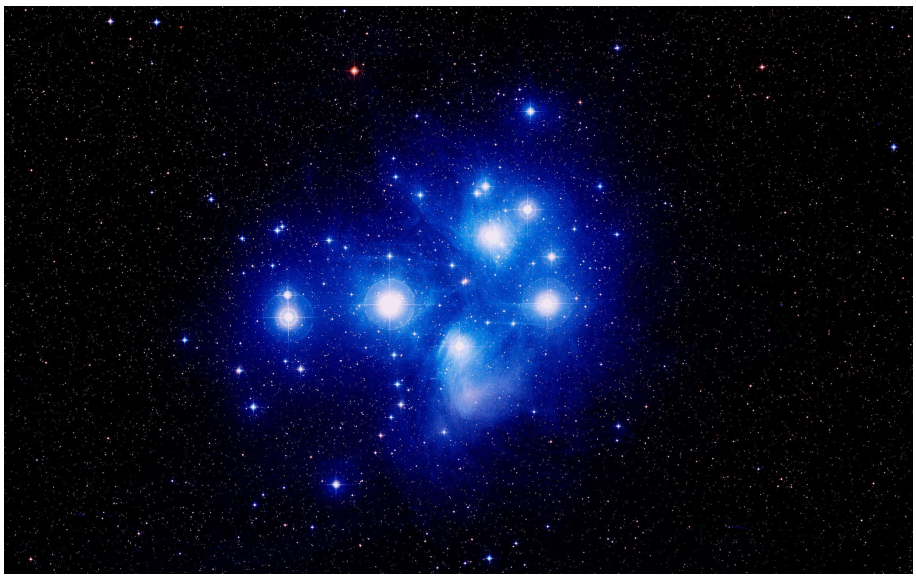
Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Note

What happened to discrete chaos, the Quenouille process, and the sharp Markov property? Some history of stochastic point processes



Note no SAMBA/16/12
Authors Peter Guttorp
Thordis L. Thorarinsdottir
Date 29th February 2012

Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Norwegian Computing Center

Norsk Regnesentral (Norwegian Computing Center, NR) is a private, independent, non-profit foundation established in 1952. NR carries out contract research and development projects in information and communication technology and applied statistical-mathematical modelling. The clients include a broad range of industrial, commercial and public service organisations in the national as well as the international market. Our scientific and technical capabilities are further developed in co-operation with The Research Council of Norway and key customers. The results of our projects may take the form of reports, software, prototypes, and short courses. A proof of the confidence and appreciation our clients have in us is given by the fact that most of our new contracts are signed with previous customers.

Norsk Regnesentral
Norwegian Computing Center
Postboks 114, Blindern
NO-0314 Oslo, Norway

Besøksadresse
Office address
Gaustadalléen 23
NO-0373 Oslo, Norway

Telefon · telephone
(+47) 22 85 25 00
Telefaks · telefax
(+47) 22 69 76 60

Internett · internet
www.nr.no
E-post · e-mail
nr@nr.no

Title	**What happened to discrete chaos, the Quenouille process, and the sharp Markov property? Some history of stochastic point processes**
Authors	**Peter Guttorp , Thordis L. Thorarinsdottir**
Date	29th February 2012
Publication number	SAMBA/16/12

Abstract

The use of properties of a Poisson process to study the randomness of stars is traced back to a 1767 paper. The process was used and rediscovered many times, and we mention some of the early scientific areas. The name Poisson process was first used in print in 1940, and we believe the term was coined in the corridors of Stockholm University some time between 1936 and 1939. We follow the early developments of doubly stochastic processes and cluster processes, and describe different efforts to apply the Markov property to point processes.

Keywords	cluster point process; doubly stochastic Poisson process; Markov point process; Poisson process
Target group	
Availability	Open
Project	SFI P19 Greenhouse
Project number	220314
Research field	
Number of pages	23

1 Introduction

Several authors have written about the history of point processes. For example, Daley and Vere-Jones (2008) devote a chapter to it, in which they distinguish between two strands of history: life tables and counting problems. Cox and Isham (1980) provide a variety of historical remarks in their Bibliographic notes at the end of each chapter. Peter Diggle, in Chapter 1 of Gelfand et al. (2010), gives a brief historical outline of spatial point processes, in the context of general spatial models. Kallenberg (1983) makes some historical remarks at the end of his book. Lewis (1972) laments the variety of nomenclature used in the point process literature. Things have not really improved since then.

In this paper we will go into a little more detail about some of the original contributions, and add new aspects of the history of point processes. We first discuss the Poisson process, tracing some of the early uses of it and searching for who first named it. The history of cluster processes goes back to 1939, while doubly stochastic processes can be traced back at least to 1938. Markov point processes are nowadays almost exclusively used in the sense of Ripley and Kelly, but we also describe some earlier attempt to extend the Markov property to the plane and their consequences for point process theory. For length reasons, we leave out the important topics of multivariate and marked processes.

2 The Poisson process

The Poisson process is characterized by the distribution of the number of points in a bounded Borel set B being Poisson with mean $\Lambda(B)$, where Λ is a non-atomic finite measure (Rényi, 1967). It follows that the number of points in disjoint sets are independent. In fact, it suffices to look at the zero probability function, i.e., the probability of no points in the set B (Daley and Vere-Jones, 2008). A crucial property of the homogeneous Poisson process, having $\Lambda(B) = \lambda|B|$, is that given the number of points in a set, they are uniformly distributed over the set. This is the property most often used in the earliest applications of this process. For Poisson's contribution to the distribution, see Stigler (1982), and for a historical overview, see Chapter 9 of Haight (1967).

2.1 Some astronomy

The earliest use of a Poisson process that we are aware of is a paper by Michell (1767), over a decade before Poisson was born (see Todhunter (1965), pp. 332-335). The paper examines

> *what it is probable would have been the least apparent distance of any two or more stars, any where in the whole heavens, upon the supposition that they had been scattered by mere chance, as it might happen.*

More precisely, what is the probability that any particular star should happen to be within one degree of another star? The probability is the ratio of a square degree to the surface

area of the sphere, or 1/13131.25. Subtracting this number from 1, and raising it to the power n of the number of stars of not less brightness than those in question, gives the probability that none of the n stars is within one degree of the given star.

Michell takes as an example the six brightest stars in the Pleiades, there being about 1500 stars "equal in splendor to the faintest of these", i.e. of similar or larger apparent brightness. Starting from the star Maya, the five other stars (Taygeta, Electra, Merope, Alcyone and Atlas), are respectively at distances of 11, 19.5,24.5, 27 and 49 arc minutes from Maia. He then computes the probabilities that no star out of 1500 scattered at random would be within these distances of Maia. These are .998173, .988018, .982506, .977148 and .926766, respectively. Subtracting each of these from 1 yields the respective probabilities that "such stars would be found within the distances above specified from the star Maya", and he now multiplies these together, "to determine the probability that these events should all take place at the same time", and finally multiplies by 1500 "to find the probability, that it should not have happened any where in the whole heavens" (he probably means the probability that it should have happened somewhere), which is about 1 in 496 000.

