Marked Point Processes in Discrete Time

Karl Sigman * Ward Whitt†

September 22, 2018

Abstract

Motivated by our interest in a periodic version of Little's law in discrete time, we present a general framework for stationary marked point processes in discrete time. It is built on a sequence $\{(t_j, k_j) : j \in \mathbb{Z}\}$ of times $t_j \in \mathbb{Z}$ and marks $k_j \in \mathbb{K}$, with batch arrivals (e.g., $t_j = t_{j+1}$) allowed. We start with a careful analysis of the sample paths. We show a topological equivalence between three different representations for a marked point process in discrete time. Then we develop discrete analogs of the familiar stationary stochastic constructs in continuous time: time-stationary and point-stationary random marked point processes, Palm distributions, inversion formulas and Campbell's theorem with an application to the derivation of a periodic-stationary Little's Law. Along the way we provide examples to illustrate interesting features of the discrete-time theory.

Keywords Marked Point Processes; Discrete-Time Stochastic Processes, Batch Arrival Processes, Queueing Theory; Periodic Stationarity.

AMS 2010 Subject Classification: Primary: 60G10; 60G55
Secondary: 60K25; 90B22

*Department of IEOR, Columbia University, Mudd Building, 500 West 120th Street, New York, NY 10027
†Department of IEOR, Columbia University, Mudd Building, 500 West 120th Street, New York, NY 10027

1 Introduction

This paper was motivated by [23], which established a periodic Little's Law (PLL) for a discrete-time periodic queueing model with batches. There is a substantial literature on discrete-time queues, largely motivated by communication and computer systems, as can be seen from the papers [9] and [7] and the recent book [6], but the PLL in discrete time was motivated by data analysis of a hospital emergency department in [22].

In [23] a sample-path version of PLL was first presented, and then was used (almost surely) to prove a periodic stationary version (Theorem 3 in [23]). We suspected that such a proof could be provided directly using stationary marked point process theory, but we did not immediately see an appropriate framework. We then decided to carefully put together such a framework that lends itself naturally to queueing and related applications; that is what is presented here. Our framework allows us, in particular, to give a direct proof of PLL in a periodic stationary setting using Palm distributions (Proposition 4.5), but we go beyond that initial goal.

We should hasten to say that marked point process theory is very well developed in continuous-time; e.g., see [10],[2], [20], [4] and [13], but it is not well developed in discrete time, let alone with batch arrivals allowed. We want the points $\{t_j : j \in \mathbb{Z}\}$ in time, $t_j \in \mathbb{Z}$, to be allowed to satisfy

$$\cdots \le t_{-2} \le t_{-1} \le 0 \le t_0 \le t_1 \le t_2 \le \cdots , \tag{1}$$

thus allowing batches, as opposed to the *simple* case

$$\cdots < t_{-2} < t_{-1} < 0 \le t_0 < t_1 < t_2 < \cdots \tag{2}$$

Even in continuous-time, many books on the subject state early on that they are assuming throughout that all point processes considered are simple. (An exception is [4]; they allow batches and even devote a chapter to it in the context of queueing models). Of course one can model a batch arrival process as a simple one in which the times at which the batches arrive forms a simple point process, and the batch size and labeling is placed in a mark (that is what is done in Chapter 7 devoted to batches in [4] for some queueing models); but this has a variety of disadvantages, including non-topological equivalence with the sequence approach above. Other approaches for batches have been developed (see in particular [15] and [16] where batch arrival processes are expressed as the superposition of a finite or countably infinite number of simple point processes and the Rate Conservation Law is used) but we did not find them as accessible or intuitive as we thought a framework should be. And of course in the literature there are scattered papers using batches in special queueing models and using special methods in their analysis (see for example [12], [21] and various examples in the books [27] (Pages 68, 267, 281, and 400 (Problem 8-5)), and [4] Chapter 7.

In spirit, our approach in the present paper for random marked point processes follows that in [20], in which stationary distributions are viewed as Cesàro averages; but here we deal with discrete time and allow batches. (See Remark 3.1 for further elaboration.)

One of the advantages in discrete time is that a point process can be defined by a sequence $\{x_n : n \in \mathbb{Z}\}$ where $x_n \in \mathbb{N}$ denotes the number of points at time n; since both $\mathbb{Z}$ and $\mathbb{N}$ are discrete hence endowed with the discrete topology, all subsets of $\mathbb{Z}$ and $\mathbb{N}$ are Borel measurable. In continuous-time, the analogue is treating a point process as a counting measure $N(A) =$ the number of points that fall in A for bounded Borel sets $A \subset \mathbb{R}$; one must deal with more complicated measure-theoretic details if they take that route. For a marked point process $\{(t_j, k_j)\}$ with mark space $\mathbb{K}$ this measure is defined on Borel sets $B \subset \mathbb{R} \times \mathbb{K}$, where $N(B)$ denotes the number of pairs (t_j, k_j) that fall in B. This measure approach then uses vague convergence of measures for its topology. The main advantage of the measure approach is that it easily generalizes to allow the time line $\mathbb{R}$ for point location to be replaced by general (non-ordered) spaces.

