Off the Beaten Path Tutorial: Stochastic Processes and Simulations – Volume 1

Vincent Granville, Ph.D.
Data Shaping Solutions, LLC

Anacortes, WA, February 2022

Note: External links (in blue) and internal references (in red) are clickable throughout this document. Keywords highlighted in orange are indexed; those in red are both indexed and in the glossary section.

Contents

About this Textbook

This scratch course on stochastic processes covers significantly more material than usually found in traditional books or classes. The approach is original: I introduce a new yet intuitive type of random structure called perturbed lattice or Poisson-binomial process, as the gateway to all the stochastic processes. Such models have started to gain considerable momentum recently, especially in sensor data, cellular networks, chemistry, physics and engineering applications. I present state-of-the-art material in simple words, in a compact style, including new research developments and open problems. I focus on the methodology and principles, providing the reader with solid foundations and numerous resources: theory, applications, illustrations, statistical inference, references, glossary, educational spreadsheet, source code, stochastic simulations, original exercises, videos and more.

Below is a short selection highlighting some of the topics featured in the textbook. Some are research results published here for the first time.

GPU clustering	Fractal supervised clustering in GPU (graphics processing unit) using image filtering techniques akin to neural networks, automated black-box detection of the number of clusters, unsupervised clustering in GPU using density (gray levels) equalizer
Inference	New test of independence, spatial processes, model fitting, dual confidence regions, minimum contrast estimation, oscillating estimators, mixture and surperimposed models, radial cluster processes, exponential-binomial distribution with infinitely many parameters, generalized logistic distribution
Nearest neighbors	Statistical distribution of distances and Rayleigh test, Weibull distribution, properties of nearest neighbor graphs, size distribution of connected components, geometric features, hexagonal lattices, coverage problems, simulations, model-free inference
Cool stuff	Random functions, random graphs, random permutations, chaotic convergence, perturbed Riemann Hypothesis (experimental number theory), attractor distributions in extreme value theory, central limit theorem for stochastic processes, numerical stability, optimum color palettes, cluster processes on the sphere
Resources	28 exercises with solution expanding the theory and methods presented in the textbook, well documented source code and formulas to generate various deviates and simulations, simple recipes (with source code) to design your own data animations as MP4 videos – see ours on YouTube

This first volume deals with point processes in one and two dimensions, including spatial processes and clustering. The next volume in this series will cover other types of stochastic processes, such as Brownian-related and random, chaotic dynamical systems. The point process which is at the core of this textbook is called the Poisson-binomial process (not to be confused with a binomial nor a Poisson process) for reasons that will soon become apparent to the reader. Two extreme cases are the standard Poisson process, and fixed (non-random) points on a lattice. Everything in between is the most exciting part.

Target Audience

College-educated professionals with an analytical background (physics, economics, finance, machine learning, statistics, computer science, quant, mathematics, operations research, engineering, business intelligence), students enrolled in a quantitative curriculum, decision makers or managers working with data scientists, graduate students, researchers and college professors, will benefit the most from this textbook. The textbook is also intended to professionals interested in automated machine learning and artificial intelligence.

It includes many original exercises requiring out-of-the-box thinking, and offered with solution. Both students and college professors will find them very valuable. Most of these exercises are an extension of the core material. Also, a large number of internal and external references are immediately accessible with one click, throughout the textbook: they are highlighted respectively in red and blue in the text. The material is organized to facilitate the reading in random order as much as possible and to make navigation easy. It is written for busy readers.

The textbook includes full source code, in particular for simulations, image processing, and video generation. You don't need to be a programmer to understand the code. It is well documented and easy to read, even for people with little or no programming experience. Emphasis is on good coding practices. The goal is to help you quickly develop and implement your own machine learning applications from scratch, or use the ones offered in the textbook. The material also features professional-looking spreadsheets allowing you to perform interactive statistical tests and simulations in Excel alone, without statistical tables or any coding. The code, data sets, videos and spreadsheets are available on my GitHub repository.

The content in this textbook is frequently of graduate or post-graduate level and thus of interest to researchers. Yet the unusual style of the presentation makes it accessible to a large audience, including students and professionals with a modest analytic background (a standard course in statistics). It is my hope that it will

entice beginners and practitioners faced with data challenges, to explore and discover the beautiful and useful aspects of the theory, traditionally inaccessible to them due to jargon.

About the Author

Vincent Granville, PhD is a pioneering data scientist and machine learning expert, co-founder of Data Science Central (acquired by a publicly traded company in 2020), former VC-funded executive, author and patent owner. Vincent's past corporate experience includes Visa, Wells Fargo, eBay, NBC, Microsoft, CNET, InfoSpace and other Internet startup companies (one acquired by Google). Vincent is also a former post-doct from Cambridge University, and the National Institute of Statistical Sciences (NISS). He is currently publisher at DataShaping.com. He makes a living as an independent researcher working on stochastic processes, dynamical systems, experimental math and probabilistic number theory.

