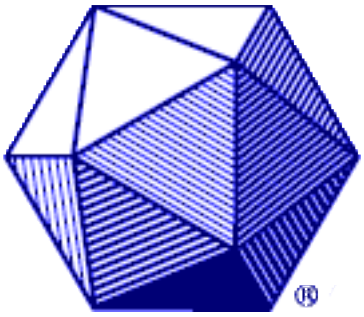

$x \to f(x - b)$ as b ranges over R^n. It is not too difficult to show (see, for example, [1], or [2]), that f is an exponential polynomial if and only if U_f is finite dimensional. So, if X_f is finite dimensional, as $U_f \subset X_f$, f is an exponential polynomial. Now if f contains any exponential $\exp(a \cdot x)$ with $a \neq 0$ in R^n, we would find that since X_f contains dilations as well as translations, X_f would be of infinite dimension. Hence $a = 0$ and so f is a polynomial.

We conclude by noting a related result of Loewner [2] that if $n > 1$ and if $f \in C(R^n)$ is such that its transforms under all isometries of R^n span a finite dimensional space, then f is a polynomial. Thus, if R_f denotes the subspace of $C(R^n)$ spanned by all translations of f and all orthogonal transformations $O_p f$: $x \to f(Px)$, where P is any orthogonal real matrix, the following holds.

THEOREM 3. *Let $n > 1$ and $f \in C(R^n)$. A necessary and sufficient condition that f be a polynomial is that R_f be finite dimensional.*

This theorem was first shown by Loewner [2] using exponential polynomials. Whether it can also be shown by methods similar to those used in the main proof of our Theorem 2 is an interesting conjecture to us.

Acknowledgements. The authors thank Dr. R. Nillsen of the University of Wollongong, and Professor J. Wilker of the University of Toronto for their helpful discussions.

References

1. P. G. Laird, On characterizations of exponential polynomials, Pacific J. Math., 80 (1979) 503–507.
2. C. Loewner, On some transformation semigroups invariant under Euclidean or non-Euclidean isometries, J. Math. Mech., 8 (1959) 393–409.

POISSON APPROXIMATIONS AND THE DEFINITION OF THE POISSON PROCESS

TIMOTHY C. BROWN
Department of Mathematics, Monash University, Clayton, Victoria 3168, Australia

1. Introduction. This article has two purposes. The first is to expose two interesting, connected facts about the Poisson distribution, both of which deserve to be better known. The first of these is that it is almost as easy to provide error bounds for certain Poisson distribution approximations as it is to derive these approximations via convergence arguments. The second one is that these bounds make it extremely easy to prove an elegant qualitative characterization of the Poisson process, due to Prékopà [**20**].

The second aim of the article is to advertise two important general ideas in probability theory. Both ideas have been extremely useful in recent research, but authors of elementary textbooks usually consider these ideas to be too advanced for inclusion. On the other hand, the exposition here only requires that the reader know some undergraduate real analysis and the most trivial facts about probability. Thus, I hope that the *simple* examples here might make the ideas more widely known.

Coupling, a technique introduced by Doeblin [**5**], is one basic idea in this article. By this we mean the study of one or two specific distributions, via the construction of two random variables (or vectors or processes). The *coupled* random variables are designed so that not only do they have

Timothy C. Brown: I did my Ph.D. at Cambridge University under G. K. Eagleson. After three years as a lecturer at the University of Bath, U.K., I returned to my native land in 1981. My chief extramathematical interests are family, religion, and music.

specified (*marginal*) distributional properties, but also it is possible to analyse their joint (*point-by-point*) behaviour. We shall use coupling in Section 2 to provide the error bounds. A (random!) sample of other areas in which coupling has proved fruitful are: the convergence of transition probabilities of Markov chains [**5**], [**19**], the renewal theorem [**15**], interacting particle systems [**23**], the F. K. G. inequality of statistical physics [**10**], simulation algorithms for spatial processes [**16**], random walks on finite groups, rapidly mixing Markov chains and the number of shuffles needed to thoroughly mix a pack of cards [**2**], and the tail behaviour of birth-and-death and stochastically monotone processes [**1**]. For an excellent survey of coupling methods for Markov processes, the reader may like to consult [**8**].