The measure approach in continuous-time is what most books use (exceptions include [4] and [20]): a *random* point process is defined as a random counting measure. In continuous time, with batches allowed, the measure approach and sequence approach are not topologically equivalent (they are if no batches are allowed; see for example Theorem D.3, Page 185 in [20]).

One of the results we prove (Proposition 2.2) is that in *discrete time* we can obtain a topological equivalence between $\{x_n\}$ and $\{t_j\}$, hence allowing us to freely move back and forth at leisure between the two representations, and with marks included.

The layout of our paper is as follows: In Section 2, we introduce the canonical space of marked point processes (e.g., non-random case), and in particular show that it forms a Polish space (e.g., a separable topological space that can be endowed with a complete metric). We give two other representations (inter-arrival time, counting sequence), and show a homeomorphism between all three. We also introduce point and time shift operators to prepare for the remaining chapters in which we work with random marked point processes utilizing ergodic theory. We include a summary of notation in Section 2.8 to help the reader. In Section 3, we introduce time and point random stationary marked point processes, introduce the Palm distribution and give inversion formulas between the time and point stationary versions. Several examples are given to gain intuition. In Section 4, we prove a Campbell's Theorem, and a Periodic Campbell's Theorem; applications are given to Little's Law and Periodic Little's Law. Finally, in Section 5 we briefly discuss the notion of Palm distributions in the non-ergodic case.

2 Point processes in discrete time

With $\mathbb{Z} = \{\cdots -2, -1, 0, 1, 2, \cdots\}$ denoting the integers, a discrete-time *point process (pp)* is a sequence of points $\psi \stackrel{\text{def}}{=} \{t_j\} = \{t_j : j \in \mathbb{Z}\}$ with the points in time $t_j \in \mathbb{Z}$ satisfying the following two conditions:

C1:

$$t_j \to +\infty \text{ and } t_{-j} \to -\infty \text{ as } j \to \infty. \tag{3}$$

C2: The points are non-decreasing and their labeling satisfies

$$\cdots \le t_{-2} \le t_{-1} \le 0 \le t_0 \le t_1 \le t_2 \le \cdots \tag{4}$$

with the proviso that $t_0 = 0$ if $t_{-1} = 0$.

The space of all point processes ψ is denoted by $\mathcal{M} \subset \mathbb{Z}^{\mathbb{Z}} = \prod_{j=-\infty}^{\infty} \mathbb{Z}$; a subspace of the product space. We endow $\mathbb{Z}$ with the discrete topology (e.g., all subsets of $\mathbb{Z}$ are open sets), and $\mathbb{Z}^{\mathbb{Z}}$ with the product topology and associated Borel $\sigma-$ algebra $\mathcal{B}\Big(\mathbb{Z}^{\mathbb{Z}}\Big)$. ($\mathcal{M}$ is not a closed subset of $\mathbb{Z}^{\mathbb{Z}}$.) $\mathcal{B}(\mathcal{M}) = \mathcal{M} \cap \mathcal{B}\Big(\mathbb{Z}^{\mathbb{Z}}\Big)$ are the Borel sets of $\mathcal{M}$.

Conditions C1 and C2 above ensure that there are an infinite number of points lying in both the positive time axis and the negative time axis, but that only a finite number of them can fall in any given time $n \in \mathbb{Z}$, and hence in any given bounded subset of time $A \subset \mathbb{Z}$. C_2 also ensures that *batches* are allowed, $t_j = t_{j+1}$, that is, one or more points can occur at any given time n but also ensures that the labeling of points at time $n = 0$ rules out such examples as $t_{-2} = t_{-1} = 0 < t_0 = 1$: *If there is a batch at the origin, then it must include the point* t_0.

Because of the product topology assumed, convergence of a sequence of pps, $\psi_m = \{t_{m,j}\} \in \mathcal{M},\ m \geq 1$, to a pp $\psi = \{t_j\} \in \mathcal{M}$, as $m \to \infty$, is thus equivalent to each coordinate converging; $\lim_{m\to\infty} t_{m,j} = t_j$ for each $j \in \mathbb{Z}$. Because $\mathbb{Z}$ is discrete, however, this is equivalent to: For each $j \in \mathbb{Z}$, there exists an $m = m_j \geq 1$ such that $t_{m,j} = t_j,\ m \geq m_j$.

2.1 Marked point processes

A marked point process (mpp) is a sequence of pairs, $\{(t_j, k_j) : j \in \mathbb{Z}\}$, where $\{t_j\} \in \mathcal{M}$ is a point process and $\{k_j\} \in \mathbb{K}^{\mathbb{Z}}$, where $\mathbb{K}$ is called the *mark space* and is assumed a complete separable metric space (CSMS) with corresponding Borel $\sigma-$ algebra $\mathcal{B}(\mathbb{K})$: Associated with each arrival point $t_j \in \mathbb{Z}$ is a *mark* $k_j \in \mathbb{K}$.

