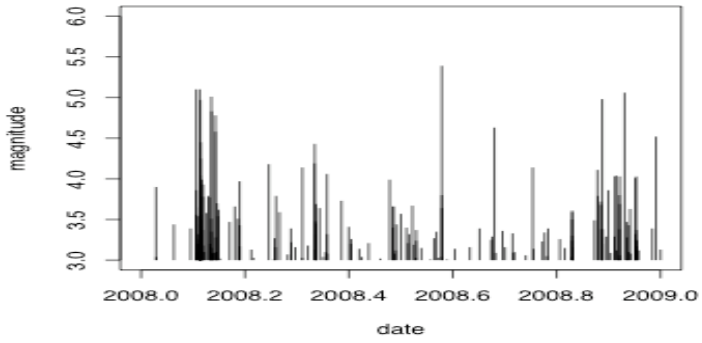
Introduction to point processes.

Frederic Paik Schoenberg

Abstract. Point process models are useful for describing phenomena occurring at random locations and/or times. Following a review of basic concepts, some important models are surveyed including Poisson processes, renewal processes, Hawkes processes, and Markovian point processes. Techniques for estimation, simulation and residual analysis for point processes are also briefly discussed.

A point process is a random collection of points falling in some space. In most applications, each point represents the time and/or location of an event, such as a lightning strike or earthquake. When modeling purely temporal data, the space in which the points fall is simply a portion of the real line.

Figure 1 shows an example of origin times of Southern California earthquakes of magnitude at least 3.0 occurring in 2008, obtained from the Southern California Earthquake Data Center. Point process data arise in a wide variety of scientific applications, including epidemiology, ecology, forestry, mining, hydrology, astronomy, ecology, and meteorology.



FORMAL DEFINITIONS

There are several ways of characterizing a point process. The usual approach is to define a point process $\boldsymbol{N}$ as a random measure on a complete separable metric space S taking values in the non-negative integers $\boldsymbol{Z}^+$ (or infinity). In this framework the measure $N(A)$ represents the number of points falling in the subset A of S. Attention is typically restricted to the case where N may contains only finitely many points on any bounded subset of S.

The random measure definition above is perhaps not the most intuitive means of characterizing a point process, and several alternatives exist. Consider the case of a temporal point process, e.g. the times of events occurring between time 0 and time T.

One may characterize N as an ordered list $\{\tau_1, ..., \tau_n\}$ of event times. One may alternatively convey the equivalent information about N via the inter-event times $\{u_1, ..., u_n\}$, where $u_i = \tau_i - \tau_{i-1}$, with the convention that $\tau_0 = 0$.

A temporal point process N may alternatively be described by a counting process $N(t)$, where for any time t between 0 and T, $N(t)$ is the number of points occurring at or before time t. The resulting process $N(t)$ must be non-decreasing and right-continuous, and take only non-negative integer values. One could thus define a temporal point process as any non-decreasing, right-continuous $\boldsymbol{Z}^+$-valued process. This sort of definition has been used by [1,2,3], and others.

The counting process definition of a point process lends itself naturally to discussions of martingales and distributional theory. On the other hand, the random measure formulation has the advantage of generalizing immediately to point processes in higher dimensions and in abstract spaces. For simplicity, the remainder of this article will deal only with purely temporal point processes, where the domain S is simply a portion of the real (time-)line.

A point process is called *simple* if with probability one, all its points τ_i occur at distinct times. A closely related concept is orderliness: a point process N is *orderly* if for any time t, $P\{N(t, t+\Delta t] > 1\}/\Delta t \to 0$ as $\Delta t \to 0$.

The points of a point process are typically indistinguishable other than by their times and/or locations. When there is other important information to be stored along with each point, such as a list of times of wildfire origin times along with their corresponding burn areas, or the earthquake times and magnitudes shown in Figure 1, the result may be viewed as a *marked* point process.

The relationship between time series and point processes is worth noting. Many datasets that are traditionally viewed as realizations of (marked) point processes could in principle also be regarded as time series, and vice versa. For instance, a sequence of earthquake origin times is typically viewed as a temporal point process; however, one could also store such a sequence as a time series consisting of zeros and ones, with the ones corresponding to earthquakes. The main difference is that for a point process, the points can occur at any times in a continuum, whereas the time intervals are discretized in the time series case.

POINT PROCESS MODELS

The most important point process model is the *Poisson process*, which is a simple point process N such that the number of points in any set follows a Poisson distribution and the numbers of points in disjoint sets are independent. That is, N is a Poisson process if $N(A_1), ..., N(A_k)$ are independent Poisson random variables, for any disjoint, measurable subsets $A_1, ..., A_k$ of S.