The other central idea here is that the path properties of stochastic processes are crucial in their definition. As an example, let us consider Brownian motion. In the simplest case, this is a model for the position of a particle executing a "random" motion in one dimension. In introductory courses, the model is often introduced as a stochastic process with certain joint normal distributions for the positions of the particle at different times. However, it seems initially that this definition hardly corresponds at all to the physical idea of a particles motion. On the other hand, the following model assumptions are easily interpretable:

(a) each path of the particle is continuous,
(b) the path of the particle in the time interval $[0, t]$ is independent of the path in $[t, t + s]$, apart from the common position at time t, and
(c) the mean position at time t is 0, while the variance of the position is t.

It is a consequence of a celebrated result of Lévy [**14**, p. 78] that precisely formulated versions of (a), (b), and (c) are enough to guarantee that the model *is* Brownian motion. More recently, the profound work of Stroock and Varadhan [**22**] has showed that characterizations like Lévy's can be extended to a large class of processes satisfying (a). As a result, some open questions on the existence and uniqueness of solutions of partial differential equations were settled. In Section 3 we show that it is very simple to prove the characterization of the ordinary Poisson process that is analogous to (a), (b) and (c). That proof is then easily modified to yield the much more general characterization of Prékopà [**20**].

2. Poisson Approximations. In this section, we give some results about the error in approximating the distribution of certain random variables by Poisson distributions. Let us first recall that the Poisson distribution with parameter $\lambda \geqslant 0$ attributes probability

$$P(J; \lambda) = \sum_{j \in J} e^{-\lambda} \lambda^j / j!$$

to the set $J \subseteq \{0, 1, 2, \dots\}$.

Suppose that $A_1, \dots, A_n$ are *independent* events in a probability space $(\Omega, \mathscr{F}, \mathbf{P})$. For example, in a random sample taken with replacement, A_i might represent the event that the ith person sampled has a certain disease. The random variable S, which gives the number of $A_1, \dots, A_n$ that occur, is often of interest; in the example, this would be the number in the sample who have the disease. Formally,

$$S = \sum_{i=1}^{n} IA_i \tag{1}$$

where IA_i is the indicator (or characteristic) random variable for the event A_i (that is, $IA_i(\omega) = 1$ if $\omega \in A_i$, and $IA_i(\omega) = 0$ otherwise). If the probabilities

$$p_i = \mathbf{P}(A_i), \quad i = 1, \dots, n,$$

are all the same, then S has a binomial distribution. On the other hand, if the p's vary, then the distribution is much more complicated. However, in both cases the distribution of S can be

approximated by a Poisson distribution in certain circumstances; intuitively these are when each event is "rare," i.e., each p_i is small. More precisely, we have the following theorem due to Le Cam [13].

THEOREM 1. *For any set* $J \subseteq \{0,1,2,\dots\}$,

$$|\mathbf{P}(S \in J) - P(J;\Sigma p_i)| \leqslant \Sigma p_i^2$$

where here, and for the rest of this section, unmarked sums are over $i = 1,\dots,n$.

We present a modification of a proof of Theorem 1 due to Serfling [21]. Here is the key tool:

THE COUPLING LEMMA. *If* X *and* Y *are any random variables and* J *is any (Borel) set of* $\mathbb{R}$, *then*

$$|\mathbf{P}(X \in J) - \mathbf{P}(Y \in J)| \leqslant \mathbf{P}(X \neq Y).$$

The Coupling Lemma is in itself trivial, as you can see from the proof. However, the quantities in Theorem 1 are *very* difficult to bound analytically and the power of the Coupling Lemma is that it turns this analytic problem into one of probabilistic *construction*. For, provided X is *any* random variable with the same distribution as S, and Y is *any* random variable with Poisson (Σp_i) distribution, the Coupling Lemma tells us that $\mathbf{P}(X \neq Y)$ is a bound for the left side in Theorem 1. The key point here is that the *joint* law of X and Y can be arbitrary, subject only to their having the correct marginals. As in Euclidean geometry, clever construction is just what the doctor ordered! An advantage of the proof of Theorem 1 as an introduction to coupling is that, in contrast to many other applications of coupling, the required coupling is quite straightforward.