Vincent published in Journal of Number Theory, Journal of the Royal Statistical Society (Series B), and IEEE Transactions on Pattern Analysis and Machine Intelligence, among others. He is also the author of multiple books, including "Statistics: New Foundations, Toolbox, and Machine Learning Recipes", "Applied Stochastic Processes, Chaos Modeling, and Probabilistic Properties of Numeration Systems" with a combined reach of over 250,000, as well as "Becoming a Data Scientist" published by Wiley. For details, see my Google Scholar profile, here.

1 Poisson-binomial or Perturbed Lattice Process

I introduce here one of the simplest point process models. The purpose is to illustrate, in simple English, the theory of point processes using one of the most elementary and intuitive examples, keeping applications in mind. Many other point processes will be covered in the next sections, both in one and two dimensions. Key concepts, soon to be defined, include:

Category	Description	Book sections
Top parameters	Intensity λ – granularity of the process	4.4, 3.2.1
	Scaling factor s – quantifies point repulsion or mixing	3.1.1, 3.2.1
	Distribution F – location-scale family, with $F_s(x) = F(x/s)$	1.1, 3.2.2
Properties	Stationarity and ergodicity	1.4, 5.3
	Homogeneity and anisotropy	1.4.4
	Independent increments	1.4.3, 3.1.3
Core distributions	Interarrival times T	1.2, 4.2
	Nearest neighbor distances	3.4, 5.4
	Point count $N(B)$ in a set B	4.3, 5.3
	Point distribution (scattering, on a set B)	1.2
Type of process	Marked point process	1.5.1
	Cluster point process	2.1, 2.1.2
	Mixtures and interlacings (superimposed processes)	1.5.3, 3.4.3
Topology	Lattice space (index space divided by λ)	2.1, 4.7
	State space (where the points are located)	2.1
	Index space (hidden space of point indices: $\mathbb{Z}$ or $\mathbb{Z}^2$)	4.7, 2.2
Other concepts	Convergence to stationary Poisson point process	1.3, 4.6
	Boundary effects	3.5
	Dimension (of the state space)	1.2
	Model identifiability	3.3

I also present several probability distributions that are easy to sample from, including logistic, uniform, Laplace and Cauchy. I use them in the simulations. I also introduce new ones such as the exponential-binomial distribution (the distribution of interarrival times), and a new type of generalized logistic distribution. One of the core distributions is the Poisson-binomial with an infinite number of parameters. The Poisson-binomial process is named after that distribution, attached to the point count (a random variable) counting the number of points found in any given set. By analogy, the Poisson point process is named after the Poisson distribution for its point count. Poisson-binomial processes are also known as perturbed lattice point processes. Lattices, also called grids, are a core topic in this textbook, as well as nearest neighbors.

Poisson-binomial processes are different from both Poisson and binomial processes. However, as we shall see and prove, they converge to a Poisson process when a parameter called the scaling factor (closely related to the variance), tends to infinity. In recent years, there has been a considerable interest in perturbed lattice point processes, see [62, 68]. The Poisson-binomial process is lattice-based, and indeed, perturbed lattice point processes and Poisson-binomial processes are one and the same. The name "Poisson-binomial" has historical connotations and puts emphasis on its combinatorial nature, while "perturbed lattice" is more modern, putting emphasis on topological features and modern applications such as cellular networks.

Poisson-binomial point processes with small scaling factor s are good at modeling lattice-based structures such as crystals, exhibiting repulsion (also called inhibition) among the points, see Figure 3. They are also widely used in cellular networks, see references in Section 2.1.

1.1 Definitions

A point process is a (usually infinite) collection of points, sometimes called events in one dimension, randomly scattered over the real line (in one dimension), or over the entire space in higher dimensions. The points are denoted as X_k with $k \in \mathbb{Z}$ in one dimension, or (X_h, X_k) with $(h, k) \in \mathbb{Z}^2$ in two dimensions. The random variable X_k takes values in $\mathbb{R}$, known as the state space . In two dimensions, the state space is $\mathbb{R}^2$. The points are assumed to be independently distributed, though not identically distributed. Later in this textbook, it will be evident from the context when we are dealing with the one or two dimensional case.

In one dimension, the Poisson-binomial process is characterized by infinitely many points X_k, $k \in \mathbb{Z}$, each centered around k/λ, independently distributed with

$$P(X_k < x) = F\Big(\frac{x - k/\lambda}{s}\Big), \tag{1}$$

where

- The parameter $\lambda > 0$ is called the intensity; it represents the granularity of the process. The expected number of points in an interval of length $1/\lambda$ (in one dimension) or in a square of area $1/\lambda^2$ (in two dimensions), is equal to one. This generalizes to higher dimensions. The set $\mathbb{Z}/\lambda$ (or $\mathbb{Z}/\lambda \times \mathbb{Z}/\lambda$ in two dimensions) is the underlying lattice space of the process (also called the grid), while $\mathbb{Z}$ (or $\mathbb{Z}^2$ in two dimensions) is called the index space. The difference between state and lattice space is illustrated in Figure 22.
- The parameter $s > 0$ is the scaling factor, closely related to the variance. It determines the degree of mixing among the X_k's. When $s = 0$, $X_k = k/\lambda$ and the points are just the lattice points; there is no randomness. When s is infinite, the process becomes a classic stationary Poisson point process of intensity λ^d, where d is the dimension.
- The cumulative distribution function (CDF) $F(x)$ is continuous and belongs to a family of location-scale distributions [Wiki]. It is centered at the origin ($F(0) = \frac{1}{2}$), and symmetric ($F(x) = 1 - F(-x)$). Thus it has zero expectation, assuming the expectation exists. Its derivative, denoted as $f(x)$, is the density function; it is assumed to be unimodal (it has only one maximum), with the maximum value attained at $x = 0$.

