorem 2 excludes such cases, by assumption. If condition (ii) is replaced by a continuity-in-probability assumption, the theorem may go through and cover the instantaneous case. In particular, if $\{P_r\}$ is a stochastic semigroup on binary rational times, it is conceivable that $P_\bullet(i,j)$ automatically extends to a continuous function on $[0,\infty)$. This would be a good substitute for Lemma 2.1. (That there may also be non-measurable extensions is one of the charming complications.)

Turn back now to Theorem 1 for Poisson processes. As pointed out by David Aldous and Persi Diaconis, there is an interesting connection with the theory of the Laplace transform, analogous to the idea of using de Finetti's theorem for coin tossing to solve the Hausdorff moment problem (Feller, 1971, p.228). Let L be a function on $[0,\infty)$. The question to be addressed is this: when is there a probability μ on $[0,\infty)$ such that $L(t) = \int_0^\infty e^{-\lambda t}\,\mu(d\lambda)$?

Necessary conditions are that $L(0) = 1$ and L is C_∞ while $L' \leq 0$, $L'' \geq 0$, etc. According to Bernstein's theorem, these conditions are also sufficient (Feller, 1971, p.439). For a probabilistic proof, we want to construct a process $\{X_t\}$ with exchangeable increments, whose sample functions are counting functions, and $L(t) = P(X_t = 0)$. This seems hard to do directly; instead, we make a "completely exchangeable" process of trees $T_0, T_1, T_2, \ldots$. More specifically, $T_n = \{X_{ns}\}$, where $X_{ns} = 0$ or 1, and the node ns consists of the non-negative integer n followed by a finite string s (perhaps empty) of 0's and 1's. These T_n are required to be exchangeable. Also, each T_n splits into T_{n0} and T_{n1}: the fragments $T_{10}, T_{11}, T_{20}, T_{21}, \ldots$ are required to be exchangeable too. And so on.

We require that each variable be the maximum of the variables at the two successor nodes, so $X_{ns} = X_{ns0} \vee X_{ns1}$. Finally, we require that

$$P\{\text{first } j \text{ variables at level } k \text{ are } 0\} = L(j/2^k). \tag{4.1}$$

Here, the nodes are ordered lexicographically. For instance, the first three nodes at level 0 are 0, 1, 2; the first six nodes at level 1 are 00, 01, 10, 11, 20, 21; and so forth. The nodes correspond to sub-intervals of $[0,\infty)$; e.g., the node n corresponds to the interval $[n, n+1]$, the node $n0$ to $[n, n+\frac{1}{2}]$, the node $n1$ to $[n+\frac{1}{2}, n+1]$, etc. The idea is that $X_{ns} = 0$ iff there is no dot in the corresponding interval for the counting process—which is yet to be constructed.

What has to be checked is that (4.1)—and exchangeability—specifies the joint distributions consistently down to level k; then the Kolmogorov consistency theorem can be used to get the infinite tree. For example, why does (4.1) give the full joint distribution at level 0? For instance,

$$\begin{aligned} P\{X_0 = 1,\ X_1 = \cdots = X_{N-1} = 0\} \\ &= P\{X_1 = \cdots = X_{N-1} = 0\} - P\{X_0 = \cdots = X_{N-1} = 0\} \\ &= L(N-1) - L(N) \end{aligned}$$

by (4.1) with $k = 0$—and exchangeability; $L(N-1) \geq L(N)$ because $L' \leq 0$. And so forth.

Why is (4.1) consistent for levels 0 and 1? Construct the level 1 variables

$$X_{00}, X_{01}, X_{10}, X_{11}, X_{20}, X_{21}, \ldots$$

to be exchangeable and satisfy (4.1); then define the level 0 variables as

$$X_0 = X_{00} \vee X_{01},\ X_1 = X_{10} \vee X_{11},\ X_2 = X_{20} \vee X_{21}, \ldots.$$

Now check that the level 0 variables are exchangeable, and

$$\begin{aligned} P\{X_0 = \cdots = X_{N-1} = 0\} &= P\{X_{00} = X_{01} = \cdots = X_{N-1,0} = X_{N-1,1} = 0\} \\ &= L\big[2(N-1)/2\big] = L(N-1). \end{aligned}$$

(The second equality follows from (4.1) with $k = 1$.)