Proof of Coupling Lemma. Without loss of generality $\mathbf{P}(X \in J) > \mathbf{P}(Y \in J)$. In this case, the required modulus is bounded by

$$\mathbf{P}(X \in J) - \mathbf{P}(X \in J \quad \text{and} \quad Y \in J) = \mathbf{P}(X \in J \quad \text{and} \quad Y \notin J),$$

and the latter is clearly bounded by $\mathbf{P}(X \neq Y)$.

The coupling we present relies on a fundamental construction in probability. This is the construction showing that a probability space which supports a uniform random variable also has one of any other distribution. It is also important for the computer simulation of random variables; usually the computer has an algorithm which produces uniform random variables and the construction can (at least in principle) then turn these into any desired distribution. We describe the construction only in the simpler discrete case and refer the interested reader to [3, p. 159] for the general case. Suppose that U is a uniform random variable over $(0,1)$ (i.e., its distribution is Lebesgue measure on $(0,1)$). Suppose also that the desired distribution Q attributes probability $q(i)$ to $\{i\}$ $(\in \{0,1,2,\dots\})$. For $x \in (0,1)$, let

$$\hat{Q}(x) = i$$

for the unique $i \in \{0,1,2,\dots\}$ such that

$$\sum_{j=0}^{i-1} q(j) < x \leqslant \sum_{j=0}^{i} q(j),$$

where the sum on the left with $i = 0$ is by convention 0. The constructed random variable is then $\hat{Q}(U)$. It is immediate that $\hat{Q}(U)$ *does* have distribution Q, because for $i \in \{0,1,2,\dots\}$,

$$\mathbf{P}(\hat{Q}(U) = i) = \mathbf{P}\left(U \in \left[\sum_{j=0}^{i-1} q(j), \quad \sum_{j=0}^{i} q(j)\right]\right).$$

Recall that we wish to construct random variables X and Y with prescribed distributions, say Q and R, so that $\mathbf{P}(X \neq Y)$ is as small as possible. The construction above gives us an obvious attempt for such similar random variables. We simply take a *single* uniform random variable U

and set $X = \hat{Q}(U)$ and $Y = \hat{R}(U)$. Unfortunately, $\mathbf{P}(X \neq Y)$ is generally still hard to calculate if we use such X and Y in Theorem 1. However, for $n = 1$ the calculation is easy. In this case, let B be the distribution of S; B is said to be a *Bernoulli* distribution with parameter $p = p_1$. Let P denote the Poisson distribution with parameter p. Then, $\mathbf{P}(\hat{B}(U) \neq \hat{P}(U)$ is the length of the dotted intervals in Fig. 1, and this is

$$(e^{-p} - (1-p)) + (1 - e^{-p}(1+p)) \leqslant p^2,$$

as required, since

$$1 - p \leqslant e^{-p}, \tag{2}$$

an inequality which also justifies the crucial ordering of the points marked in Fig. 1.