We denote the space of all mpps by

$$\mathcal{M}_K = \mathcal{M} \times \mathbb{K}^{\mathbb{Z}}, \tag{5}$$

a product space. Noting that a pp is a special case of a mpp when $\mathbb{K}$ is a set of one point $\{k\}$, *we will still use w.l.o.g. the notation $\psi \in \mathcal{M}_K$ to denote an mpp.*

Typical examples for a mark space are $\mathbb{K} = \mathbb{R}^d$, or $\mathbb{K} = \mathbb{N}$, but one can even allow $\mathbb{K} = \mathbb{R}^{\mathbb{Z}}$, so as to accommodate an entire infinite sequence as a mark. In many examples, the mark is a way of adding in some further information about the point it represents. A simple example in a queueing model context : t_j denotes the time of arrival of the j^{th} customer and $k_j = s_j$ denotes the service time of the customer, or $k_j = w_j$ denotes the sojourn time of the customer, or the pair $k_j = (s_j, w_j)$.

Remark 2.1 Since both $\mathcal{M}$ and $\mathbb{K}^{\mathbb{Z}}$ are separable, the Borel $\sigma-$ algebra $\mathcal{B}(\mathcal{M}_K)$ is equal to the product of the individual Borel $\sigma-$ algebras, $\mathcal{B}(\mathcal{M}) \times \mathcal{B}(\mathbb{K}^{\mathbb{Z}})$.

Remark 2.2 We are using a *two-sided* framework meaning that we allow an infinite past $\{t_j : j \leq 0\}$ in time as well as an infinite future $\{t_j : j \geq 0\}$ in time. A *one-sided* framework refers to the infinite future case only, and it can be considered on its own if need be.

2.2 Polish space framework

In this section, we provide a deeper analysis of the space $\mathcal{M}$ of point processes by showing that it is a *Polish space*, e.g., it is metrizable as a *complete* separable metric space (CSMS) for some metric. We then obtain as a Corollary (Corollary 2.1) that the space of all marked point processes $\mathcal{M}_K$ is thus Polish. This then allows one to apply standard weak convergence/tightness results/techniques to random marked point processes when needed, such as use of Prohorov's Theorem (see for example Section 11.6 , Theorem 11.6.1, p. 387 in [26] in the general context of stochastic processes.)

First observe that $\mathbb{Z}$ in the discrete topology (e.g., all subsets are open) is a CSMS, metrizable with the standard Euclidean metric it inherits as a subspace of $\mathbb{R}$, $|i-j|,\ i \in \mathbb{Z},\ j \in \mathbb{Z}$. (It is a closed subset of $\mathbb{R}$, and its subspace topology is precisely the discrete topology.) Now $\mathbb{Z}^{\mathbb{Z}}$ is a closed subset of the CSMS $\mathbb{R}^{Z}$, hence is a CSMS. That $\mathbb{R}^{Z}$ is a CSMS: see for example, Example 3, Page 265 in [17]. (More generally, the countable product of Polish spaces is Polish in the product topology.)

For example, for $\mathbb{Z}^{\mathbb{Z}}$, one could replace $|i-j|$ by $d(i,j) \stackrel{\text{def}}{=} \min\{|i-j|, 1\}$, its bounded equivalent, and then use the metric

$$D(\psi_1, \psi_2) = \sup_{j \in \mathbb{Z}} \frac{d(t_{1,j}, t_{2,j})}{1+|j|}.$$

Another well-known equivalent metric is based on,

$$D(\psi_1, \psi_2) = \sum_{j=1}^{\infty} \frac{|t_{1,j} - t_{2,j}|}{1 + |t_{1,j} - t_{2,j}|} 2^{-j},$$

while before hand bijectively mapping $\mathbb{Z}$ to $\{1, 2, 3, \ldots\}$ to re-index. Both these metrics generate the product topology and make $\mathbb{Z}^{\mathbb{Z}}$ complete.

We are now ready for

Proposition 2.1 *The space of point processes $\mathcal{M} \subset \mathbb{Z}^{\mathbb{Z}}$ is a Polish space. (In particular, it is a Borel measurable subset of $\mathbb{Z}^{\mathbb{Z}}$.)*

Proof : That $\mathcal{M}$ is a separable metric space follows since it is a subspace of a CSMS; the same metric can be used so that it is a metric space (and in general, a subspace of a separable metric space is separable). But under this metric $\mathcal{M}$ is not complete since $\mathcal{M}$ is not a closed subset of $\mathbb{Z}^{\mathbb{Z}}$. Thus it suffices to prove that the subset $\mathcal{M}$ is a G_δ subset of $\mathbb{Z}^{\mathbb{Z}}$, that is, it is of the form

$$\mathcal{M} = \cap_{i=1}^{\infty} B_i, \tag{6}$$

where each $B_i \subset \mathbb{Z}^{\mathbb{Z}}$ is an open set.