Multiplying the five probabilities together is an error, as the spheres are not disjoint. If one instead calculates the probabilities of one star in each of the spherical shells, and otherwise proceeds as did Michell, the answer becomes 1 in 14 509 269. Michell's conclusion is that stars are not scattered by mere chance, but rather that there is some cause for this. Newcomb (1860) points out that Michell's resulting probability was too high, but takes a different approach to calculating the probability, "to be used in so many more problems". He derives the Poisson distribution as the binomial approximation for large n and small p (there is no mention of any earlier work), shows that it is a distribution, and applies it to calculating the probability of seeing six stars in a square degree, which he finds to be 1 in 7 790 638. Perhaps we would be inclined to compute the probability of finding at least six stars in a square degree, which in this case is 1 in 7 831 614. Stigler (1982) considers this the first application of the Poisson distribution to data. Pólya (1919) calculates the average nearest neighbor distance for uniformly scattered stars, with no reference to any earlier calculations.

Seidel (1876) posed the following elementary probability problem:

> *An event A can occur in an unlimited, or in a very large and not known, number of instances which are mutually independent and all equally possible; we denote the probabilities (assumed finite and not close to one) that it does not occur, occur precisely once, precisely twice, etc., by y_0, y_1, y_2... One of these numbers are given; how do we determine the other?*

As an example, Seidel mentions the number of comets with luminosity exceeding a minimum value that are seen in a year. He computes the Poisson approximation to the binomial for $p = \lambda/N$, and notes that if the events occur in a given time interval or set, the probability corresponding to a three-fold increase in the time interval or set will be of the same form, but with a λ which is three times larger. Given that on average over several years there are 6.9 comet sightings per year, then the case where we observe 6 sightings

is more likely than observing 7 by a factor of 7:6.9.

Seidel only refers to Bernoulli's work, and not in detail. However, it would not be surprising if he was aware of the work of Clausius described in the next section. They both entered University of Berlin in 1840 as students of Dirichlet (O'Connor and Robertson, 2000a,b).

2.2 The kinetic theory of heat

At the beginning of the 19th century, heat was considered a weight-less substance, caloric, which permeated different pieces of matter in differing degrees depending on their temperature. Among the eminent scientists of the day espousing this point of view were Carnot, Laplace and Poisson. In the middle of the nineteenth century a different idea started to emerge, espoused above all by Krönig, Clausius, Joule and Maxwell. Instead of assuming the presence of a caloric, they advocated a kinetic theory, saying that heat only was movement of molecules, and consequent collisions with other molecules or the sides of the container. This explained the gas laws, relating temperature, pressure and volume, in a very simple fashion, and had actually been suggested by Daniel Bernoulli a century earlier (Brush (2003) is the source for the description in this paragraph).

The simplest kinetic theory assumed that molecules move with constant velocities and direction until they collide with another molecule or the side of the vessel containing the substance. This was criticized (e.g., by Buijs-Bullet (1858)) on the grounds that this theory would imply that two gases would mix rapidly, in contrast to observations. For example, when food is brought into a room, if molecules are moving rapidly in straight lines, the smell ought to reach each part of the room nearly simultaneously, which it of course does not. In order to explain this phenomenon, Clausius (1858) set out to calculate the mean distance between collisions. If a molecule tends to collide fairly quickly with another molecule, instead of moving a long distance in a straight line, the criticism would be rendered invalid.

His assumptions are set out in an earlier paper about the kinetic theory (Clausius, 1857). They include (a) the space actually filled with molecules of gas is infinitesimal compared with the whole space occupied by the gas; (b) the duration of an impact is infinitesimal compared to the interval between two successive collisions; and (c) the influence of the molecular forces is infinitesimal.

The third assumption can be interpreted to require that the portion of the path for which molecular forces are changing the motion of a molecule must be infinitesimal compared to the portion where such forces can be considered inactive. However, since the theory is based upon molecular forces being active only at very small distances, a path which is long in comparison to the sphere of influence of a molecule may yet, in absolute terms, be very small. Clausius now attempts to compute "How far on average can the molecule move, before its center of gravity comes into the sphere of influence of another molecule." If this average distance is small, most molecules will not be able to travel very far before they collide with another or at least have their path bend, and consequently gases will

not mix very rapidly.

In order to compute this *mean free path* of a molecule, Clausius first argues that the mean free path among moving molecules is no greater than among stationary ones, and so proceeds to discuss the case where only one molecule moves and all other are stationary (the mean free length is then reduced by 3/4 compared to the case where all the molecules are moving with the same velocity). He now makes the crucial assumption that

> *there is a space containing a great number of molecules, and that these are not regularly arranged, the only condition being that the density is the same throughout, i.e., in equal parts of the space there are the same number of molecules.*

Later text seems to indicate that what Clausius means is really that the *average* numbers of molecules are the same in equal parts of the space. The space is now divided into parallel layers perpendicular to the direction of movement of the single molecule, and the probability that the molecule freely passes through a layer of thickness x is a^x, where a is the probability of free passage through a layer of unit thickness. This, of course, subsumes independence between occurrences in different layers. He sets $\alpha = -\log a$ to get the probability $e^{-\alpha x}$, the zero probability of a Poisson process, probably the first derivation of this quantity, which will be rediscovered repeatedly.

The reference to this work is given in several places, such as Cox and Isham (1980) and Gelfand et al. (2010). In Daley and Vere-Jones (2008) a reference is instead given to Boltzmann (1868), a reference originating from Haight (1967), which however appears to be incorrect. The Boltzmann paper derives the asymptotic distribution of the maximum of uniform random variables.

2.3 The distribution of blood cells over a microscope slide

Abbe (1879) describes a new instrument, designed by C. Zeiss, to count the number of blood cells in a sample. He is interested in the reliability of the instrument. Abbe is studying the variability about the mean number n of blood cells in a defined region, and uses the Poisson distribution to describe it. He says

> *this rule can similarly be used whether the situation is a random spatial distribution of counts, as in this case, or the counts are distributed in time. ... Its validity is based on the assumption that n is small compared to the square root of the number corresponding to the largest possible value in the set of possible numbers, and also, if it concerns spatial counts, the mean number in the volume is only a small fraction of the square root of the number that would completely fill this volume.*

Abbe does not mention the independence between counts in disjoint sets, and does not give an argument for the Poisson distribution of counts, although he determines that it is a proper distribution.