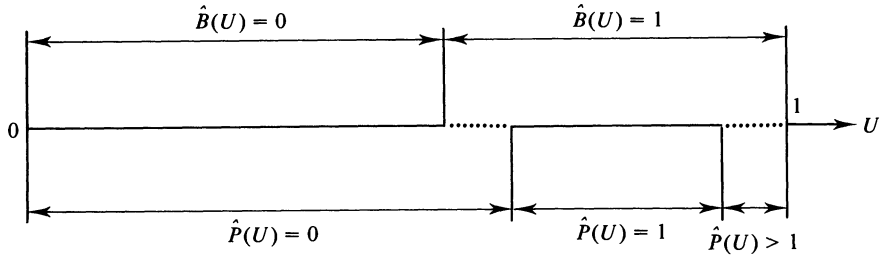

FIG. 1

Parenthetically, we remark that $(\hat{B}(U), \hat{P}(U))$ is called a *maximal coupling*, because $\mathbf{P}(\hat{B}(U) \neq \hat{P}(U))$ actually achieves equality in the Coupling Lemma for $J = \{0, 2, 3, \dots\}$. As a consequence, $p(1 - e^{-p})$ is actually the best possible bound in Theorem 1 for the case $n = 1$. While there is a method to construct a maximal coupling for *any* distributions Q and R [**12**], even for discrete distributions it is not true that $(\hat{Q}(U), \hat{R}(U))$ is always a maximal coupling. For an example, consider R equal to the Bernoulli(1/2) distribution and Q given by $Q\{0\} = Q\{1\} = \frac{1}{4}$, $Q\{2\} = \frac{1}{2}$; here $\mathbf{P}(\hat{R}(U) \neq \hat{Q}(U)) = \frac{3}{4}$, while the maximum of $|R(J) - Q(J)|$ is $\frac{1}{2}$. Nevertheless, $(\hat{Q}(U), \hat{R}(U))$ does always have a weaker maximal similarity property, namely that it maximises the *correlation* between all pairs of random variables with distributions R and Q (see, for example, [**24**], Theorem 2.5).

Proof of Theorem 1. Let $U_1, \dots, U_n$ be independent uniform random variables over $(0, 1)$. Let B_i and $P_i (i = 1, \dots, n)$ denote the Bernoulli and Poisson distributions both with parameters p_i, and denote $\hat{B}_i(U_i)$ by X_i and $\hat{P}_i(U_i)$ by Y_i. Clearly, (X_i, Y_i), $i = 1, \dots, n$, are independent random vectors. Therefore, ΣX_i has the same distribution as S. Moreover, by a standard result, ΣY_i has the $P(.; \Sigma p_i)$ distribution [**6**, p. 237]. At the same time, $\Sigma X_i \neq \Sigma Y_i$ implies $X_i \neq Y_i$ for some i in $\{1, \dots, n\}$. Combining these facts with the Coupling Lemma produces

$$\begin{aligned} |\mathbf{P}(S \in J) - P(J; \Sigma p_i)| &\leqslant \mathbf{P}\left(\bigcup_{i=1}^{n} \{X_i \neq Y_i\} \right) \\ &\leqslant \Sigma \mathbf{P}(X_i \neq Y_i). \end{aligned}$$

Applying the argument for $n = 1$ to each term in the sum completes the proof.

Often it is not convenient to have the parameter of the approximating Poisson distribution equal to Σp_i.

COROLLARY 2. *For any* $J \subseteq \{0, 1, 2, \dots\}$ *and any* $\lambda > 0$,

$$|\mathbf{P}(S \in J) - P(J; \lambda)| \leqslant |\lambda - \Sigma p_i| + \left(\max_i p_i\right) \Sigma p_i.$$

The corollary follows from the triangle inequality, Theorem 1, and the inequalities

$$\Sigma p_i^2 \leqslant \left(\max_i p_i\right)\Sigma p_i$$

and

$$|P(J;\lambda) - P(J;\Sigma p_i)| \leqslant |\lambda - \Sigma p_i|. \tag{3}$$

To see the truth of (3), we use an argument of Freedman [7]. Suppose that $\Sigma p_i > \lambda$. Let Z_1 and Z_2 be independent Poisson random variables with parameters λ and $\Sigma p_i - \lambda$. By the Coupling Lemma, the left side of (3) is bounded by $\mathbf{P}(Z_1 \neq Z_1 + Z_2)$, which in turn is bounded by

$$\mathbf{P}(Z_2 \neq 0) \leqslant 1 - e^{-(\Sigma p_i - \lambda)}.$$

The inequality (2) now gives (3). A symmetrical argument gives (3) if $\Sigma p_i < \lambda$.

To close this section we mention that the coupling technique can also produce bounds similar to Theorem 1 for dependent events and for multinomial distributions. The interested reader is referred to [**21**], [**4**], and [**18**].

3. Defining the Poisson Process. The bounds of the previous section are now applied to see that the general Poisson process arises from purely qualitative assumptions. But first we show how the ideas of that section provide a natural way to introduce the Poisson process in an elementary course (i.e., without $(\Omega, \mathscr{F}, \mathbf{P})$!).