To this end define, for $i \geq 1$, subsets $B_i \subset \mathbb{Z}^{\mathbb{Z}}$ as those sequences $\{t_j\} \in \mathbb{Z}^{\mathbb{Z}}$ satisfying

1. $t_{-1} < 0$ if $t_0 > 0$.
2. $$t_{-i} \leq \cdots \leq t_{-1} \leq 0 \leq t_0 \leq t_i \leq \cdots \leq t_i.$$
3. There exists a $j > i$ and a $j' < -i$ such that $t_j > t_i$ and $t_{-j'} < t_{-i}$.

From Conditions C_1 and C_2 defining $\mathcal{M}$ it is immediate that

$$\mathcal{M} = \cap_{i=1}^{\infty} B_i.$$

We will now show that each B_i can be expressed as

$$B_i = B_i^+ \cap B_i^-,$$

where both B_i^+ and B_i^- are open sets, hence (finite intersection of open sets is always open) confirming (6) thus completing the proof.

For each subset B_i defined above ($i \geq 1$) we let B_i^+ be the union (indexed by $j \geq 1$) over all subsets $S_{i,j}^+ \subset \mathbb{Z}^{\mathbb{Z}}$ of the form

1. $t_{-1} < 0$ if $t_0 > 0$.

2.
$$t_{-i} \leq \cdots \leq t_{-1} \leq 0 \leq t_0 \leq t_i \leq \cdots \leq t_i.$$

3. $t_{i+j} > t_i$.

Each such subset $S^+_{i,j} \subset \mathbb{Z}^{\mathbb{Z}}$ is open because it is a finite dimensional subset (all finite dimensional subsets are open; discrete topology consequence). Hence being the union of open sets, B^+_i is open. Similarly B^-_i is the union over all open sets $S^-_{i,j}$, which are defined similarly to $S^+_{i,j}$ except with *3.* replaced by $t_{-i-j} < t_{-i}$.

■

In general, the finite or countable product of Polish spaces is Polish (in the product topology), and hence since the mark space $\mathbb{K}$ is assumed a CSMS, $\mathbb{K}^{\mathbb{Z}}$ is Polish. From Proposition 2.1, $\mathcal{M}$ is Polish and hence the product of the two, $\mathcal{M} \times \mathbb{K}^{\mathbb{Z}}$, is Polish too:

Corollary 2.1 *The space of all marked point processes*

$$\mathcal{M}_K = \mathcal{M} \times \mathbb{K}^{\mathbb{Z}}$$

is a Polish space in the product topology.

2.3 Random marked point processes Ψ

In the case of a random mpp, that is, when the points t_j and marks k_j are random variables we will denote it by upper case letters

$$\Psi = \{(T_j, K_j)\}. \tag{7}$$

A rmpp Ψ has sample paths in $\mathcal{M}_K$. We will denote the distribution of such a Ψ by $P(\Psi \in \cdot)$ defined on all Borel sets $\mathcal{E} \in \mathcal{B}(\mathcal{M}_K)$; $P(\Psi \in \mathcal{M}_K) = 1$.

2.4 The interarrival-time representation ϕ for a marked point process

Interarrival times $\mathbf{u} = \{u_j\} = \{u_j : j \in \mathbb{Z}\}$ of a pp $\psi \in \mathcal{M}$ are defined by $u_j \stackrel{\text{def}}{=} t_{j+1} - t_j,\ j \in \mathbb{Z}$, and thus

$$t_j = t_0 + u_0 + \cdots + u_{j-1},\ j \geq 1,\ t_{-j} = t_0 - (u_{-1} + \cdots + u_{-j+1}),\ j \geq 1. \tag{8}$$

The equality $u_j = 0$ means that both t_j and t_{j+1} occur at the same time (e.g., occur in the same batch).

We call $\phi = \phi(\psi) \stackrel{\text{def}}{=} \{t_0, \mathbf{u}\}$ the *interarrival-time representation* of a pp $\psi \in \mathcal{M}$.

As a consequence of (8), ψ and ϕ uniquely determine one another.

Such ϕ form a subspace $\mathcal{N} \subset \mathbb{N} \times \mathbb{N}^{\mathbb{Z}}$; the product space. Given any element $\{t_0, \mathbf{u}\}\} \in \mathbb{N} \times \mathbb{N}^{\mathbb{Z}}$, the only restrictions on it so as to uniquely define a pp $\psi \in \mathcal{M}$ using (8) is that $t_0 = 0$ if $t_0 - u_{-1} = 0$, and

$$\sum_{j=0}^{\infty} u_j = \infty,\ \sum_{j=1}^{\infty} u_{-j} = \infty.$$

That is what defines the subspace $\mathcal{N}$. We thus have a bijection mapping between $\mathcal{M}$ and $\mathcal{N}$; $\psi \mapsto \phi$. This bijection immediately extends to marked point processes, $\psi \in \mathcal{M}_K$, by adjoining in the marks $\{k_j\} \in \mathbb{K}^{\mathbb{Z}}$ yielding the product space $\mathcal{N}_K = \mathcal{N} \times \mathbb{K}^{\mathbb{Z}}$; then $\phi = \phi(\psi) \stackrel{\text{def}}{=} \{t_0, (\mathbf{u}, \mathbf{k})\} \in \mathcal{N}_K = \mathcal{N} \times \mathbb{K}^{\mathbb{Z}}$.