The behavior of a simple temporal point process N is typically modeled by specifying its *conditional intensity*, $\lambda(t)$, which represents the infinitesimal rate at which events are expected to occur around a particular time t, conditional on the prior history of the point process prior to time t. Formally, the conditional intensity associated with a temporal point process N may be defined via the limiting conditional expectation

$$\lambda(t) = lim_{\Delta t \to 0} \, E\{N[t, t + \Delta t) \mid H_t\} / \Delta t,$$

provided this limit exists, where H_t is the history of the point process N over all times strictly prior to time t. Some authors instead define λ as the conditional probability

$$\lambda(t) = lim_{\Delta t \to 0} \, P\{N[t, t + \Delta t) > 0 \mid H_t] / \Delta t \, ,$$

since the two definitions are equivalent for orderly point processes. As all finite-dimensional distibutions of N are uniquely determined by the conditional intensity [4,5], in modeling N it suffices to prescribe a model for λ.

For a temporal point process originating at time 0, one may define the *compensator* $A(t)$ as the integral of the conditional intensity from time 0 to time t. The compensator may equivalently be defined as the unique non-negative non-decreasing predictable process $A(t)$ such that $N[0,t) - A(t)$ is a martingale [1,2].

λ may be estimated nonparametrically [6,7,8,9], or via a parametric model. In general, $\lambda(t)$ depends not only on t but also on the times τ_i of preceding events. When N is a Poisson process, however, λ is deterministic; i.e. $\lambda(t)$ depends only on t.

A stationary Poisson process has constant conditional rate: $\lambda(t) = \alpha$, for all t. This model posits that the risk of an event is the same at all times, regardless of how frequently such events have occurred previously. For a non-stationary Poisson process, $\lambda(t)$ is some function of t. A generalization is the Cox process, or doubly-stochastic Poisson process, which is a Poisson process whose intensity function is randomly generated.

Another important elementary type of temporal point process is the *renewal process*. A renewal process is a simple point process such that the inter-event times $\{u_1, ..., u_n\}$ are independent (typically i.i.d.) random variables. In the i.i.d. case, the density function governing each inter-event time is called the *renewal density*. Note that the renewal density f is ordinarily taken to be a density on the half-line, i.e. $f(t) = 0$ for $t < 0$. Renewal models embody the notion that the hazard of an event occurring at a particular time depends only on the time since the most recent event. In fire hazard analysis, for example, such a model is consistent with the theory of fuel loading followed by complete fuel depletion in the event of a fire.

A point process N may be called *self-exciting* if $cov\{N(s,t), N(t,u)\} > 0$ *for* $s < t < u$. N is *self-correcting* if instead this covariance is negative. Thus the occurrence of points in a self-exciting point process causes other points to be more likely to occur, whereas in a self-correcting process, the points have an inhibitory effect. By definition, a Poisson process is neither self-exciting nor self-correcting.

Self-exciting point process models are often used in seismology to model events that are temporally clustered. A commonly used example is the *Hawkes process*, where the conditional intensity is given by

$$\lambda(t) = \mu(t) + \sum_{i:\, \tau_i < t} \nu(t-\tau_i),$$

where $\mu(t)$ represents the deterministic background rate and the function ν governs the clustering density. An example where $\mu(t)$ is constant was offered by [10]. Hawkes models are commonly used in seismology, where they are sometimes called epidemic-type aftershock sequence (ETAS) models, encompassing the notion that earthquakes can have aftershocks, and those aftershocks can have aftershocks, etc. A form of the clustering density ν that is commonly used in modeling earthquake aftershocks is the Omori-Utsu law [11]

$$\nu(t) = \kappa(t+c)^{-p},$$

which corresponds to power-law decay in the clustering behavior over time.

Alternative versions of clustered point processes are formed by generating a sequence of parents and then placing clusters of points (offspring) around each parent. For the Neyman–Scott process, for instance, the offspring points are independently and identically distributed around the parents, and for the Bartlett-Lewis the offspring points are each generated via a renewal process originating at the corresponding parent.

Self-correcting models are used in ecology, forestry and other fields to model occurrences that are well-dispersed. Such models may be useful in describing births of species, for example, or in seismology for modeling earthquake catalogs after aftershocks have been removed [12]. An important example is the Markovian point process model (so called because the conditional rate obeys the Markov property) where

$$\lambda(t) = g\{t, N[0,t)\}.$$

For example, the function g may be selected so that λ takes the form

$$\lambda(t) = \exp\{\alpha + \beta\,(t - \rho\, N[0,t))\},$$

where α, β, and ρ are constants [13,14,15]. The parameters α and β govern the background rate and trend in the occurrences, while the product $\beta\rho$ represents the decrease in conditional rate of future events caused by each event, which may be due to diminished fuel load in the case of wildfire modeling, for instance, or the release of strain energy in the seismological case.