Student (W. S. Gosset) (1907) studies a similar problem (yeast cells rather than blood cells), and, after deriving the marginal Poisson distribution of counts in a square, remarks "In applying this to actual cases it must be noted that we have not taken into account any

"interference" between the particles; there has been supposed the same chance of a particle falling on an area which already has several particles as on one altogether unoccupied. Clearly if m [the average number of cells per square] be large this will not be the case, but with the dilutions usually employed this is not of any importance." He then proceeds to estimate the correlation between each square and its four nearest neighbors, which in his material is very close to 0.

2.4 Insurance mathematics and approximations

In a 1903 dissertation at the University of Uppsala in Sweden, Filip Lundberg (Lundberg, 1903) lays the foundations to what would later become the Stockholm school of insurance mathematics. He introduces the concept of a continuous risk, characterized by being a sum of independent partial risks, or risk elements. Each of the risk elements is associated with a risk sum a_i and a risk coefficient $p_i{}'$ in such a fashion that during a time element dt there are two possibilities, (a) the risk sum a_i comes due with probability $dp_i = p_i{}'$; or (b) the risk sum a_i does not come due with probability $1 - dp_i$.

Lundberg is interested in the distribution of the total risk $x = \sum_i a_i m_i$ where the m_i are non-negative integers. Writing $f(x)$ for the density of x, he derives a variational equation

$$\delta f(x) = \int (f(x-a) - f(x))\delta p(a) da \tag{1}$$

where $\delta p(a) = p_i$ when $a = a_i$. In the special case where all the a_i are equal to a unit value, writing $P = \int p(a)da = \sum a_i p_i$, we see that

$$\frac{\partial f(x,P)}{\partial P} = f(x-1,P) - f(x,P) \tag{2}$$

This is the forward equation for the Poisson process, later derived (as a special case) by Kolmogorov (1931) in his fundamental paper on Markov processes. When the a_i are different, the resulting process is a compound Poisson process (Cramér (1969) describes Lundberg's work and puts it in the context of modern insurance mathematics).

2.5 Telephones, earthquakes, and radioactive decay

In 1909, the Danish telephone engineer A. K. Erlang published some probability calculations (Erlang, 1909) relating to telephone processes. In the paper he develops the Poisson distribution for the number of incoming calls to an exchange in a given period of time, discusses briefly the possibility of temporally non-homogeneous processes, and outlines the first example of a $M/M/1$ queue. The paper does not give many references, but Erlang describes how the Director of the Copenhagen Telephone Company, Mr. F. Johannsen, has been using probability methods in solving many practical problems, and set others to work on similar investigations.

Whitworth (1886) also calculates the zero probability function for a Poisson process on the positive half line, and mentions an application to earthquakes. He clearly considers this his own invention. von Schweider (1905) quotes the "well-known free path distribution" and deduces the standard error of the average alpha particle count in a fixed time interval

based on the Poisson distribution. Bateman, in his Note following Rutherford and Geiger (1910), refers to both Whitworth's and von Schweider's results, and derives the forward equations for the transition probabilities in the same fashion as Lundberg (1903). The paper by Rutherford and Geiger deals with the counts of alpha particles in 1/8 minute intervals.

2.6 Where did the name come from?

From a mathematical point of view, the Poisson process was invented by Filip Lundberg (Haight (1967) only indirectly mentions Lundberg, and mistakenly ascribes the derivation of the forward equations to Charlier (1905), who derives the Poisson distribution in his B-series approximation, but does not deal with temporal or spatial developments). But who came up with the name? Lundberg's dissertation (as well as his subsequent work) does not name the process. Like de Finetti (1929), Kolmogorov (1931) and Lévy (1939), he just notes that the marginal distribution is Poisson. Fry (1928), in his excellent book on engineering uses of probability, uses the term "Poisson law" in two senses: as the probability distribution, and as the process, but does not seem to need to distinguish the senses. Khinchin (1933) also develops the Poisson process, which he calls "the elementary discontinuous process", while Doob (1937) calls it "the time series with Poisson distribution". Wiener (1938) associates a stochastic process in his harmonic analysis framework with the name Poisson, when he develops the "discrete or Poisson chaos", establishing that a linear resonator set in motion by Poisson chaos has the same spectrum as the response to a single impulse.

On the other hand, Feller (1940) uses the name Poisson process as if it were well-known, and Ove Lundberg's dissertation the same year (Lundberg, 1940) does the same. The name may have been in common use in Stockholm (where Feller worked 1934-1939 and Lundberg was a PhD student), but not until after Cramér finished his book (Cramér, 1937) in 1936, since he also just notes the marginal Poisson distribution. In later editions he uses the term Poisson process. Also, in his lecture (Feller, 1936) at the 1936 International Congress of Mathematicians in Oslo, Feller describes but does not name the Poisson process. Grimmet and Stirzaker (2001) say that Feller came up with the name "before 1940" which is consistent with our findings. Since neither Kolmogorov, Khintchin, nor Doob use that nomenclature, it must not have been standard in the Soviet Union or in the United States. In the British literature we have not found the terminology before 1947 in a book review by Kendall, while Lévy uses the term in a 1948 letter. It is likely that the term was adopted by most participants at the first major meeting after World War II, the Berkeley Symposium on Probability and Mathematical Statistics in 1945, where it was used in lectures by Doob and Feller. The term point process originated with Connie Palm in his 1943 dissertation (Palm, 1943). Eggenberger and Pólya (1923) may be the first authors to use the spatial term point patterns, in discussing "die statistische Verteilung von Punktgesamtheiten im Raume."

3 Cluster processes

A general cluster process is constructed from a primary process of cluster centers, each point of which generates a secondary process of cluster points. The resulting superposition of points (which may or may not include the points from the primary process) is a cluster process. Usual simplifying assumptions include the primary process being a Poisson process (yielding a Poisson cluster process), the secondary processes being conditionally independent of each other (yielding an independent cluster process) in which case a general formula for the probability generating functional can be derived (Daley and Vere-Jones, 2008, eq. 6.3.6). In what follows we shall, unless otherwise specified, assume that the cluster processes considered are independent cluster processes.

3.1 Neyman-Scott Poisson cluster processes

Neyman (1939) describes a stochastic model for the location of potato beetle larvae on a field. The structure is based on the beetle laying masses of eggs in random locations. When the eggs hatch, the larvae can move independently away from the egg mass. In this paper, Neyman mainly uses the structure to develop a family of over-dispersed (compared to the Poisson distribution) discrete distributions, so-called contagious distributions, these being the marginal distributions of the numbers of eggs in a given part of the field. Since it is a marginal distribution, it depends on the dispersion (larva movement) only through an integral. The distribution of egg mass numbers is taken to be Poisson or binomial in the paper (although a general formula is developed for the characteristic function of the number of larvae in a given part of the field). It would be straightforward to develop the joint distribution of numbers of larvae in disjoint parts of the field by replacing a binomial distribution by a multinomial in the calculation, but Neyman does not do that.