For random marked point processes Ψ, we will denote the interarrival-time representation by $\Phi = \{T_0, (\mathbf{U}, \mathbf{K})\}$.

2.5 The counting measure and counting sequence $\mathbf{x} = \{x_n\}$ for a point process

Given a pp $\psi \in \mathcal{M}$, if $A \subset \mathbb{Z}$ is a bounded subset, then we let

$$c(A) \stackrel{\text{def}}{=} \sum_{j \in \mathbb{Z}} I\{t_j \in A\}$$

denote the total number of points that fall in A; in particular we let $x_n \stackrel{\text{def}}{=} c(\{n\})$, denote the number of points that fall in time slot $n \in \mathbb{Z}$,

$$x_n = \sum_{j \in \mathbb{Z}} I\{t_j = n\}, \ n \in \mathbb{Z}. \tag{9}$$

Thus $c(\cdot)$ defines a measure on the subsets of $\mathbb{Z}$ called the *counting measure* of ψ, and the sequence $\mathbf{x} \stackrel{\text{def}}{=} \{x_n\} = \{x_n : n \in \mathbb{Z}\} \in \mathbb{N}^{\mathbb{Z}}$ is called the *counting sequence* of ψ. The space of all such counting sequences of pps $\psi \in \mathcal{M}$ is denoted by $\mathcal{X} \subset \mathbb{N}^{\mathbb{Z}}$, a proper subspace of the product space.

Let $c(n) \stackrel{\text{def}}{=} x_0 + \ldots + x_n, \ n \geq 0$, denote the cumulative number of points from time 0 up to and including time n; $\{c(n) : n \geq 0\}$ is called the forward *counting process*. In our framework, $c(0) = c(\{0\}) = x_0 > 0$ is possible; the number of points at the origin can be non-zero. Moreover,

$$\sum_{n=0}^{\infty} x_n = \infty, \ \sum_{n=0}^{\infty} x_{-n} = \infty, \tag{10}$$

since $t_j \to +\infty$ and $t_{-j} \to -\infty$ as $j \to \infty$ as required from C2.

When $x_n > 0$ we say that a *batch* occurred at time n. When $x_n \in \{0, 1\}, \ n \in \mathbb{Z}$, we say that the point process is *simple*; at most one arrival can occur at any time n.

We extend our counting measure $c(A)$ for a $A \subset \mathbb{Z}$ to include the marks of a marked point process so as to be a measure on $\mathcal{B}(\mathbb{Z} \times \mathbb{K})$ via

$$c(B) = \sum_{j \in \mathbb{Z}} I\{(t_j, k_j) \in B\}, \ B \in \mathcal{B}(\mathbb{Z} \times \mathbb{K}).$$

The measure $c(B)$ counts the number of pairs (t_j, k_j) that fall in the set $B \in \mathcal{B}(\mathbb{Z} \times \mathbb{K})$. For Borel sets of the form $B = A \times K$ where $A \subset \mathbb{Z}$ and $K \in \mathcal{B}(\mathbb{K})$,

$$c(A \times K) = \sum_{t_j \in A} I\{k_j \in K\},$$

the number of points in A that have marks falling in K. Then $c(A) = c(A \times \mathbb{K})$ ($K = \mathbb{K}$ (the entire mark space)) then gives back the counting measure as before of just the $\{t_j\}$.

For a random mpp Ψ, we denote the counting measure by $C(\cdot)$, and the counting sequence by $\mathbf{X} = \{X_n\}$.

2.6 A counting sequence representation $(\mathbf{x}, \hat{\jmath}_0)$ for point processes

A sequence $\mathbf{x} \in \mathcal{S}$ appears at first sight to be an equivalent way of defining a point process, in the sense that there should be a bijection between the two representations $\mathbf{x} = \{x_n\}$ and $\psi = \{t_j\}$. When $x_0 = 0$ this is true because from (4) the points themselves are then uniquely labeled

$$\cdots \leq t_{-2} \leq t_{-1} < 0 < t_0 \leq t_1 \leq t_2 \leq \cdots \tag{11}$$

$t_0 > 0$ is the first positive point, and $t_{-1} < 0$ is the first negative point; all else then follows. But when $x_0 > 0$ the points are not uniquely labeled. For example if $x_0 = 2$, we could have $t_{-1} < 0 = t_0 = t_1 < t_2$ or $t_{-2} < t_{-1} = 0 = t_0 < t_1$. Both possibilities satisfy (4). So whereas the mapping $\psi \longmapsto \mathbf{x}$ is unique, the inverse mapping is not.