Consider next the tail σ-field of the tree. Clearly, T_n is a 1–1 function of (T_{n0}, T_{n1}). So the tail σ-field of $\{T_n\}$ equals the tail σ-field of $\{T_{ns}\}$, where ns is lexicographically ordered, along any fixed level k. (The ns at level k consists of strings of length $k+1$, beginning with a non-negative integer and continuing with 0's and 1's.)

Condition on the tail σ-field Σ of $\{T_n\}$. Given Σ, we have at level k a set of iid 0–1 variables X_{ns}; each is 0 with conditional probability $p_{k,\omega}$. Clearly, $p_{k,\omega} = p_{k+1,\omega}^2$, and then $p_{k,\omega} = p_{0,\omega}^{1/2^k}$. If $p_{0,\omega} = 0$ then $p_{k,\omega} = 0$ for all k. In this case, let $\lambda_\omega = \infty$. On the other hand, if $p_{0,\omega} > 0$, let $\lambda_\omega = -\log p_{0,\omega}$, so $p_{k,\omega} = \exp(\lambda_\omega/2^k)$ and $0 \leq \lambda_\omega < \infty$. Of course, given Σ, the relationship of nodes in level k to their children at level $k+1$ remains as it was: $X_{ns} = X_{ns0} \vee X_{ns1}$.

In effect, then, we have a version of de Finetti's theorem for our trees. As a consequence, for $j \geq 1$,

$$\begin{aligned} L(j/2^k) &= P\{\text{first } j \text{ variables at level } k \text{ are } 0\} \\ &= \int_{\lambda<\infty} \exp(-\lambda_\omega j/2^k)\, P(d\omega). \end{aligned}$$

On $\{\lambda = \infty\}$, all variables are 1 and $L = 0$. Thus, for $t \geq 0$,

$$L(t) = \int_{\lambda<\infty} e^{-\lambda_\omega t}\, P(d\omega).$$

Let $t \downarrow 0$ to see that $P(\lambda < \infty) = L(0+) = 1$. This completes the proof of the sufficiency part of Bernstein's theorem, using the version of de Finetti's theorem for trees that was sketched above—but not Theorems 1 or 2.

These ideas go back to Choquet (1953-54); also see Kendall (1974) and Matheron (1975). For a derivation through the Martin boundary, see Watanabe (1960). For connections with point processes, see Kurtz (1974), Matthes, Kerstan and Mecke (1978), or Kallenberg (1986, chapter 9). There are references below to a number of other works on exchangeability; some discuss current research, others provide useful reviews; also listed are some papers that initiated major lines of activity.

Acknowledgments. I would like to thank David Aldous, Persi Diaconis, Steve Evans, Lucien Le Cam, and Jim Pitman for useful conversations. An anonymous referee was also very helpful.

REFERENCES

ACCARDI, L. AND LU, Y. G. (1993). A continuous version of de Finetti's theorem. *Ann. Probab.* **21** 1478–1493.

ALDOUS, D. (1985). *Exchangeability and Related Topics.* Springer Lecture Notes in Mathematics **1117**.

BLACKWELL, D. (1958). Another countable Markov process with only instantaneous states. *Ann. Math. Statist.* **29** 313–6.

BÜHLMANN, H. (1960). Austauschbare stochastische Variabeln und ihre Grenzwertsätze. *Univ. Calif. Publ. Statist.* **3** 1–36.

CHOQUET, G. (1953–4). Theory of capacities. *Ann. Inst. Fourier.* **5** 131–295.

CHUNG, K. L. (1967). *Markov Chains with Stationary Transitions.* 2nd ed., Springer, New York.

DIACONIS, P. (1988). Recent progress on de Finetti's notions of exchangeability. *Bayesian Statistics* **3** 111–125. J. M. Bernardo, M. H. Degroot, D. V. Lindley and A. F. M. Smith, eds. Oxford University Press.

DIACONIS, P., EATON M. L., AND LAURITZEN, S. L. (1992). Finite de Finetti theorems in linear models and multivariate analysis. *Scand J. Statist.* **19** 289–315.

DIACONIS, P. AND FREEDMAN, D. (1980a). de Finetti's theorem for Markov chains. *Ann. Prob.* **8** 115–30.