In two dimensions, Formula (1) becomes

$$P[(X_h, Y_k) < (x, y)] = F\Big(\frac{x - h/\lambda}{s}\Big)F\Big(\frac{y - k/\lambda}{s}\Big). \tag{2}$$

Typical choices for F are

$$\text{Uniform: } F(x) = \frac{1}{2} + \frac{x}{2} \text{ if } -1 \le x \le 1, \text{ with } F(x) = 1 \text{ if } x > 1 \text{ and } F(x) = 0 \text{ if } x < -1$$
$$\text{Laplace: } F(x) = \frac{1}{2} + \frac{1}{2}\text{sgn}(x)(1 + \exp(-|x|))$$
$$\text{Logistic: } F(x) = \frac{1}{1 + \exp(-x)}$$
$$\text{Cauchy: } F(x) = \frac{1}{2} + \frac{1}{\pi}\arctan(x)$$

where $\text{sgn}(x)$ is the sign function [Wiki], with $\text{sgn}(0) = 0$. Despite the appearance, I use the standard form of these well-known distributions, when the location parameter is zero, and the scaling factor is $s = 1$. It looks unusual because I define them via their cumulative distribution function (CDF), rather than via the more familiar density function. Throughout this textbook, I use the CDF and its inverse (the quantile function) for simulation purposes.

F	Uniform	Logistic	Laplace	Cauchy	Gaussian
$\text{Var}[F_s]$	$s^2/3$	$\pi^2 s^2/3$	$2s^2$	∞	s^2

Table 1: Variance attached to F_s, as a function of s

Table 1 shows the relationship between s and the actual variance, for the distributions in question. I use the notation $F_s(x) = F(x/s)$ and $f_s(x)$ for its density, interchangeably throughout this textbook. Thus, $F(x) = F_1(x)$ and $f(x) = f_1(x)$. In orther words, F is the standardized version of F_s. In two dimensions, I use

$F(x,y) = F(x)F(y)$, assuming independence between the two coordinates: see Formula (2).

Remark: The parameter s is called the scaling factor because it is proportional to the variance of F_s, but visually speaking, it represents the amount of repulsion among the points of the process. See visual impact of a small s in Figure 3, and of a larger one in Figure 4.

1.2 Point Count and Interarrival Times

An immediate result is that $F_s(x - k/\lambda)$ is centered at k/λ. Also, if $s = 0$, then $X_k = k/\lambda$. If s is very small, X_k is very close to k/λ most of the time. But when s is large, the points X_k's are no longer ordered, and the larger s, the more randomly they are permutated (or shuffled, or mixed) on the real line.

Let $B = [a, b]$ be an interval on the real line, with $a < b$, and $p_k = P(X_k \in B)$. We have:

$$\begin{aligned} p_k &= F_s(b - t_k) - F_s(a - t_k) \\ &= F\Big(\frac{b - k/\lambda}{s}\Big) - F\Big(\frac{a - k\lambda}{s}\Big) \end{aligned} \tag{3}$$

This easily generalizes to two dimensions based on Formula (2). As a consequence, the integer-valued random variable $N(B)$ counting the number of points of the process in a set B, known as the counting measure [Wiki] or point count , has a Poisson-binomial distribution of parameters $p_k, k \in \mathbb{Z}$ [Wiki]. The only difference with a standard Poisson-binomial distribution is that here, we have infinitely many parameters (the p_k's). Basic properties of that distribution yield:

$$\mathrm{E}[N(B)] = \sum_{k=-\infty}^{\infty} p_k \tag{4}$$

$$\mathrm{Var}[N(B)] = \sum_{k=-\infty}^{\infty} p_k(1 - p_k) \tag{5}$$

$$P[N(B) = 0] = \prod_{k=-\infty}^{\infty} (1 - p_k) \tag{6}$$

$$P[N(B) = 1] = \Big(\sum_{k=-\infty}^{\infty} \frac{p_k}{1 - p_k} \Big) \cdot P[N(B) = 0] \tag{7}$$

It is more difficult, though possible, to obtain the higher moments $\mathrm{E}[N^r(B)]$ or $P[N(B) = r]$ in closed form if $r > 2$. This is due to the combinatorial nature of the Poisson-binomial distribution. But you can easily obtain approximated values using simulations.