The particular case where the egg mass sizes are Poisson distributed is studied by Thomas (1949), without allowing any movement of larvae. It was not until Neyman and Scott applied the structure to clusters of galaxies that the actual location of the points, once they are moved away from the cluster centers, were considered (Neyman and Scott, 1952). A variety of researchers (e.g. Bartlett (1954); Skellam (1952); Thompson (1954, 1955)) have worked out details about special cases. Frequently the Neyman-Scott approach is cited using Neyman and Scott (1958), but this is not where the process was introduced, not even in the context of galaxy models. The apparent clustering of galaxies is discussed by many astronomers, such as Bok (1934), while the idea that galaxies are clustered in reality, not just apparently, is suggested by Charlier (1922).

The independent and identically distributed distance of the secondary points from their cluster center characterizes the Neyman-Scott cluster process structure. The name Thomas process has been used in the literature for a Neyman-Scott Poisson cluster process with Poisson cluster sizes and bivariate normal dispersion. As pointed out by Diggle et al. (1976), this is a generalization of the distribution derived by Thomas, and they propose the name generalized Thomas process. This Neyman-Scott Poisson cluster process (re-

gardless of the dispersion distribution) is also a particular doubly stochastic Poisson process, known as a shot noise process (Exercise 6.3.8 in Daley and Vere-Jones (2005)), and it seems hardly necessary to give the process with normal dispersion a separate name.

3.2 Other dispersion or cluster size mechanisms

Bartlett (1963) in the context of traffic and Lewis (1964) in the context of computer failures independently develop a Poisson cluster process in which the secondary processes are renewal processes, now called a Bartlett-Lewis process. A straightforward generalization to more than one dimension is to have the secondary processes be random walks.

There are no examples in the literature of secondary processes that are not independent (Laslett (1978) mentions the possibility), although it is not difficult to think of some. For example, in the linear case one could let the cluster sizes be determined by the value of a birth-death-immigration process at the times of the primary points.

3.3 Non-Poisson cluster processes

There is really no particular reason (except mathematical convenience) to use a Poisson process for the cluster centers. Neyman and Scott (1952) use a "quasi-uniform" distribution of cluster centers, which could be a Poisson process or a doubly stochastic Poisson process. Neyman and Scott (1958) outline a hierarchy of kth order cluster processes, having a $k-1$th order cluster process as its primary process. Apparently clusters of galaxies need at least a second order cluster process.

As discussed, e.g. by Le Cam (1960) and Kavvas and Delleur (1981), in modeling precipitation it would be natural to have cluster centers correspond to front systems, which, since they are physically separated, would need an inhibitory mechanism (see e.g. Istok and Boersma (1989)). A physical description of stratiform precipitation can be found in Hobbs and Locatelli (1978).

Similarly, in modeling forests (such problems go at least back to Pólya (1918), describing the longest line of visibility in a random forest), there is an apparent tendency for large trees, even in virgin forests, to be less dispersed than what would be predicted by a Poisson process. On the other hand, processes of seed dispersal can create clusters of smaller trees (e.g., Guttorp (1991), Rathbun and Cressie (1994) and Jalilian et al. (2011)).

4 Doubly stochastic processes

Le Cam (1947) uses the characteristic functional (originating with Kolmogorov (1935)) to analyze a precipitation model which is a Poisson process where the rate itself is a random process. Such processes are called doubly stochastic Poisson processes. The name was introduced by Bartlett (1963). Other names for the same process are the Quenouille process (Brillinger, pers. comm.) after a spatial sampling paper (Quenouille, 1949) that introduces the doubly stochastic mechanism, and the Cox process after Cox (1955), where a piecewise constant rate function was suggested as a model for stops of a loom due to

weft break. The latter name apparently was coined by Krickeberg (1972).

Commonly used particular cases of doubly stochastic Poisson processes are ones where the rate process is $\exp Z_t$ where Z_t is a Gaussian process (called a log Gaussian Cox process by the same misnomer as for the lognormal distribution) and one where the rate is a shot noise process. The later is a linear filter of a Poisson process, $X_t = \int_{-\infty}^{t} w(t-u)dN(u)$, and was first considered by Campbell (1909). It is interesting that a doubly stochastic Poisson process where the rate is a Poisson driven shot noise process is equivalent to a Neyman-Scott Poisson cluster process with Poisson distributed cluster size distribution and dispersion density proportional to w.

The Gauss-Poisson process, introduced by Newman (1970), is defined through its characteristic functional, and determined by its first two factorial moments. It was proposed as a generalization of a Poisson process that maintains invariance under superposition. Daley and Vere-Jones (2005, Chapter 6) point out that this process is both a doubly stochastic Poisson process and a Poisson cluster process of Bartlett-Lewis type. It also shares with the Poisson process the fact that the property of first order stationarity is equivalent to that of full stationarity (Milne and Westcott, 1972).

The first general characterization of doubly stochastic Poisson processes was given by Mecke (1968) who defines the process through a mixture of Poisson process,

$$X = \int X_\lambda Q(d\lambda),$$

where X_λ is a Poisson process with intensity λ and Q is a probability measure on a suitable σ-algebra of subsets of the set of all Radon measures λ. Let D_q denote an operator that independently selects the occurrences of a point process X with probability q, for example $D_q X_\lambda = X_{q\lambda}$, and let Π denote the set of all point processes such that $D_q\Pi = \{D_q X : X \in \Pi\}$. Mecke then shows that the set

$$\bigcap_{0<q<1} D_q\Pi$$

is the set of all doubly stochastic Poisson processes.

Matérn (1971) poses the question

> *...one may ask if it is possible to construct a doubly stochastic Poisson process with a nontrivial variable density function in such a way that the numbers $N(B)$ will be distributed in the negative binomial distribution.*

In his reply to the discussion of the same paper, Matérn notes that the number of points in the bounded Borel set B, $N(B)$, will have a negative binomial distribution if and only if the random intensity measure has a gamma distribution. More generally, Diggle and Milne (1983) investigate whether there exist spatial point processes with negative binomial counts that are stationary, ergodic, and orderly. They consider two classes of processes: doubly stochastic processes with a gamma distributed random intensity measure and cluster processes with Poisson cluster centers, cluster sizes following a logarithmic

distribution, and a degenerate dispersion process that places all the offsprings at the cluster center. Under the assumption that the gamma measure is a completely random measure, Diggle and Milne (1983) show that a doubly stochastic process of the first type is necessarily also in their second class of cluster processes.

We are not aware of any further examples of doubly stochastic non-Poisson processes, although the processes determined by its conditional intensity (see section 6) can be thought of that way.