TRANSFORMATIONS

Some important operations on point processes include superposition, thinning, and rescaling. Graphically, these operations correspond, respectively, to overlaying points of one process onto those of another point process, deleting certain points, and stretching out the time axis.

The Poisson process frequently arises as a limiting process resulting from transformations of point processes. For example, under quite general conditions, when k independent copies of a temporal point process are superposed and the time axis is rescaled by a factor of k, the resulting process converges to a Poisson process (for details see [16,17]). Similar

results can be obtained for randomly thinned point processes [18] or rescaled point processes [19].

ESTIMATION AND SIMULATION

The parameter vector θ for a point process model with conditional intensity $\lambda(\theta)$ is typically estimated by maximizing the log-likelihood function

$$L(\theta) = \sum_i \log \lambda(\tau_i ; \theta) - \int_S \lambda(t)dt.$$

The consistency, asymptotic normality, and efficiency of maximum likelihood estimators, have been established under standard conditions [20,21], and standard errors may be obtained via the Hessian of the log-likelihood [22,23,24].

Alternatively, simulations may be useful for obtaining approximate standard errors and for other types of inference. An effective simulation method for point processes based on random thinning was devised in [25,26]. The procedure, which works for point processes whose conditional rate λ is bounded (or locally bounded) above by some constant b, involves simulating a (locally) stationary Poisson process with intensity b and thinning it, keeping each point τ_i with probability $\lambda(\tau_i) / b$.

MODEL EVALUATION

The fit of a parametric point process model is often assessed using a likelihood score such as the Akaike Information Criterion (AIC), which is defined as $2p - 2L(\theta)$, where p is the number of fitted parameters. The AIC rewards a model for higher likelihood and penalizes a model for overfitting; lower AIC indicates better fit.

Residual analysis methods for point processes often involve comparing the number of points observed with the integrated conditional intensity within specified portions of S. Methods for normalizing and inspecting such residuals are described in [27].

Another useful technique for evaluating point process models is via rescaling. The method essentially involves rescaling the time axis of the observed point process N at time t by a factor of $\lambda(t)$. More specifically, if the points are observed from time 0 to time T, then each point τ_i is moved to the new time $A(\tau_i)$, where A is the compensator. The resulting process M is a stationary Poisson process with unit rate, provided the original point process is simple [19]. Similarly, one may inspect residuals obtained by randomly thinning the process: that is, keeping each point τ_i independently with probability inversely proportional to $\lambda(\tau_i)$. As with rescaled residuals, the resulting thinned residuals will be distributed according to a stationary Poisson process [28]. In practice, one may use the estimated intensity or compensator in place of the true intensity or compensator and inspect the rescaled or thinned process for uniformity. Several tests exist for this purpose, with different uses depending on the alternative hypotheses (e.g. [3,29,30,31,32,33]). Some of the most useful are tests based on second and higher order properties [34,35,36,37]. There are also more general second-order tests for point processes that do not rely on a stationary Poisson null hypothesis [38,39,40].

For further reading on point processes, see [5,17].

References:

[1] Jacod J. Multivariate point processes: predictable projections, Radon-Nikodym derivatives, representation of martingales. Z. Wahrsch. Verw. Gebiete 1975; 31: 235-253.

[2] Brémaud P. *Point Processes and Queues: Martingale Dynamics*. 1981; Springer-Verlag, New York.

[3] Andersen P, Børgan O, Gill R, Keiding N. *Statistical Models Based on Counting Processes*. 1993; Springer-Verlag, New York.

[4] Fishman PM, Snyder D. (1976) The statistical analysis of space-time point processes. IEEE Transactions on Information Theory 1976; 22: 257–274.

[5] Daley DJ, Vere-Jones D. *An Introduction to the Theory of Point Processes, Volume I: Elementary Theory and Methods, 2nd ed*. 2003; Springer, New York.

[6] Diggle P. A kernel method for smoothing point process data. Appl. Stat. 1985; 34: 138-147.

[7] Guttorp P, Thompson M. Nonparametric estimation of intensities for sampled counting processes. J. Royal Stat. Soc. B 1990; 52: 157-173.

[8] Vere-Jones D. Statistical methods for the description and display of earthquake catalogs. In S*tatistics in the Environmental and Earth Sciences*, Walden A, Guttorp P, eds., 1992; Edward Arnold, London, pp. 220-246.

[9] Brillinger D. Some wavelet analyses of point process data. 31st Asilomar Conference on Signals, Systems and Computers, 1998; IEEE Computer Society, pp. 1087-1091.

[10] Hawkes A, Adamopoulos L. Cluster models for earthquakes -- regional comparisons. Bull. Int. Statist. Inst. 1973; 45(3): 454-461.