Another fundamental, real-valued random variable, denoted as T or $T(\lambda, s)$, is the interarrival times between two successive points of the process, once the points are ordered on the real line. In two dimensions, it is replaced by the distance between a point of the process, and its nearest neighbor. Thus it satisfies (see Section 4.2) the following identity:

$$P(T > y) = P[N(B) = 0],$$

with $B =]X_0, X_0 + y]$, assuming it is measured at X_0 (the point of the process corresponding to $k = 0$). See Formula (38) for the distribution of T. In practice, this intractable exact formula is not used; instead it is approximated via simulations. Also, the point X_0 is not known, since the X_k's are in random order, and retrieving k knowing X_k is usually not possible. The indices (the k's) are hidden. However, see Section 4.7. The fundamental question is whether using X_0 or any X_k (say X_5), matters for the definition of T. This is discussed in Section 1.4 and illustrated in Table 4.

Finally, the point distribution is also of particular interest. In one dimension, this distribution can be derived from the distribution of interarrival times: the distance between two successive points. For instance, for a stationary Poisson process on the real line (that is, the intensity λ does not depend on the location), the points in any given set B are uniformly and independently distributed in B, and the interarrival times have an exponential distribution of expectation $1/\lambda$. However, for Poisson-binomial processes, there is no such simple result. If s is small, the points are more evenly spaced than the laws of pure randomness would dictate, see Figure 3. Indeed, the process is called repulsive: it looks as if the points behave like electrical charges, all of the same sign, exercising repulsive forces against each other. Despite this fact, the points are still independently distributed. To the contrary, cluster processes later investigated in this textbook, exhibit point attraction: it looks as if the points are attracted to each other.

Remark: A binomial process is defined as a finite set of points uniformly distributed over a domain B of finite area. Usually, the number of points is itself random, typically with a binomial distribution.

1.3 Limiting Distributions, Speed of Convergence

I prove in Theorem 4.5 that Poisson-binomial processes converge to ordinary Poisson processes. In this section, I illustrate the rate of convergence, both for the interarrival times and the point count in one dimension.

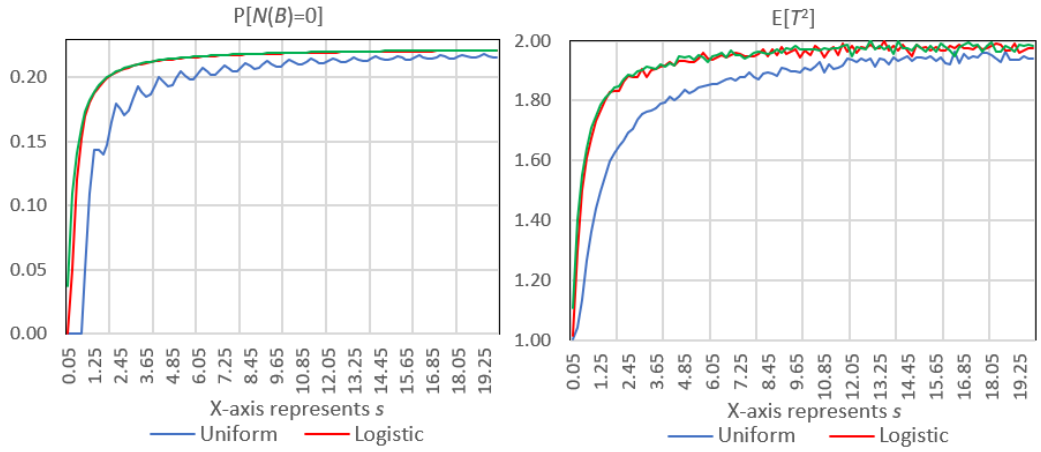

Figure 1: Convergence to stationary Poisson point process of intensity λ

In Figure 1, we used $\lambda = 1$ and $B = [-0.75, 0.75]$; $\mu(B) = 1.5$ is the length of B. The limiting values (combined with those of Table 3), as $s \to \infty$, are in agreement with $N(B)$'s moments converging to those of a Poisson distribution of expectation $\lambda\mu(B)$, and T's moments to those of an exponential distribution of expectation $1/\lambda$. In particular, it shows that $P[N(B) = 0] \to \exp[-\lambda\mu(B)]$ and $E[T^2] \to 2/\lambda$ as $s \to \infty$. These limiting distributions are features unique to stationary Poisson processes of intensity λ.

Figure 1 illustrates the speed of convergence of the Poisson-binomial process to the stationarity Poisson process of intensity λ, as $s \to \infty$. Further confirmation is provided by Table 3, and formally established by Theorem 4.5. Of course, when testing data, more than a few statistics are needed to determine whether you are dealing with a Poisson process or not. For a full test, compare the empirical moment generating function (the estimated $\mathrm{E}[T^r]$'s say for all $r \in [0, 3]$) or the empirical distribution of the interarrival times, with its theoretical limit (possibly obtained via simulations) corresponding to a Poisson process of intensity λ. The parameter λ can be estimated based on the data. See details in Section 3.