5 Markovian processes

Three types of Markov properties for point processes appeared in the literature over a period of three decades, from 1948 to 1977. The earliest approaches concern univariate processes in time while each type has also been considered for spatial or, more generally, multivariate processes. Interestingly, the three properties define three disjoint model classes.

5.1 Markov dependent interpoint intervals

It was Wold (1948, 1949) who first investigated a Markov property for point process models through his work on processes with interpoint intervals that form a Markov chain. He thereby extends the arsenal of available models past the Poisson process and renewal processes both of which appear as limiting cases in his framework. Let $\{W_j\}$ denote the process of interpoint intervals. Under Wold's model, the intervals take the form of an m-dependent Markov sequence such that the distribution of W_{j+m} given $\{W_i : i \leq j+m-1\}$ does not depend on $\{W_j : i < j\}$. These processes are now called Wold processes. When $\{W_j\}$ is assumed to be stationary with $m = 1$, a marginal interval density $f(\cdot)$ for W_j and a conditional interval density $f(\cdot|w)$ for W_j given $W_{j-1} = w$ define a Wold process if

$$f(v) = \int_0^\infty f(v|w)f(w)dw.$$

Wold analyzes this model in the case where the conditional density is an exponential density with rate $\lambda(w) = aw^{-1/2}$ and Cox (1955) considers $\lambda(w) = a_0 + a_1 w$.

Lampard (1968) discusses a first order Wold process which is more tractable than Wold's and Cox's models, and notes that it was being used in neurophysiological experiments. Lampard's process is a counter system with two counters, A and B, which input is a pair of independent Poisson processes, X_{up} and X_{down}, with different mean rates. At time t_j, counter A has $N_j + r + 1$ counts in it and is switched to receive pulses from X_{down} while counter B has zero counts and is switched to receive pulses from X_{up}. The situation continues until counter A reaches zero at which time t_{j+1}, an additional count of $r + 1$ is added instantaneously to counter B for it to have $N_{j+1} + r + 1$ counts in it. The roles are then switched. Most statistical properties of this process can be obtained in closed form (Lampard, 1968; Lee and Ong, 1986). However, the general intractability of Wold processes is a limitation of the model framework. The three models mentioned

here, and their statistical properties, are discussed in detail in Lawrence (1972), a more recent overview and further examples are given in Daley and Vere-Jones (2005).

Several other models have been proposed that are strongly related to Wold processes. The semi-Markov process is a model with k distinct distribution functions $F_1, \ldots, F_k$ from which the interpoint intervals may be drawn where the current distribution function is selected according to a Markov chain. This process was proposed independently by Lévy (1954a,b), Smith (1955) and Takács (1954). Its properties are discussed in Cox (1963), Cox and Lewis (1966), Cox and Isham (1980), and Lawrence (1972). The semi-Markov process has e.g. been applied to model spike trains recorded from nerve cells (Ekholm, 1972). Moran (1967) proposes a point process where the interpoint intervals are a sequence of independent pairs $\{W_{j1}, W_{j2}\}$ which each have some bivariate distribution F. He in particular discusses the case when F is such that W_{j1} and W_{j2} are marginally exponentially distributed with unit mean and their sum $W_j = W_{j1} + W_{j2}$ has density $\lambda e^{-\lambda}$, see also Cox and Isham (1980) and Lawrence (1972). Moran called his process a non-Markovian quasi-Poisson process; it has since been called Moran's process.

Isham (1977) extended this notion of Markovian dependency to higher dimensions. Here, the finite point process $X^n = \{X_j\}_{j=1}^n$ is a Markov sequence of points in a subset of $\mathbb{R}^k$ described through its spherical polar coordinates $\{n^{1/k}R_j, \Theta_j\}$, where Θ_j is a $(k-1)$-dimensional vector. The radial coordinates are rescaled such that the average density of events in X^n is the same for all n. As a special case, Isham considers the case $k = 2$ where the radial coordinates are independent with $\pi R_j \sim \mathcal{U}([0, 1/\lambda])$ while the angular variables form a 1-dependent Markov sequence with

$$f(\theta_{j+1}|\theta_j) = \frac{1}{2\pi}\big[1 + 2\rho\cos(\theta_{j+1} - \theta_j)\big], \quad |\rho| < \frac{1}{2}$$

if $j > 1$ and θ_1 uniformly distributed on $[0, 2\pi)$. It then follows that Θ_j is marginally uniformly distributed over $[0, 2\pi)$. The points in the process are thus marginally independent and the process X^n converges to a Poisson process as $n \to \infty$.

5.2 Lévy's Markov property

In 1948, Lévy defined a Markov property for processes on the real plane as follows: a process Z is said to have the *sharp Markov property* with respect to a given set in the plane if the σ-field generated by the values of the process on the set and the σ-field generated by the values outside the set are conditionally independent given the σ-field generated by the values on the boundary of the set (Lévy, 1948a). In a second paper that was published together with the theory part, Lévy provides a handful of very specific examples of processes fulfilling the sharp Markov property (Lévy, 1948b) and he concludes his theory paper by asking "Do there exist nondegenerate two-parameter processes with the sharp Markov property with respect to a sufficiently large class of sets?" (Merzbach and Nualart, 1990)

In the point process setting, the process Z corresponds to the counting process associated with a point process X. The first existence result for point processes is Carnal (1979) and Carnal and Walsh (1991), who show that the Poisson sheet fulfills the sharp Markov

property with respect to bounded relatively convex sets on the positive quadrant of the real plane. The Poisson sheet is defined by $Z(\mathbf{t}) = N(R_t)$, where N denotes the counting process for the unit Poisson point process on the positive quadrant of the plane and $R_t = (\mathbf{0}, \mathbf{t}]$ is a rectangle in this space. Later, Dalang and Walsh (1992) show that any two-parameter process on the positive quadrant of the plane of independent increments with no Gaussian part and where the map $\mathbf{t} \mapsto Z(\mathbf{t})$ is continuous in probability satisfies the sharp Markov property relative to all bounded sets. Merzbach and Nualart (1990) and Russo (1984) have previously proved special cases of this result for smaller families of sets. It follows from Theorem 2 in Wang (1981) that if a point process X (or its associated counting process) fulfills the sharp Markov property, it is either a Poisson process or a doubly stochastic Poisson process.