The only problem we have to address then is how to keep track of the labeling of the points in x_0 when it is a batch, $x_0 > 0$, to ensure a unique mapping $\mathbf{x} \longmapsto \psi$.

Once that labeling is secure, the remaining points from $\{x_n : n \neq 0\}$ are uniquely labeled by (4). Note that if $x_0 > 0$ then in particular $t_0 = 0$ (recall condition C2). If $x_0 = 1$, then we are done, since then $t_{-1} < 0 = t_0 < t_1$ and all else follows from (4). So let us consider $x_0 > 1$.

We can write $x_0 = \imath_0 + \jmath_0$, where $\imath_0$ denotes the number of points in the batch, if any, labeled ≤ -1, and $\jmath_0 - 1$ denotes the number of points in the batch, if any, labeled ≥ 1. For example if $x_0 = 3$ and the 3 points are labeled $t_{-1} = t_0 = t_1 = 0$, then $\imath_0 = 1$ and $\jmath_0 = 2$. If the 3 points are labeled $t_{-2} = t_{-1} = t_0 = 0$, then $\imath_0 = 2$ and $\jmath_0 = 1$. Finally, if the 3 points are labeled $t_0 = t_1 = t_2$, then $\imath_0 = 0$ and $\jmath_0 = 3$. In general, $\jmath_0 \geq 1$ and $\imath_0 \geq 0$. When $x_0 > 0$, we view $\jmath_0$ as the number of points *in front of and including* t_0 in the batch, and $\imath_0$, the number of points *behind* t_0 in the batch. The idea is to imagine the batch as a bus with labeled seats. If $\imath_0 = b$ and $\jmath_0 = a$ then the $x_0 = b + a$ points are labeled $t_{-b} = t_{-b+1} = \cdots = 0 = t_0 = t_1 = \cdots = t_{a-1}$. The reader will notice the similarity of $\jmath_0$ and $\imath_0$ to the *forwards* and *backwards* recurrence time in (say) renewal theory; but here they do not represent time, they represent batch sizes/positions.

As our general solution to the labeling problem, we thus introduce

$$\hat{\jmath}_0 \stackrel{\text{def}}{=} \begin{cases} \jmath_0 & \text{if } x_0 > 0 \ , \\ 0 & \text{if } x_0 = 0, \end{cases} \tag{12}$$

Then we can consider a point process $\psi \in \mathcal{M}$ to be uniquely defined by $(\mathbf{x}, \hat{\jmath}_0)$. For example, if $x_0 = 2$ and $\jmath_0 = 2$, then $t_{-1} < 0 = t_0 = t_1 < t_2$ (e.g., $\imath_0 = 0$), whereas if $x_0 = 2$ and $\jmath_0 = 1$, then $t_{-2} < t_{-1} = 0 = t_0 < t_1$ (e.g., $\imath_0 = 1$). We denote by $\mathcal{S} \subset \mathcal{X} \times \mathbb{N} \subset \mathbb{N}^{\mathbb{Z}} \times \mathbb{N}$ the subspace of all $(\mathbf{x}, \hat{\jmath}_0)$ constructed from mpps $\psi \in \mathcal{M}$.

For a random point process Ψ, we use the notation $(\mathbf{X}, \hat{J}_0) = (\{X_n\}, \hat{J}_0)$, J_0, I_0 and so on for the counting sequence representation.

Remark 2.3 An important special case of $(\mathbf{x}, \hat{\jmath}_0) \in \mathcal{S}$ is when $\hat{\jmath}_0 = x_0$. This is the case when if $x_0 = a > 0$, then the points in the batch at time 0 are labeled $0 = t_0, \ldots, t_{a-1}$; t_0 is the first point in the batch, and $t_{-1} < 0$.

2.7 Extending the counting sequence representation to include marks

We now turn to extending the counting sequence representation to include marks, and obtain the subspace $\mathcal{S}_K$ of such marked representations.

We extend our $(\mathbf{x}, \hat{\jmath}_0) = (\{x_n\}, \hat{\jmath}_0) \in \mathcal{S}$ representation from Section 2.6 to $((\mathbf{x}, \overline{\mathbf{k}}), \hat{\jmath}_0) = (\{(x_n, \overline{k}_n)\}, \hat{\jmath}_0) \in \mathcal{S}_K$ by letting $\overline{k}_n = (k_1(n), \ldots, k_{x_n}(n))$ denote the list of associated marks of the x_n points (when $x_n > 0$). The labeling of the marks is automatically determined: if, for example $x_0 = b + a > 0$ with $\jmath_0 = a$ and $\imath_0 = x_0 - \jmath_0 = b$, then the $b + a$ marks are attached to $t_{-b}, \ldots, t_0 \ldots, t_{a-1}$ via $k_{-b} = k_1(0), \ \ldots \ k_{a-1} = k_{a+b}(0)$.