In Figure 1, the values of $\mathrm{E}[T^2]$ are more volatile than those of $P[N(B) = 0]$ because they were estimated via simulations; to the contrary, $P[N(B) = 0]$ was computed using the exact Formula (6), though truncated to 20,000 terms. The choice of a Cauchy or logistic distribution for F makes almost no difference. But a uniform F provides noticeably slower, more bumpy convergence. The Poisson approximation is already quite good with $s = 10$, and only improves as s increases. Note that in our example, $N(B) > 0$ if $s = 0$. This is because $X_k = k$ if $s = 0$; in particular, $X_0 = 0 \in B = [-0.75, 0.75]$. Indeed $N(B) > 0$ for all small enough s, and this effect is more pronounced (visible to the naked eye on the left plot, blue curve in Figure 1) if F is uniform. Likewise, $E[T^2] = 1$ if $s = 0$, as $T(\lambda, s) = \lambda$ if $s = 0$, and here $\lambda = 1$.

The results discussed here in one dimension easily generalize to higher dimensions. In that case B is a domain such as a circle or square, and T is the distance between a point of the process, and its nearest neighbor. The limit Poisson process is stationary with intensity λ^d, where d is the dimension.

1.4 Properties of Stochastic Point Processes

In this section, we review key features of point processes in general, applied to the Poisson-binomial process introduced in Section 1.1. A more comprehensive yet elementary presentation of some of these concepts (except those in Section 1.5), for the one-dimensional case and for traditional stochastic models (Markov chains, renewal, birth and death, queuing and Poisson processes), is found in any textbook on the subject, for instance in "Introduction to Stochastic Models" by R. Goodman [34].

1.4.1 Stationarity

There are various definitions of stationarity [Wiki] for point processes. The most common one is that the distribution of the point count $N(B)$ depends only on $\mu(B)$ (the length or area of B), but not on its location. The Poisson-binomial process is not stationary. Assuming $\lambda = 1$, if s is small enough, the point count distribution attached to (say) $B_1 = [0.3, 0.8]$ is different from that attached to $B_2 = [5.8, 6.3]$, despite both intervals having

the same length. This is obvious if $s = 0$: in that case $N(B_1) = 0$, and $N(B_2) = 1$. However, if $B_1 = [a, b]$ and $B_2 = [a+k/\lambda, b+k/\lambda]$, then $N(B_1)$ and $N(B_2)$ have the same distribution, regardless of $k \in \mathbb{Z}$; see Theorem 4.1 for a related result. So, knowing the theoretical distribution of $N([x, x + 1/\lambda])$ for each $0 \leq x < 1/\lambda$ is enough to know the distribution of $N(B)$ on any interval B. Since λ is unknown when dealing with actual data, it must be estimated using techniques described in Section 3. This generalizes to two dimensions, with the interval $N([x, x+1/\lambda])$ replaced by the square $N([x, x+1/\lambda]) \times N([y, y+1/\lambda])$, with $0 \leq x, y < 1/\lambda$. Statistical testing is discussed in [55], also available online, here.

The interarrival times T face fewer non-stationarity issues, as evidenced by Theorem 4.3, Table 4, and Exercise 5. It should be favored over the point count $N(B)$, when assessing whether your data fit with a Poisson-binomial, or a Poisson point process model. In particular, it does not depend, for practical purposes, on the choice of X_0 in the definition of T in Section 1.2. The definition could be changed using (say) X_5, or any other X_k instead of X_0, with no impact on the theoretical distribution.

1.4.2 Ergodicity

This brings us to the concept of ergodicity. It is heavily used in the active field of dynamical systems: see [15, 19, 41] and my book [36] available here. I will cover dynamical systems in details, in my upcoming book on this topic. For Poisson-binomial point processes, ergodicity means that you can estimate a quantity in two different ways:

- using one very long simulation of the process (a large n in our case),
- or using many small realizations of the process (small n), and averaging the statistics obtained in each simulation

Ergodicity means that both strategies, at the limit, lead to to same value. This is best illustrated with the estimation of $\mathrm{E}[T]$, or its higher moments. The expectation of the interarrival times T is estimated, in most of my simulations, as the average distance between a point X_k, and its nearest neighbor to the right, denoted as X'_k. It is computed as an average of $X'_k - X_k$ over $k = -n, \ldots, n$ with $n = 3 \times 10^4$, on a single realization of the process. The same methodology is used in the source code provided in Section 6. Likewise, $\mathrm{E}[T^2]$ is estimated as the average $(X'_k - X_k)^2$ in the same way.

Table 4 is an exception. There I used 10^4 realizations of a same Poisson-binomial process. In each realization I computed, among others, $T_0 = X'_0 - X_0$. This corresponds to the actual definition of T provided in Section 1.2. Then I averaged these T_0's over the 10^4 realizations to get an approximated value for T. It turns out that both methods lead to the same result. This is thanks to ergodicity, as far as T is concerned. I may as well have averaged $T_5 = X'_5 - X_5$ over the 10^4 realizations, and end up with the same result for $\mathrm{E}[T]$. Note that not all processes are ergodic. The difference between stationarity and ergodicity is further explained here.