5.3 Ripley-Kelly Markov property

Markov point processes are nowadays usually considered to be the class of processes that satisfy the spatial Markov property described by Ripley and Kelly (1977). Ripley and Kelly's Markov property is the point process analog of a Markov random field (Besag, 1974). Let $\sim$ denote a neighborhood relation on the set S on which the point process is defined. A Markov point process is defined through properties of its density function. The density f with respect to the unit rate Poisson process on S should satisfy the following: Whenever $f(x) > 0$, it holds for all $\xi \in S \setminus x$ that $f(x \cup \xi)/f(x)$ depends only on x through the neighbors of ξ, that is, points $\eta \in x$ with $\eta \sim \xi$. The Hammersley-Clifford characterization offers an alternative representation (Ripley and Kelly, 1977), see also Isham (1981). It states that the density function f is the density of an (orderly) Markov point process if and only if there exists and interaction function φ such that

$$f(x) = \prod_{y \subset x} \varphi(y), \tag{3}$$

where an interaction function is a non-negative function such that $\varphi(x) = 1$ whenever there exist $\xi, \eta \in x$ with $\xi \nsim \eta$. These processes have been thoroughly studied in the recent literature, see e.g. Daley and Vere-Jones (2008, Chapter 10), Møller and Waagepetersen (2004, Chapter 6), and van Lieshout (2000).

Point processes whose densities fulfill Equation (3) are much older than their formal characterization. The simplest nearest neighbor interaction potential is of the form $v^{1(r \leq R}$ where $0 \leq v \leq 1$ measures the strength of inhibition for points within a radius R (Strauss, 1975). Such processes occur also in statistical physics as model for large interacting particle systems and are in this context usually called Gibbs processes, see e.g. Preston (1976) or Ruelle (1969). A brief overview of Markov point process models with their roots in statistics is given in Daley and Vere-Jones (2008, Chapter 10) who also discuss extensions of the Hammersley-Clifford characterization above.

6 Intensity-based processes

A different way of constructing point process models on the positive half-line originated with the idea of a random hazard function (Gaver, 1963). The idea led to the concept of a conditional intensity function (Yashin, 1970), and the subsequent realization that the conditional intensity determines a regular point process uniquely (Papangelou, 1972). Among the important consequences is the possibility of writing down an explicit likelihood, tools for simulating these processes through a random time change of a unit rate Poisson process, and the consequent possibility of a residual analysis (Section 7.2 of Daley and Vere-Jones (2005) describes much of this).

The first process defined through its conditional intensity was the self-exciting process (Hawkes, 1971), but even Wold's process of Markov intervals can be written in this form. More generally, Aalen's (Aalen, 1978; Andersen et al., 1993) work on multiplicative processes originated in this idea.

On the line, doubly stochastic processes are of course defined through their conditional intensity, given a σ-algebra generated by the history of the point process and of the process determining the rate.

7 Concluding remarks

In this paper we have attempted to go through systematically the history of the main classes of univariate unmarked point processes. We have rediscovered an early use of the Poisson process, and a forgotten definition of Markov point processes. In addition, we have traced the first use of some terms in the point process literature, including the term itself, the names Poisson process and Cox process, and the term point pattern. Occasionally some workers in the field of point processes may be a little surprised by some of the first occurrences we have managed to document, but there is of course no guarantee that there may not be an earlier instance in a literature we have not had access to.

Acknowledgments

The authors are grateful for discussions with and references from David Brillinger, Daniela Cocchi, David Cox, Geoffrey Grimmet, Georg Lindgren, Stephen Stigler and David Vere-Jones. Part of the work was done when the second author was visiting the University of Washington in 2011 and another part when the first author visited Heidelberg University in 2012.

References

Aalen, O. (1978). Nonparametric inference for a family of counting processes. *Ann. Stat.*, 6(4):701–726.

Abbe, E. (1879). Über Blutkörper-Zählung. *Jena Zeitschrift von Medizin und Naturwissenschaften*, 13:98–105.

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York.

Bartlett, M. (1954). Processus stochastiques ponctuels. *Ann. I. H. Poincaré*, 14:35–60.

Bartlett, M. (1963). The spectral analysis of point processes. *J. Roy. Stat. Soc. B*, 25(2):264–296.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. Roy. Statist. Soc. Ser. B*, 36:192–236.

Bok, B. J. (1934). The apparent clustering of external galaxies. *Harvard College Observatory Bulletin*, (895):1–8.

Boltzmann, L. (1868). Studien über das Gleichgewicht der lebendigen Kraft zwischen bewegten materiellen Punkten. *Sitzungsberichte Mathematische und Naturwissenschaftliche Klasse Kaiserliche Akademie der Wissenschaften*, 58:517–560.

Brush, S. G. (2003). *The Kinetic Theory of Gases*. Imperial College Press, London.

Buijs-Bullet, C. H. D. (1858). Über die Art von Bewegung, welche wir Wärme und Elektricität nennen. *Ann. Phys.*, 103:240.

Campbell, N. (1909). The study of discontinuous phenomena. *Proc. Cambr. Phil. Soc.*, 15:117–136.

Carnal, E. (1979). *Processus markoviens à plusieurs parametres*. PhD thesis, École Polytechnique Fédérale die Lausanne.

Carnal, E. and Walsh, J. B. (1991). Markov properties for certain random fields. In Mayer-Wolf, E., Merzbach, E., and Schwartz, A., editors, *Stochastic analysis: liber amicorum for Moshe Zakai*, pages 91–110. Academic, San Diego.

Charlier, C. V. L. (1905). Die zweite Form des Fehlergesetzes. *Ark. Mat. Astr. Fys.*, 2:1–8.

Charlier, C. V. L. (1922). How an infinite world may be built up. *Ark. Mat. Astr. Fys.*, 16(22).

Clausius, R. (1857). Über die Art der Bewegung, welche wir Wärme nennen. *Ann. Phys.*, 100:353–80.

Clausius, R. (1858). Über die mittlere Länge der Wege, welche bei Molekularbewegung gasförmigen Körper von den einzelnen Molecülen zurück gelegt werden, nebst einigen anderen Bemerkungen über die mechanischen Wärmetheorie. *Ann. Phys.*, 105:239–258.

Cox, D. R. (1955). Some statistical methods connected with series of events (with discussion). *J. Roy. Stat. Soc. B*, 17:129–164.

Cox, D. R. (1963). Some models for series of events. *Bull. I. S. I.*, 40:737–746.

Cox, D. R. and Isham, V. (1980). *Point processes*. Chapman & Hall, London.

Cox, D. R. and Lewis, P. A. (1966). *Statistical Analysis of Series of Events*. Methuen, London.

Cramér, H. (1937). *Random variables and probability distributions*. Cambridge University Press, Cambridge.

Cramér, H. (1969). Historical review of Filip Lundberg's work on risk theory. *Skand. Aktuar. (Suppl.)*, 52:6–12.

Dalang, R. C. and Walsh, J. B. (1992). The sharp Markov property of Lévy sheets. *Ann. Probab.*, 20(2):591–626.

Daley, D. and Vere-Jones, D. (2005). *An Introduction to the Theory of Point Processes*, volume I: Elementary Theory and Methods. Springer, 2nd edition.