Feller [**6**, p. 159] cites many examples where the Poisson distribution provides an adequate fit to experimental data. We shall consider the first of Feller's examples and construct a mathematical model for it. In this example, a radioactive substance emits α-particles; the number N_t of emissions reaching a given space up to time $t > 0$ is recorded. The following assumptions appear not unreasonable:

(1) that $\{N_t\}$ is a *point process on* $\mathbb{R}^+$, by which we mean that each N_t is a random variable and that a graph of N_t versus t is always an increasing step function with jumps of size 1 and $N_0 = 0$ (the assumption that the jumps in the graph are of size 1 corresponds to the idea that no two emissions occur at *exactly* the same time),

(2) that $\{N_t\}$ is *completely random*, by which we mean, informally, that information concerning the emissions in one interval of time tells us nothing about the emissions in a disjoint interval of time or, formally, that for each integer n and $0 < t_1 < \cdots < t_n$,

$$N_{t_1}, N_{t_2} - N_{t_1}, \ldots, N_{t_n} - N_{t_{n-1}}$$

are independent (lack of complete randomness corresponds to the idea that past information can tell us something about future emissions),

(3) that $\{N_t\}$ is *stationary* (*in increments*), by which we mean that $N_t - N_s$ has the same distribution as $N_v - N_u$ provided $t - s = v - u$ (of course, in the long term, this cannot be true as the radioactive substance decays, but it will be *approximately* true over intervals short compared to the half life).

THEOREM 3. *If* $\{N_t\}$ *is a stationary, completely-random point process on* $\mathbb{R}^+$, *then each* N_t *has a Poisson distribution.*

Before proving this theorem, it is instructive to compare Theorem 3 with treatments of the Poisson process in textbooks. Firstly, it seems almost universal that in the first treatment of the Poisson process, assumption (1) is replaced by a stronger and practically less interpretable assumption concerning the probability of two or more emissions in a short time interval. Complete randomness is usually assumed, but the simple assumption (3) is usually replaced by the assumption that there exists λ so that for all t,

$$\mathbf{P}(N_{t+h} - N_t = 1) = \lambda h + o(h)$$

as $h \to 0$. It initially puzzled me as a student that a model for radioactive emissions should be formulated thus; perhaps I was not alone. In any event, with these usual assumptions there are two ways to show that N_t has a Poisson distribution. Either some differential equations are formulated and solved (in which case the proof is longer than the proof of Theorem 3) or the Poisson approximation to the binomial is used (in which case the proof's structure is much the same as that below, but the stronger assumptions do not make the proof very much easier). Finally, in revising the manuscript, I discovered that Feller [**6**, p. 157] *sketches* a similar proof to the one below; it is hoped that the detailed treatment here will convince others that this is *the* natural way to introduce the Poisson process to students.

Proof. We will define for each $n = 1,2,\ldots$ a random variable S_n of the form of S in (1). It is then shown that

$$\mathbf{P}(N_t \neq S_n) \to 0 \tag{4}$$

as $n \to \infty$, and also that there exists λ such that for each m in $\{0,1,2,\ldots\}$

$$\mathbf{P}(S_n = m) \to e^{-\lambda}\lambda^m/m! \tag{5}$$

as $n \to \infty$ along a subsequence of $\mathbb{N}$. The inequality

$$|\mathbf{P}(N_t = m) - e^{-\lambda}\lambda^m/m!| \leqslant |\mathbf{P}(N_t = m) - \mathbf{P}(S_n = m)| + |\mathbf{P}(S_n = m) - e^{-\lambda}\lambda^m/m!|$$

and the Coupling Lemma then finish the proof.

To define S_n we consider the partition $\mathscr{P}_n$ of $(0, t]$ into n (left-open, right-closed) intervals of equal length. For $i = 1,\ldots,n$, we let A_{ni} be the event that the ith interval of $\mathscr{P}_n$ has *at least one* emission in it and let S_n be the number of $A_{n1},\ldots,A_{nn}$ which occur.