1.4.3 Independent Increments

A one dimensional point process is said to have independent increments or independent interarrival times if the point counts $N(B_1), N(B_2)$ for any two non-overlapping time intervals B_1, B_2 are independent. It is shown in some textbooks, for instance [71] (available online here), that the only stationary renewal process with independent increments is the stationary Poisson process. The proof is simple, and based on the fact that the only distribution having the memoryless property, is the exponential one. Another definition of independent increments [Wiki] is based on the independence of the successive interarrival times. If combined with "identically distributed", it allows you, for Poisson-binomial process, to choose any arbitrary k to define the interarrival times as the random variable $T = X'_k - X_k$, where X'_k is the closest neighbor point of X_k, to the right on the real line. These two definitions of "independent increments" are not equivalent, since the first one based on point count, is measured at arbitrary locations, while the second one, based on interarrival times, is (in one dimension) the interdistance between actual points of the process. The point count and interarrival times are related by the identity $P(T > y) = P[N(B_0) = 0]$, where $B_0 =]X_0, X_0 + y]$, see Section 4.2.

An off-the-beaten-path test of independence is discussed in Section 3.1.3, precisely to assess the assumption of independent increments, on simulated data. A related concept is the memoryless property [Wiki].

1.4.4 Homogeneity

An ordinary Poisson point process (the limit, as $s \to \infty$, of a Poisson-binomial process) is said to be homogeneous if the intensity λ does not depend on the location. In the case of the Poisson process, homogeneity is equivalent to stationarity. Even for non-homogenous Poisson processes, the point count $N(B_1)$ and $N(B_2)$, attached to two disjoint sets $B1, B_2$, are independently (though not identically) distributed. This is not the case for Poisson-binomial processes, not even for those that are homogeneous.

Poisson-binomial processes investigated so far are homogeneous. I discuss non-homogeneous cases in Sections 1.5.3, 1.5.4 and 2.1. A non-homogeneous Poisson-binomial process is one where the intensity λ depends on the index k attached to a point X_k.

1.5 Transforming and Combining Multiple Point Processes

I discuss here a few types of point process operations, including translation, rotation, superimposition, mixtures of point processes and marked point processes. Cluster point processes, a particular type of superimposed processes, are not introduced in this section: they are treated in detail in Section 2.1 and 3.4. Another type of operation called thinning (see [59], available online here), is not described in this textbook.

1.5.1 Marked Point Process

In one dimension, a marked point process is similar to a couple of paired time series. It has two components: the base process, modeled here as a Poisson-binomial process, and a "mark" (a random variable) attached to each point. In one dimension, the base process typically represents time occurrences of events, and marks represent some feature attached to each event. The definition easily generalizes to any dimension.

An example is the highest yearly flood occurrences for a particular river, over a long time period, say 200 years. Due to yearly recurrence, a Poisson-binomial process where the intensity is $\lambda = 1$ and the time unit is a year, is better suited than a standard Poisson process. The marks measure the intensity of each maximum yearly flood. Another, 3-D example, is the position of atoms in a crystal. The marks may represent the type of atom.

Formally, in one dimension, a marked point process is a (usually infinite) set of points (X_k, Y_k) with $k \in \mathbb{Z}$. The definition easily generalizes to higher dimensions. Typically, the Y_k's (the marks) are independently distributed, and independently distributed from the underlying process (X_k). The underlying process can be a Poisson-binomial process.

1.5.2 Rotation, Stretching, Translation and Standardization

In two dimensions, rotating a Poisson-binomial process is equivalent to rotating its underlying lattice attached to the index space. Rotating the points has the same effect as rotating the lattice locations, because F (the distribution attached to the points) belongs to a family of location-scale distributions [Wiki]. For instance, a $\pi/4$ rotation will turn the square lattice into a centered-square lattice [Wiki], but it won't change the main properties of the point process. Both processes, the original one and the rotated one, may be indistinguishable for all practical purposes unless the scaling factor s is small, creating model identifiability [Wiki] issues. For instance, the theoretical correlation between the point coordinates (X_h, Y_k) or the underlying lattice point coordinates $(h/\lambda, k/\lambda)$, measured on all points, remains equal to zero after rotation, because the number of points is infinite (this may not be the case if you observe points through a small window, because of boundary effects). Thus, a Poisson-binomial process has a point distribution invariant under rotations, on a macro-scale. This property is called anisotropy [Wiki]. On a micro-scale, a few changes occur though: for instance the two-dimensional version of Theorem 4.1 no longer applies, and the distance between the projection of two neighbor points on the X or Y axis, shrinks after the rotation.

Applying a translation to the points of the process, or to the underlying lattice points, results in a shifted point process. It becomes interesting when multiple shifted processes, with different translation vectors, are combined together as in Section 1.5.3. Theorem 4.1 may not apply to the shifted process, though it can easily be adapted to handle this situation. One of the problems is to retrieve the underlying lattice space of the shifted process. This is useful for model fitting purposes, as it is easier to compare two processes once they have been standardized (after removing translations and rescaling). Estimation techniques to identify the shift are discussed in Section 3.4.