Daley, D. and Vere-Jones, D. (2008). *An Introduction to the Theory of Point Processes*, volume II: General Theory and Structure. Springer, 2nd edtion edition.

de Finetti, B. (1929). Sulla possibilità di valori eccezionali per una legge di incrementi aleatori. *Rendiconti della R. Academia Nazionali dei Lincei*, 10:548–553.

Diggle, P. J., Besag, J., and Gleaves, J. T. (1976). Statistical analysis of spatial point patterns by means of distance methods. *Biometrics*, 32:659–667.

Diggle, P. J. and Milne, R. K. (1983). Negative binomial quadrat counts and point processes. *Scand. J. Statist.*, 10:257–267.

Doob, J. L. (1937). Stochastic processes depending on a continuous parameter. *T. Am. Math. Soc.*, 42:107–140.

Eggenberger, F. and Pólya, G. (1923). Über die Statistik verketteter Vorgänge. *Z. Angew. Math. Mech.*, 1:279–289.

Ekholm, A. (1972). A generalization of the two-state two-interval semi-Markov model. In Lewis, P. A. W., editor, *Stochastic point processes: statistical analysis, theory and applications*, pages 199–252. Wiley, New York.

Erlang, A. K. (1909). Sandsynlighedsregning og telefonsamtaler. *Nyt Tidsskrift for Mathematik, Afd. B*, 20:33–40.

Feller, W. (1936). Über die Theorie der stochastischen Prozesse. In *Comptes Rendus du Congrès International des Mathématiciens Oslo*, volume 2, pages 192–196, Oslo. Brøggers.

Feller, W. (1940). On the integro-differential equations of purely discontinuous Markoff processes. *T. Am. Math. Soc.*, 48:488–515.

Fry, T. C. (1928). *Probability and its Engineering Uses*. Van Nostrand, New York.

Gaver, D. P. (1963). Random hazard in reliability problems. *Technometrics*, 5:211–216.

Gelfand, A. E., Diggle, P., Fuentes, M., and Guttorp, P. (2010). *Handbook of spatial statistics*. Chapman & Hall/CRC.

Grimmet, G. and Stirzaker, D. (2001). *Probability and Random Processes*. Oxford University Press, Oxford, third edition.

Guttorp, P. (1991). Spatial processes in ecology. In Besag, J. and Simpson, J., editors, *Spatial Statistics and Digital Image Analysis*, pages 129–146. National Academy Press, Washington.

Haight, F. A. (1967). *Handbook of the Poisson Distribution*. Wiley, New York.

Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58:83–90.

Hobbs, P. V. and Locatelli, J. D. (1978). Rainbands, precipitation cores and generating cells in a cyclonic storm. *J. Atmos. Sci.*, 35:230–241.

Isham, V. (1977). A Markov construction for a multidimensional point process. *J. Appl. Prob.*, 14:507–515.

Isham, V. (1981). An introduction to spatial point processes and Markov random fields. *Int. Stat. Rev.*, 49:21–43.

Istok, J. D. and Boersma, L. (1989). A stochastic cluster model for hourly precipitation data. *J. Hydrol.*, 106:257–285.

Jalilian, A., Guan, Y., and Waagepetersen, R. (2011). Decomposition of variance for spatial Cox processes. *Scand. J. Stat.*, page To appear.

Kallenberg, O. (1983). *Random Measures*. Akademie-Verlag, Berlin.

Kavvas, M. L. and Delleur, J. W. (1981). A stochastic cluster model for daily rainfall sequences. *Water Resour. Res.*, 17:1151–1160.

Khinchin, A. (1933). Asymptotische Gesetze de Wahrscheinlichkeitsrechning. *Ergebnisse der Mathematik und ihrer Grenzgebiete*, 2:1–77.

Kolmogorov, A. (1931). Über die analytischen Methoden in der Wahrscheinlichkeitsrechning. *Math. Ann.*, 104:415–458.

Kolmogorov, A. N. (1935). La transformation de Laplace dans les escapes linéaires. *C. R. Acad. Sci. Paris*, 200:1717–1718.

Krickeberg, K. (1972). The Cox process. In *Sympos. Math.*, volume 9, pages 151–167. I. N. D. A. M.

Lampard, D. G. (1968). A stochastic process whose successive intervals between events form a first order Markov chain: I. *J. Appl. Probab.*, 5:648–668.

Laslett, G. M. (1978). Mixing of cluster point processes. *J. Appl. Probab.*, 15:715–725.

Lawrence, A. J. (1972). Some models for stationary series of univariate events. In Lewis, P. A. W., editor, *Stochastic point processes: statistical analysis, theory and applications*, pages 199–252. Wiley, New York.

Le Cam, L. (1947). Un instrument d'étude des fonctions aléatoires, la fonctionelle caractéristique. *C. R. Acad. Sci., Paris*, 224:710–711.

Le Cam, L. M. (1960). A stochastic description of precipitation. In Neyman, J., editor, *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, vol. 3*, pages 165–186, California, Berkeley.

Lee, P. A. and Ong, S. H. (1986). Higher-order and non-stationary properties of Lampard's stochastic reversible coutner system. *Statistics*, 17(2):261–278.

Lévy, P. (1939). Sur certains processus stochastiques homogènes. *Compositio Mathematica*, 7:283–339.

Lévy, P. (1948a). Chaînes doubles de Markoff et fonctions alétoires de deux variables. *C. R. Acad. Sci., Paris*, 226:53–55.

Lévy, P. (1948b). Exemples de processus double de Markoff. *C. R. Acad. Sci. Paris*, 226:307–308.

Lévy, P. (1954a). Processus semi-markoviens. In *Porc. Internat. Congr. Math.*, volume 3, pages 416–426. Amsterdam.

Lévy, P. (1954b). Systemèmes semi-Markoviens à au plus infinité dénombrable d'états possibles. In *Porc. Internat. Congr. Math.*, volume 2, page 294. Amsterdam.

Lewis, P. A. W. (1964). A branching Poisson process model for the analysis of computer failure patters (with Discussion. *J. Roy. Stat. Soc. B*, 26:398–456.

Lewis, P. A. W. (1972). Introduction. In Lewis, P. A. W., editor, *Stochastic point processes: statistical analysis, theory and applications*, pages xiii–xxii. Wiley, New York.

Lundberg, F. (1903). *I. Approximerad framställning af sannolikhetsfunktionen. II. Återförsäkring af kollektivrisken*. Almqvist & Wiksell, Uppsala.

Lundberg, O. (1940). *On random processes and their application to sickness and accident statistics*. Almqvist & Wiksell, Uppsala.