Let D be the random variable which gives the minimum time between emissions in $(0, t]$. If $N_t \neq S_n$, then at least one interval of $\mathscr{P}_n$ has two or more emissions in it, so that $D \leqslant 1/n$. Hence, $\mathbf{P}(N_t \neq S_n)$ is bounded above by $\mathbf{P}(D \leqslant 1/n)$. Because $\{N_t\}$ is a point process, $D > 0$. This implies that the distribution function of D is 0 at 0 and, since it is also right continuous, $\mathbf{P}(D \leqslant 1/n) \to 0$ and (4) follows.

By complete randomness, the events $A_{n1},\ldots,A_{nn}$ are independent. To establish (5), we may thus apply Corollary 2 with $S = S_n$ and $J = \{m\}$. By stationarity, $\mathbf{P}(A_{n1}) = \cdots = \mathbf{P}(A_{nn}) = p_n$, say, and thus the bound of Corollary 2 is $|np_n - \lambda| + (np_n)^2/n$. Hence, it suffices to show that $\{np_n\}$ is a bounded sequence, for, in that case, the Bolzano-Weierstrass theorem gives us a subsequence along which np_n converges.

The random variable N_t is zero if, and only if, *each* A_{ni} does *not* occur. Thus, by complete randomness and stationarity,

$$\mathbf{P}(N_t = 0) = (1 - p_n)^n. \tag{6}$$

A consequence of this is that if $\mathbf{P}(N_t = 0)$ were 0, then it would be certain that each A_{ni} would occur. But then, since $S_n \leqslant N_t$ for all n, we would have $N_t = \infty$, contradicting the assumption that $\{N_t\}$ is a point process. Thus we may take negative logs in (6) to conclude that $n\{-\ln(1 - p_n)\}$ equals the finite number $L = -\ln \mathbf{P}(N_t = 0)$. The inequality $x \leqslant -\ln(1 - x)$ $(x \leqslant 1)$ now tells us that L is an upper bound for np_n. Since clearly $np_n \geqslant 0$, the required boundness is now established.

A couple of remarks on this proof seem desirable. There are two methods to avoid the use of subsequences in the proof. One way (Feller [**5**, p. 157]) uses the fact that $-n\ln(1 - p_n)$ converges only if np_n converges. The other way would be to define $\mathscr{P}_n$ to have 2^n intervals, in which case it is easy to show that np_n increases with n. Strictly speaking, we only need the usual binomial *convergence* to Poisson in the proof and not the bound of Corollary 2; but the bound is the easiest

way to establish the convergence needed for the next theorem.

The parameter λ introduced in the proof clearly depends on t, and so we denote it as $\lambda(t)$. It is now fairly easy to show that $\lambda(t)$ is linear in t. First, the fact that N_t has a Poisson $\lambda(t)$ distribution implies that

$$\lambda(t) = -\ln \mathbf{P}(N(t) = 0) = n\lambda(t/n) \tag{7}$$

from (6). Let i and j be arbitrary integers > 0. Setting $t = 1$ and $n = j$ in (7) gives $\lambda(1) = j\lambda(1/j)$. If we set $t = i/j$ and $n = i$, the same equation gives $\lambda(i/j) = i\lambda(1/j)$, and from this and the previous equation we deduce that $\lambda(t) = t\lambda(1)$, for all rational t. To extend the latter to real t, we consider the distribution function, F, of the time, T_1, to the first emission. Since $\{T_1 > t\} = \{N_t = 0\}$, (7) gives $\lambda(t) = -\ln(1 - F(t))$ and the right continuity of F yields the required extension.

We now generalise Theorem 3 to the setting of point processes on $\mathbb{R}^p (p \geqslant 1)$ and drop the assumption of stationarity (with the requisite topological facts it is straightforward to replace $\mathbb{R}^p$ by a locally compact, second countable, Hausdorff space as in [**11**, p. 48]). Since $\mathbb{R}^p$ has a more complicated order structure, we need to describe our process differently and an easy way to do this is by specifying an appropriate sample space, Ω.