[11] Ogata Y. Statistical models for earthquake occurrences and residual analysis for point processes. J. Amer. Statist. Assoc. 1988; 83: 9-27.

[12] Vere-Jones D. Earthquake predicton -- a statistician's view. J. Phys. Earth 1978; 26: 129-146.

[13] Isham V, Westcott M. A self-correcting point process. Stoch. Proc. Appl. 1979; 8: 335-347.

[14] Ogata Y, Vere-Jones D. Inference for earthquake models: a self-correcting model. Stoch. Proc. Appl. 1984; 17: 337-347.

[15] Vere-Jones D, Ogata Y. On the moments of a self-correcting process. J. Appl. Prob. 1984; 21: 335-342.

[16] Kallenberg O. *Random Measures, 3rd ed*. 1983; Akademie-Verlag, Berlin.

[17] Daley DJ, Vere-Jones D. *An Introduction to the Theory of Point Processes Volume II: General Theory and Structure, 2nd ed*., 2008; Springer, New York.

[18] Westcott M. Simple proof of a result on thinned point processes. Ann. Probab. 1976; 4: 89-90.

[19] Meyer P. Démonstration simplifée d'un théorème de Knight. Lecture Notes in Mathematics 1971; 191: 191-195.

[20] Brillinger D. Statistical inference for stationary point processes. Stochastic Processes and Related Topics 1975; 1: 55-100.

[21] Ogata Y. The asymptotic behavior of maximum likelihood estimators for stationary point processes. Ann. Inst. Stat. Math. 1978; 30: 243-262.

[22] Kutoyants Y. *Parameter Estimation for Stochastic Processes*. 1984; Helderman, Berlin.

[23] Dzhaparidze K. On asymptotic inference about intensity parameters of a counting process. Proc. 45th Session Int. Stat. Inst. 1985; 4(23.2): 1-15.

[24] Rathbun S, Cressie N. Asymptotic properties of estimators for the parameters of spatial inhomogeneous Poisson point processes. Adv. Appl. Prob. 1994; 26: 122-154.

[25] Lewis P, Shedler G. Simulation of nonhomogeneous Poisson processes by thinning. Naval Research Logistics Quarterly 1979; 26: 403-413.

[26] Ogata Y. On Lewis' simulation method for point processes. IEEE Trans. Inf. Theory 1981; IT-27: 23-31.

[27] Baddeley A, Turner R, Møller J, Hazelton M. Residual analysis for spatial point processes. J. Royal Stat. Soc. B 2005; 67(5): 617--666.

[28] Schoenberg FP. Multi-dimensional residual analysis of point process models for earthquake occurrences. J. Amer. Statist. Assoc. 2003; 98(464): 789--795.

[29] Diggle P. On parameter estimation and goodness-of-fit testing for spatial point patterns. Biometrics 1979; 35: 87-101.

[30] Dijkstra J, Rietjens T, Steutel F. A simple test for uniformity. Statistica Neerlandica 1984; 38: 33-44.

[31] Lisek B, Lisek M. A new method for testing whether a point process is Poisson. Statistics 1985; 16: 445-450.

[32] Lawson A. On tests for spatial trend in a non-homogeneous Poisson process. J. Appl. Stat. 1988; 15: 225-234.

[33] Arsham H. A modified one-sided K-S confidence region narrower in the tail. Comm. Stat. A 1987; 16, 17-28.

[34] Bartlett M. The spectral analysis of two-dimensional point processes. Biometrika 1964; 51, 299-311.

[35] Davies R. Testing the hypothesis that a point process is Poisson. Adv. Appl. Probab. 1977; 9: 724-746.

[36] Ripley B. Tests of `randomness' for spatial point patterns. J. Royal Stat. Soc. B 1979; 41: 368-374.

[37] Heinrich L. Goodness-of-fit tests for the second moment function of a stationary multidimensional Poisson process. Statistics 1991; 22: 245-278.

[38] Baddeley A, Møller J, Waagepetersen R. Non and semi-parametric estimation of interaction in inhomogeneous point patterns. Statistica Neerlandica 2000; 54(3): 329-350.

[39] Veen A, Schoenberg FP. Assessing spatial point process models for California earthquakes using weighted K-functions: analysis of California earthquakes. in *Case Studies in Spatial Point Process Models*, Baddeley A, Gregori P, Mateu J, Stoica R, Stoyan D (eds.), 2005; Springer, NY, pp. 293-306.

[40] Adelfio G, Schoenberg FP. Point process diagnostics based on weighted second-order statistics and their asymptotic properties. Annals of the Institute of Statistical Mathematics 2009; 61(2), 75-96.