By a standardized Poisson-binomial point process, I mean one in its canonical form, with intensity $\lambda = 1$, scaling factor $s = 1$, and free of shifts or rotations. Once two processes are standardized, it is easier to compare them, assess if they are Poisson-binomial, or perform various machine learning procedures on observed data, such as testing, computing confidence intervals, cross-validation, or model fitting. In some way, this is similar to transforming and detrending time series to make them more amenable to statistical inference. There is also some analogy between the period or quasi-period of a time series, and the inverse of the intensity λ of a Poisson-binomial process: in fact, $1/\lambda$ is the fixed increment between the underlying lattice points in the lattice space, and can be viewed as the period of the process.

Finally, a two dimensional process is said to be stretched if a different intensity is used for each coordinate for all the points of the process. It turns the underlying square lattice space into a rectangular lattice, and the homogeneous process into a non-homogeneous one, because the intensity varies locally. Observed data points

can be standardized using the Mahalanobis transformation [Wiki], to remove stretching (so that variances are identical for both coordinates) and to decorrelate the two coordinates, when correlation is present.

1.5.3 Superimposition and Mixing

Here we are working with two-dimensional processes. When the points of m independent point processes with same distribution F and same index space $\mathbb{Z}^2$ are bundled together, we say that the processes are superimposed. These processes are no longer Poisson-binomial, see Exercise 14. Indeed, if the scaling factor s is small and $m > 1$ is not too small, they exhibit clustering around each lattice location in the lattice space. Also, the intensities or scaling factors of each individual point process may be different, and the resulting combined process may not be homogeneous. Superimposed point processes also called interlaced processes.

A mixture of m point processes, denoted as M, is defined as follows:

- We have m independent point processes $M_1, \ldots, M_m$ with same distribution F and same index space $\mathbb{Z}^2$,
- The intensity and scaling factor attached to M_i are denoted respectively as λ_i and s_i $(i = 1, \ldots, m)$,
- The points of M_i $(i = 1, \ldots, m)$ are denoted as (X_{ih}, Y_{ik}); the index space consists of the (h, k)'s,
- The point (X_h, Y_k) of the mixture process M is equal to (X_{ih}, Y_{ik}) with probability $\pi_i > 0$, $i = 1, \ldots, m$.

While mixing or superimposing Poisson-binomial processes seem like the same operation, which is true for stationary Poisson processes, in the case of Poisson-binomial processes, these are distinct operations resulting in significant differences when the scaling factors are very small (see Exercise 18). The difference is most striking when $s = 0$. In particular, superimposed processes are less random than mixtures. This is due to the discrete nature of the underlying lattice space. However, with larger scaling factors, the behavior of mixed and superimposed processes tend to be similar.

Several of the concepts discussed in Section 1.5 are illustrated in Figure 2, representing a realization of m superimposed shifted stretched Poisson-binomial processes, called m-interlacing. For each individual process M_i, $i = 1, \ldots, m$, the distribution attached to the point (X_{ih}, X_{ik}) (with $h, k \in \mathbb{Z}$) is

$$P(X_{ih} < x, Y_{ik} < y) = F\Big(\frac{x - \mu_i - h/\lambda}{s}\Big) F\Big(\frac{y - \mu_i' - k/\lambda'}{s}\Big), \quad i = 1, \ldots, m$$

This generalizes Formula (2). The parameters used for the model pictured in Figure 2 are:

- Number of superimposed processes: $m = 4$; each one displayed with a different color,
- Color: red for M_1, blue for M_2, orange for M_3, black for M_4,
- scaling factor: $s = 0$ (left plot) and $s = 5$ (right plot),
- Intensity: $\lambda = 1/3$ (X-axis) and $\lambda' = \sqrt{3}/3$ (Y-axis),
- Shift vector, X-coordinate: $\mu_1 = 0, \mu_2 = 1/2, \mu_3 = 2, \mu_4 = 3/2$,
- Shift vector, Y-coordinate: $\mu_1' = 0, \mu_2' = \sqrt{3}/2, \mu_3' = 0, \mu_4' = \sqrt{3}/2$,
- F distribution: standard centered logistic with zero mean and variance $\pi^2/3$.

For simulation purposes, the points (X_{ih}, Y_{ik}) of the i-th process M_i $(i = 1, \ldots, m)$, are generated as follows:

$$X_{ih} = \mu_i + \frac{h}{\lambda} + s \cdot \log\Big(\frac{U_{ih}}{1 - U_{ih}}\Big) \tag{8}$$

$$Y_{ik} = \mu_i' + \frac{k}{\lambda'} + s \cdot \log\Big(\frac{U_{ik}}{1 - U_{ik}}\Big) \tag{9}$$

where U_{ij} are uniformly and independently distributed on $[0, 1]$ and $-n \leq h, k \leq n$. I chose $n = 25$ in the simulation – a window much larger than that of Figure 2 – to avoid boundary effects in the picture. The boundary effect is sometimes called edge effect. The unobserved data points outside the window of observations, are referred to as censored data [Wiki]. Of course, in my simulations their locations and features (such as which process they belong to) are known by design. But in a real data set, they are truly missing or unobservable, and statistical inference must be adjusted accordingly [23]. See also Section 3.5.