To make $\overline{k}_n$ mathematically rigorous, we introduce a 'graveyard state' $\Delta \notin \mathbb{K}$, to adjoin with the mark space $\mathbb{K}$; $\overline{\mathbb{K}} = \mathbb{K} \cup \{\Delta\}$. Letting d_K denote the standard bounded metric of $\mathbb{K}$ under its metric d (e.g., $d_K(x, y) = \min\{d(x, y), 1\}, \ x, y \in \mathbb{K}$), then we extend the metric to $\overline{\mathbb{K}}$ via $\overline{d}_K(x, y) = d_K(x, y), \ x, y \in \mathbb{K}, \ \overline{d}_K(x, \Delta) = 1, \ x \in \mathbb{K}, \ \overline{d}_K(\Delta, \Delta) = 0$. Then it is immediate that $\overline{\mathbb{K}}$ is a CSMS.

We then re-define $\overline{k}_n \stackrel{\text{def}}{=} (k_1(n), \ldots, k_{x_n}(n), \Delta, \Delta, \ldots) \in \overline{\mathbb{K}}^{\mathbb{N}_+}$, an infinite sequence in the product space $\prod_{i=1}^{\infty} \overline{K}$,under the product topology, where we define $\overline{k}_n = \boldsymbol{\Delta} \stackrel{\text{def}}{=} (\Delta, \Delta, \ldots) \in \overline{\mathbb{K}}^{\mathbb{N}_+}$, if $x_n = 0$.

Thus our space of all marked counting sequence representations, $((\mathbf{x}, \overline{\mathbf{k}}), \hat{\jmath}_0) = (\{(x_n, \overline{k}_n)\}, \hat{\jmath}_0)$, of mpps $\psi \in \mathcal{M}_K$ is a subspace

$$\mathcal{S}_K \subset (\mathbb{N} \times \overline{\mathbb{K}}^{\mathbb{N}_+})^{\mathbb{Z}} \times \mathbb{N}.$$

In its counting sequence representation, a random marked point process is denoted by $(\{(\mathbf{X}, \overline{K})\}, \hat{J}_0) = (\{(X_n, \overline{\mathbf{K}}_n)\}, \hat{J}_0)$.

2.8 Summary of notation

- $\mathcal{M}$: the space of all point processes $\psi = \{t_j\} = \{t_j : j \in \mathbb{Z}\}$. $\mathcal{M}_K = \mathcal{M} \times \mathbb{K}^{\mathbb{Z}}$: the space of all marked point processes $\psi = \{(t_j, k_j)\}$ with mark space $\mathbb{K}$. (Section 2.1)

- $\mathcal{N}$: the space of all point processes in the interarrival-time representation $\phi = \{t_0, \mathbf{u}\} = \{t_0, \{u_j\}\} = \{t_0, \{u_j : j \in \mathbb{Z}\}\}$, $u_j = t_{j+1} - t_j, \ j \in \mathbb{Z}$. $\mathcal{N}_K$ is the space of all marked point processes in the interarrival-time representation; $\phi = \{t_0, (\mathbf{u}, \mathbf{k})\}$. (Section 2.4)

- $\mathcal{S}$: the space of all point processes in the counting sequence representation $(\mathbf{x}, \hat{\jmath}_0)$, where $\mathbf{x} = \{x_n\} = \{x_n : n \in \mathbb{Z}\} \in \mathcal{X}$ is the counting sequence; $x_n = c(\{n\}) =$ the number of points that fall in the time-slot $n \in \mathbb{Z}$. $c(\cdot)$ is the counting measure of a pp; $\{c(n) : n \geq 0\}$ is the forward time counting process, $c(n) = x_0 + \cdots + x_n$.

  $\mathcal{S}_K$ is the space of all marked point processes in the counting sequence representation; $((\mathbf{x}, \overline{\mathbf{k}}), \hat{\jmath}_0) = (\{(x_n, \overline{k}_n)\}, \hat{\jmath}_0)$. (Sections 2.6 and 2.7)

- *Random* marked point process (rmpp) notation: $\Psi = \{(T_j, K_j)\}$, $\Phi = \{T_0, (\mathbf{U}, \mathbf{K})\} = \{T_0, \{(U_j, K_j)\}\}$, $(\mathbf{X}, \hat{J}_0) = (\{X_n\}, \hat{J}_0)$, $C(\cdot)$.

- $\mathcal{B}(\mathcal{T})$; Borel $\sigma-$ algebra of a topological space $\mathcal{T}$.

- $P(\Psi \in \cdot)$, $P(\Phi \in \cdot)$, $P((\{(\mathbf{X}, \overline{K})\}, \hat{J}_0) \in \cdot)$ corresponding distributions of a rmpp, on $\mathcal{B}(\mathcal{M}_K)$, $\mathcal{B}(\mathcal{N}_K)$, $\mathcal{B}(\mathcal{S}_K)$ respectively.

2.9 Topological equivalence of the three representations

Given that we have bijective mappings between $\mathcal{M}_K$ and $\mathcal{N}_K$ and $\mathcal{M}_K$ and $\mathcal{S}_K$, and we already have shown that under the product topology, $\mathcal{M}_K$ is a Polish space (Corollary 2.1), we can immediately conclude that $\mathcal{N}_K$ and $\mathcal{S}_K$ are Polish spaces too under the induced mapping topologies; all three are topologically equivalent.