Matérn, B. (1971). Doubly stochastic Poisson processes in the plane. In Patil, G. P., Pielou, E. C., and Waters, W. E., editors, *Statistical Ecology*, volume I, pages 195–213. Pennsylvania State University Press.

Mecke, J. (1968). Eine charakteristische eigenschaft der dopplet stochastischen poissonschen prozesse. *Z. Wahrscheinlichkeit.*, 11:74–81.

Merzbach, E. and Nualart, D. (1990). Markov properties for point processes on the plane. *Ann. Probab.*, 18(1):342–358.

Michell, J. (1767). An inquiry into the probable Parallax, and Magnitude, of the Fixed Stars, from the Quantity of Light which they afford us, and the particular Circumstances of their Situation. *Philosophical Transactions*, 57:234–264.

Milne, R. K. and Westcott, M. (1972). Further results for Gauss-Poisson processes. *Advances in Applied Probability*, 4:151–176.

Møller, J. and Waagepetersen, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman & Hall/CRC.

Moran, P. A. P. (1967). A non-markovian quasi-poisson process. *Studia Sci. Math. Hungar.*, 2:425–429.

Newcomb, S. (1860). Notes on the theory of probabilities. *The Mathematical Monthly*, 2:134–140.

Newman, D. (1970). A new family of point processes which are characterized by their second moment properties. *J. Appl. Probab.*, 7:338–358.

Neyman, J. (1939). On a new class of 'contagious' distributions applicable to entomology and bacteriology. *Ann. Math. Stat.*, 10:35–57.

Neyman, J. and Scott, E. L. (1952). A theory of the spatial distribution of galaxies. *Astrophys. J.*, 116:144–163.

Neyman, J. and Scott, E. L. (1958). Statistical approach to problems of cosmology (with discussion). *J. Roy. Stat. Soc. B*, 20:1–43.

O'Connor, J. J. and Robertson, E. F. (2000a). Phillip Ludwig von Seidel. Accessed January 2011. Available from: `"http://www-history.mcs.st-andrews.ac.uk/Biographies/Seidel"`.

O'Connor, J. J. and Robertson, E. F. (2000b). Rudolf Julius Emmanuel Clausius. Accessed January 2011. Available from: `"http://www-history.mcs.st-andrews.ac.uk/Biographies/Clausius"`.

Palm, C. (1943). *Intensitätsschwankungen im Fernsprechverkehr*, volume 44 of *Ericsson Technics*. Stockholm.

Papangelou, F. (1972). Summary of some results on point and line processes. In Leviws, P. A. W., editor, *Stochastic Point Processes*, pages 522–532, New York. Wiley.

Pólya, G. (1918). Zahlentheoretisches und Wahrscheinlichkeitstheoretisches über die Sichtweite im Walde. *Archiv für Mathematik und Physik*, 27:135–142.

Pólya, G. (1919). Zur Statistik der sphärischen Verteilung der Fixsterne. *Astron. Nachr.*, 208:175–180.

Preston, C. (1976). *Random Fields*. Lecture Notes in Mathematics 534. Springer-Verlag, Berlin.

Quenouille, M. H. (1949). Problems in plane sampling. *Ann. Math. Stat.*, 20(3):355–375.

Rathbun, S. L. and Cressie, N. (1994). A space-time survival point process for a longleaf pine forest in southern georgia. *J. Am. Stat. Assoc.*, 89:pp. 1164–1174.

Rényi, A. (1967). Remarks on the Poisson process. *Stud. Sci. Math. Hung.*, 2:119–123.

Ripley, B. D. and Kelly, F. P. (1977). Markov point processes. *J. Lond. Math. Soc.*, 15(2):188–192.

Ruelle, D. (1969). *Statistical Mechanics: rigorous results*. W. A. Benjamin, Reading, MA.

Russo, F. (1984). Étude de la propriété de Markov étroite en relation avec les processus planaires à accroissements indépendants. In *Séminaire de Probabilités XVIII. Lecture Notes in Math*, volume 1059, pages 353–378. Springer, Berlin.

Rutherford, E. and Geiger, H. (1910). The probability variations in the distribution of α-particles. *Philos. Mag.*, 20:698–707.

Seidel, L. (1876). Ueber die Probabilitäten solche Ereignisse, welche selten vorkommen, obgleich sie umbeschränkt oft möglich sind. *Sitzungsberiche der königsche bayerische Akademie der Wissenschaften, Mahematische physikalische Classe*, VI:44–50. Sitzung vom 5. Februar, 1876.

Skellam, J. (1952). Studies in statistical ecology: I. Spatial pattern. *Biometrika*, 39:346–362.

Smith, W. L. (1955). Regenerative stochastic processes. *Proc. Roy. Soc. Ser. A*, 232:6–31.

Stigler, S. M. (1982). Poisson on the Poisson distribution. *Stat. Probabil. Lett.*, 1:33–35.

Strauss, D. J. (1975). A model for clustering. *Biometrika*, 62(2):467–475.

Student (W. S. Gosset) (1907). On the error of counting with a haemacytometer. *Biometrika*, 5:351–360.

Takács, L. (1954). Some investigations concerning recurrent stochastic processes of a certain type. *Magyar Tud. Akad. Mat. Kutató Int. Kozl.*, 3:115–128. In Hungarian.

Thomas, M. (1949). A generalization of Poisson's binomial limit for use in ecology. *Biometrika*, 36:18–25.

Thompson, H. R. (1954). A note on contagious distributions. *Biometrika*, 41:267–271.

Thompson, H. R. (1955). Spatial point processes, with applications to ecology. *Biometrika*, 42:102–115.

Todhunter, I. (1965). *A History of the Mathematical Theory of Probability from the time of Pascal to that of Laplace*. Chelsea, New York.

van Lieshout, M. N. M. (2000). *Markov point processes and their applications*. Imperial College Press, London.

von Schweider, E. (1905). Über Schwankungen der radioaktiven Umwandling. In *Comptes Rendus du premier congrès international pour L'Ètude de la Radiologie et de l'Ionisation*, pages 1–3, Bruxelles.

Wang, J. G. (1981). Some remarks on processes with independent increments. In *Séminaire de Probabilités XV 1979/80*, pages 627–631. Lecture Notes in Mathematics nr. 850.

Whitworth, W. (1886). *Choice and Chance*. Deighton Bell, Cambridge, 4th edition.

Wiener, N. (1938). The homogeneous chaos. *American Journal of Mathematics*, 60:897–936.

Wold, H. (1948). On stationary point processes and Markov chains. *Skand. Aktua.*, 31:229–240.

Wold, H. (1949). Sur les processus stationnaires ponctuels. In *Le calcul des probabilités et ses applications*, pages 75–86. C.N.R.S., Paris.

Yashin, A. (1970). Filtering of jump processes. *Automatic Remote Control*, 1970:725–730.