I discuss Figure 2 in Section 1.5.4. A simple introduction to mixtures of ordinary Poisson processes is found on the Memming blog, here. In Section 3.4, I discuss statistical inference: detecting whether a realization of a point process is Poisson or not, and detecting the number of superimposed processes (similar to estimating the number of clusters in a cluster process, or the number of components in a mixture model). In Section 3.4.4, I introduce a black-box version of the elbow rule to detect the number of clusters, of mixture components, or the number of superimposed processes.

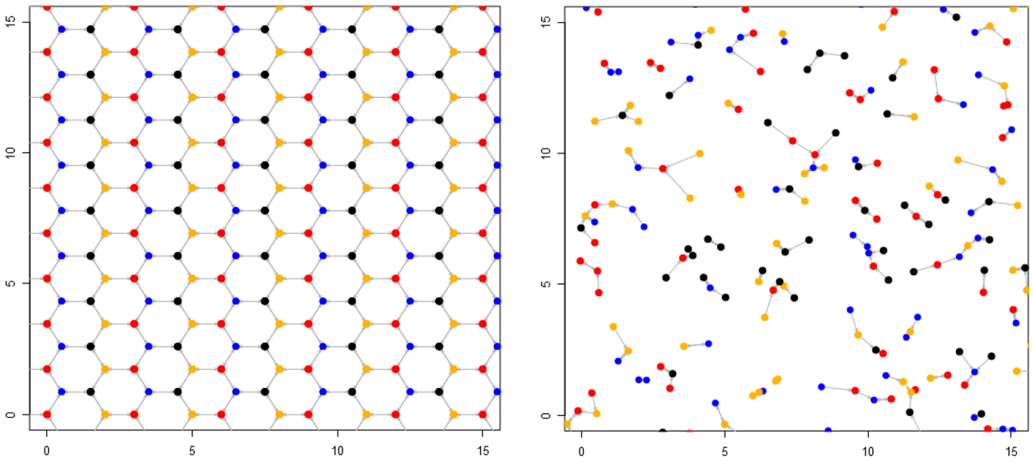

Figure 2: Four superimposed Poisson-binomial processes: $s = 0$ (left), $s = 5$ (right)

1.5.4 Hexagonal Lattice, Nearest Neighbors

Here I dive into the details of the processes discussed in Section 1.5.3. I also discuss Figure 2. The source code to produce Figure 2 is discussed in Sections 6.4 (nearest neighbor graph) and 6.7 (visualizations). Some elements of graph theory are discussed here, as well as visualization techniques.

Surprisingly, it is possible to produce a point process with a regular hexagonal lattice space using simple operations on a small number ($m = 4$) of square lattices: superimposition, stretching, and shifting. A stretched lattice is a square lattice turned into a rectangular lattice, by applying a multiplication factor to the X and/or Y coordinates. A shifted lattice is a lattice where the grid points have been shifted via a translation.

Each point of the process almost surely (with probability one) has exactly one nearest neighbor. However, when the scaling factor s is zero, this is no longer true. On the left plot in Figure 2, each point (also called vertex when $s = 0$) has exactly 3 nearest neighbors. This causes some challenges when plotting the case $s = 0$. The case $s > 0$ is easier to plot, using arrows pointing from any point to its unique nearest neighbor. I produced the arrows in question with the `arrow` function in R, see source code in Section 6.7, and online documentation here. A bidirectional arrow between points A and B means that B is a nearest neighbor of A, and A is a nearest neighbor of B. All arrows on the left plot in Figure 2 are bidirectional. Boundary effects are easily noticeable, as some arrows point to nearest neighbors outside the window. Four colors are used for the points, corresponding to the 4 shifted stretched Poisson-binomial processes used to generate the hexagon-based process. The color indicates which of these 4 process, a point is attached to.

The source code in Section 6.4 handles points with multiple nearest neighbors. It produces a list of all points with their nearest neighbors, using a hash table. A point with 3 nearest neighbors has 3 entries in that list: one for each nearest neighbor. A group of points that are all connected by arrows, is called a connected component [Wiki]. A path from a point of a connected component to another point of the same connected component, following arrows while ignoring their direction, is called a path in graph theory.

In my definition of connected component, the direction of the arrow does not matter: the underlying graph is considered undirected [Wiki]. An interesting problem is to study the size distribution, that is, the number of points per connected component, especially for standard Poisson processes. See Exercise 20. In graph theory, a point is called a vertex or node, and an arrow is called an edge. More about nearest neighbors is discussed in Exercises 18 and 19.

Finally, if you look at Figure 2, the left plot seems to have more points than the right plot. But they actually have roughly the same number of points. The plot on the right seems to be more sparse, because there are large areas with no points. But to compensate, there are areas where several points are in close proximity.

2 Applications

Applications of Poisson-binomial point processes (also called perturbed lattices point processes) are numerous. In particular, they are widely used in cellular and sensor network modeling and optimization. It also has applications in physics and crystal structures: see Figure 5 featuring man-made marble countertops. I provide many references in Section 2.1.

Here I focus on two-dimensional processes, to model lattice-based clustering. It is different from traditional clustering in two ways: clustering takes place around the vertices of the lattice space, and the number of clusters