The samples we are trying to model are random sets of points in $\mathbb{R}^p$ (in the previous example, each point is an emission). Two sensible restrictions are that each set ω is countable and that each ω has finite intersection with every bounded set. Accordingly, we take Ω to be the set of all such ω and for any bounded set B define $N(B, \omega)$ to be the cardinality of $\omega \cap B$ (in the previous example $N_t(\omega) = N((0, t], \omega)$). For B a bounded Borel set, we want $N(B) : \omega \to N(B, \omega)$ to be a random variable (so that $\{N_t\}$ is a point process on $\mathbb{R}^+$ in the example). Accordingly, we call a probability measure $\mathbf{P}$ *a point process on* $\mathbb{R}^p$ if $\mathbf{P}$ has domain a σ-algebra $\mathcal{F}$ on Ω, which is large enough so that for any $B \subseteq \mathbb{R}^p$, $N(B)$ is a random variable (hereafter, all subsets of $\mathbb{R}^p$ are assumed bounded Borel). Informally, $\mathbf{P}$ is the law that dictates how to choose the points of the process. The point process is *completely random* if $N(B_1), \ldots, N(B_n)$ are independent whenever $B_1, \ldots, B_n$ are disjoint subsets of $\mathbb{R}^p$ (that, in the previous example, this definition is no stronger than the one given is a standard measure theoretic exercise). The point process is called a *Poisson process* if each $N(B)$ has a Poisson distribution. In this case, since for $x \in \mathbb{R}^p$, $N\{x\}$ only takes the values 0 or 1, the parameter of the distribution of $N\{x\}$ must always be 0. Thus, a Poisson process has *no fixed atoms*, meaning that for each $x \in \mathbb{R}^p$, $\mathbf{P}(N\{x\} = 0) = 1$. This qualitative property appears quite reasonable in many applications; an *exact* time point at which there was a positive probability of observing a radioactive emission would be very special indeed. The remarkable fact along these lines is due essentially to Prékopà [**20**].

THEOREM 4. *A completely random point process is a Poisson process if, and only if, it has no fixed atoms.*

A proof of Theorem 4 is presented by detailing some necessary changes to that of Theorem 3; the result seems much simpler than existing proofs of Theorem 4. Since we now have defined an underlying probability space, we also show how to eliminate the nonmathematical language in the previous proof.

Proof. We have already noted that it is necessary that $\mathbf{P}$ has no fixed atoms.

We fix a subset B of $\mathbb{R}^p$ and replace every occurrence of N_t in the previous proof by $N(B)$. The major change is in the definition of the partition $\mathcal{P}_n$ which will no longer have n sets in it, nor will these sets be intervals of equal size. Let $x \in \overline{B}$ and let $D_x(\omega)$ be the minimum of $|y - x|$ for the finitely many y in $\omega \cap \overline{B}$. The assumption of no fixed atoms means that each D_x has a distribution function which is 0 at 0. Using right continuity again, we may thus choose a strictly positive $d_x \leqslant 1/n$ such that $\mathbf{P}(D_x \leqslant d_x) \leqslant 1/n$. Thus, if C_x is the open ball with centre x and radius d_x,

$$\mathbf{P}(N(C_x) \geqslant 1) \leqslant 1/n.$$

The Heine-Borel Theorem then allows us to choose a finite subcover of $\bar{B}$ from the sets $\{C_x\}_{x \in \bar{B}}$. By taking suitable intersections [**9**, p. 32], we can use this subcover to form a partition $\mathscr{P}_n = \{B_{n1}, \ldots, B_{nk}\}$ of B (where of course k actually depends on n). We now let $A_{ni} = \{N(B_{ni}) \geqslant 1\}$ $(i = 1, \ldots, k)$ and S_n be S in (1) except that now and hereafter the sum is over $i \in \{1, \ldots, k\}$.