DIACONIS, P. AND FREEDMAN, D. (1980b). Finite exchangeable sequences. *Ann. Prob.* **8** 745–64.

DIACONIS, P. AND FREEDMAN, D. (1981). Partial exchangeability and sufficiency. *Proceedings of the ISI Golden Jubilee International Conference.* Sankhyā. 205–236.

DIACONIS, P. AND FREEDMAN, D. (1987). A dozen de Finetti-style results in search of a theory. *Ann. Inst. Henri Poincaré* **23** 397–423.

DIACONIS, P. AND FREEDMAN, D. (1988). Conditional limit theorems for exponential families and finite versions of de Finetti's theorem. *J. Theoretical Prob.* **1** 381–410.

DIACONIS, P. AND FREEDMAN, D. (1990). Cauchy's equation and de Finetti's theorem. *Scand. J. Statist.* **17** 235–50.

DUBINS, L. AND FREEDMAN, D. (1979). Exchangeable processes need not be mixtures of independent, identically distributed random variables. *Z. Wahrscheinlichkeitstheorie* **48** 115–32.

DYNKIN, E. (1978). Sufficient statistics and extreme points. *Ann. Prob.* **6** 705–30.

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications.* Vol. 2, 2nd. ed. Wiley, New York.

FREEDMAN, D. (1962). Invariants under mixing which generalize de Finetti's theorem. *Ann. Math. Stat.* **33** 916–23.

FREEDMAN, D. (1963). Invariants under mixing which generalize de Finetti's theorem: continuous time parameter. *Ann. Math. Stat.* **34** 1194–1216.

FREEDMAN, D. (1983). *Markov Chains.* Springer, New York.

HEWITT, E. AND SAVAGE, L. J. (1955). Symmetric measures on cartesian products. *TAMS* **80** 470–501.

JENSEN, J. L. (1991). Uniform saddlepoint approximations and log-concave densities. *JRSS B* **53** 157–72.

KALLENBERG, O. (1973). Canonical representations and convergence criteria for processes with interchangeable increments. *Z. Wahrscheinlichkeitstheorie* **27** 23–36.

KALLENBERG, O. (1982). Characterizations and embedding properties in exchangeability. *Z. Wahrscheinlichkeitstheorie* **60** 249–81.

KALLENBERG, O. (1986). *Random Measures.* 4th ed. Akademie-Verlag and Academic Press.

KALLENBERG, O. (1992). Symmetries on random arrays and set-indexed processes. *J. Theoretical Prob.* **5** 727–765.

KENDALL, D. G. (1974). Foundations of a theory of random sets. *Stochastic Geometry* 322–376. E. F. Harding and D. G. Kendall, eds. Wiley, New York.

KÜCHLER, U. AND LAURITZEN, S. L. (1989). Exponential families, extreme point models and minimal space-time invariant functions for stochastic processes with stationary and independent increments. *Scand. J. Statist.* **16** 237–261.

KURTZ, T. G. (1974). Point processes and completely monotone set functions. *Z. Wahrscheinlichkeitstheorie* **31** 57–67.

LAURITZEN, S. L. (1988). *Extremal Families and Systems of Sufficient Statistics.* Springer Lecture Notes in Statistics **49**.

MARTIN-LÖF, P. (1984). Extreme point models in statistics. *Scand. J. Statist.* **11** 65–91.

MATHERON, G. (1974). *Random Sets and Integral Geometry.* Wiley, New York.

MATTHES, K., KERSTAN, J. AND MECKE, J. (1978). *Infinitely Divisible Point Processes.* Wiley, Chichester.

PEMANTLE, R. (1992). Vertex-reinforced random walk. *Prob. Theory Related Fields* **92** 117–136.

RESSEL, P. (1985). De Finetti-type theorems: an analytical approach. *Ann. Prob.* **13** 898–922.

RUELLE, D. (1978). *Thermodynamic Formalism.* Addison Wesley.

RYLL-NARDZEWSKI, C. (1957). On stationary sequences of random variables and the de Finetti's equivalence. *Colloquium Math.* **4** 149–156.

WATANABE, T. (1960). A probabilistic method in Hausdorff moment problem and Laplace-Stieltjes transform. *J. Math. Soc. Japan* **12** 192–206.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CA 94720
freedman@stat.berkeley.edu