To demonstrate (4), we proceed as before after making the formal definition that $D(\omega)$ is the minimum of $|y - z|$ such that y and z are both in $\omega \cap B$, and noting that $N(B, \omega) \neq S_n(\omega)$ implies there exist $\omega_i \neq \omega_j \in \omega$ such that ω_i and ω_j both lie in the same set of $\mathscr{P}_n$. To demonstrate (5), we note that even in the absence of stationarity, it still suffices to show that $\{\Sigma p_{ni}\}_n$ is bounded. This is because each B_i in $\mathscr{P}_n$ is contained in some C_x and so the maximum of $p_{n1}, \ldots, p_{nk}$ is bounded above by $1/n$. Finally, we note that $\mathbf{P}(N(B) = 0)$ now becomes the product of $(1 - p_{n1}), \ldots, (1 - p_{nk})$ and by the last statement this exceeds $(1 - 1/n)^n > 0$ for $n \geqslant 2$. Thus, we still get the finite number $-\ln \mathbf{P}(N(B) = 0)$ as an upper bound for Σp_{ni}.

4. Acknowledgement. The author is grateful to J. G. Kupka for many valuable conversations on this material.

References

1. D. J. Aldous, Tail behaviour of birth-and-death and stochastically monotone processes, Z. Wahrsch. Verw. Gebiete, 62 (1983) 361–374.

2. D. J. Aldous, Random walks on finite groups and rapidly mixing Markov chains, Séminaire de Probabilités XVII, Lecture Notes in Mathematics, 986, Springer, 1983.

3. P. Billingsley, Probability and Measure, Wiley, New York, 1979.

4. T. C. Brown, Some Poisson approximations, Ann. Probab., 11 (1983) 726–744.

5. W. Doeblin, Exposé de la théorie des chaînes simples constantes de Markov à un nombre fini d'états, Rev. Math. Union Interbalkanique, 2 (1938) 77–105.

6. W. Feller, An Introduction to Probability Theory and Its Applications, vol. 1, Wiley, New York, 1968.

7. D. Freedman, The Poisson approximation for dependent events, Ann. Probab., 2 (1974) 256–269.

8. D. Griffeath, Coupling methods for Markov processes, in Studies in Probability and Ergodic Theory (Editor G. Rota), Academic Press, 1978.

9. P. Halmos, Measure Theory, Springer-Verlag, New York, 1974.

10. R. A. Holley, Remarks on the FKG inequalities, Comm. Math. Phys., 34 (1974) 227–231.

11. O. Kallenberg, Random Measures, Academic Press, London, 1976.

12. T. Kaijser, On a new contraction condition for random systems with complete connections, Rev. Roumaine Math. Pures Appl., 26 (1981) 1075–1117.

13. L. Le Cam, An approximation theorem for the Poisson binomial distribution, Pacific J. Math., 10 (1960) 1181–1197.

14. P. Lévy, Processus Stochastiques et Mouvement Brownien, Gauthier-Villars, Paris, 1948.

15. T. Lindvall, A probabilistic proof of Blackwell's renewal theorem, Ann. Probab., 5 (1977) 482–485.

16. H. W. Lotwick and B. W. Silverman, Convergence of spatial birth and death processes, Math. Proc. Cambridge Philos. Soc., 90 (1981) 155–165.

17. K. Matthes, J. Kerstan and J. Mecke, Infinitely Divisible Point Processes, Wiley, Chichester, 1978.

18. D. R. McDonald, On the Poisson approximation to the multinomial distribution, Canad. J. Statist., 8 (1980) 116–118.

19. J. W. Pitman, Uniform rates of convergence for Markov chain transition probabilities, Z. Wahrsch. Verw. Gebiete, 29 (1974) 193–227.

20. A Prékopa, On secondary processes generated by a random point distribution of Poisson type, Ann. Univ. Sci. Budapest. Eötvös Sect. Math., 1 (1958) 153–170.

21. R. J. Serfling, A general Poisson approximation theorem, Ann. Probab., 3 (1975) 726–731.

22. D. W. Stroock and S. R. S. Varadhan, Multidimensional Diffusion Processes, Springer-Verlag, New York, 1979.

23. L. N. Vasershtein, Markov processes on countable produce spaces describing large systems of automato, Problemy Peredači Informacii, 3 (1969) 64–72.

24. W. Whitt, Bivariate distributions with given marginals, Ann. Statist., 4 (1976) 1280–1289.