This simply follows from the basic fact that if X, τ is a topological space and $f : X \longrightarrow Y$ is a bijective mapping onto a space Y, then $Y, f(\tau)$ is a topological space with topology $f(\tau) \stackrel{\text{def}}{=} \{f(A) : A \in \tau\}$. Moreover, if X, τ is Polish under a metric d_X, then $Y, f(\tau)$ is Polish under the metric $d_Y(y_1, y_2) \stackrel{\text{def}}{=} d_X(f^{-1}(y_1), f^{-1}(y_2))$. Summarizing:

Proposition 2.2 *All three representations for a marked point process,* $\psi = \{(t_j, k_j)\} \in \mathcal{M}_K, \phi = \{t_0, \{(u_j, k_j) : j \in \mathbb{Z}\}\} \in \mathcal{N}_K,\ (\{(x_n, \overline{k}_n)\}, \hat{J}_0) \in \mathcal{S}_K$ *are topologically equivalent;* $\mathcal{M}_K$*,* $\mathcal{N}_K$ *and* $\mathcal{S}_K$ *are homeomorphic Polish spaces.*

This allows us to conveniently work with any one of the three representations.

2.10 Shift mappings: the point and time shift operators

A point process can be shifted in several ways. One way is to shift to a specific point t_i and relabel that point as $t_0 = 0$ at time $n = 0$ (the present). All points labeled behind t_i become the past, and all points labeled in front of t_i become the future: Given a $\psi \in \mathcal{M}$, for each $i \in \mathbb{Z}$, we have a mapping

$$\theta_i : \mathcal{M} \longmapsto \mathcal{M}, \tag{13}$$

$\theta_i \psi \stackrel{\text{def}}{=} \{t_{i+j} - t_i : j \in \mathbb{Z}\}$, with the points denoted by $\{t_j(i) : j \in \mathbb{Z}\}$.

For any ψ, $\theta_i \psi$ always has a point at the origin, in particular $t_0(i) = 0$.

Note that $\theta_{i+1} = \theta_1 \circ \theta_i,\ i \geq 1$, so $\{\theta_i : i \geq 1\}$ is determined by just

$$\theta \stackrel{\text{def}}{=} \theta_1, \ \textit{the point-shift operator.} \tag{14}$$

Note that if $t_0 = 0$, then $\theta_0 \psi = \psi$, otherwise it moves t_0 to the origin.

For $\theta_i \psi$, all points in the same batch as t_i are relabeled as should be. For example if $i = 3$ and $t_2 = t_3 = t_4 = 6$ (batch of size 3 at time $n = 6$), we have $t_{-1}(3) = 0 = t_0(3) = t_1(3)$; the batch has been repositioned to time $n = 0$. If there is a batch at time $n = 0$, for example $t_{-2} = t_{-1} = t_0 = t_1$, a batch of size 4, then for $i = 1$, $t_{-3}(1) = t_{-2}(1) = t_{-1}(1) = t_0(1) = 0$; each batch position get shifted back by 1.

This point shift mapping translates immediately to a shift for the interevent-time representation $\phi \in \mathcal{N}$ in the same way

$$\theta_i : \mathcal{N} \longmapsto \mathcal{N},$$

$\theta_i \phi \stackrel{\text{def}}{=} \{0, \{u_{j+i} : j \in \mathbb{Z}\}\} = \{0, \{u_j(i) : j \in \mathbb{Z}\}\}$ is precisely the interevent-time representation for $\theta_i \psi$. For this reason, we use the same notation $\theta = \theta_1$ for the point-shift operator in both representations.

A second type of shift is with respect to *time*. Given a $(\mathbf{x}, \hat{\jmath}_0) \in \mathcal{S}$, for each time $m \in \mathbb{Z}$, we have a mapping

$$\zeta_m : \mathcal{S} \longmapsto \mathcal{S}, \tag{15}$$

$\zeta_m(\{x_n\},\hat{\jmath}_0) \stackrel{\text{def}}{=} (\{x_{m+n}\},x_m) = (\{x_n(m)\},x_0(m))$. ζ_m moves x_m to be the number of points at time $n=0$ and shifts the other x_n into the past and future appropriately. It also forces $t_{-1}<0$: If $x_0(m)=x_m>0$, then it's points (now moved to occur at time $n=0$) are labeled $t_0 \ldots t_{x_0-1}$.

As with point shifting, $\zeta_{m+1}=\zeta_1\circ\zeta_m,\ m\geq 1$, hence $\{\zeta_m : m\geq 1\}$ is determined by

$$\zeta \stackrel{\text{def}}{=} \zeta_1, \ \textit{the time-shift operator.} \tag{16}$$

The two operators θ and ζ are fundamental in our use of ergodic theory when we are dealing with random point processes.

Remark 2.4 When a point process is simple, then only the time shift mapping is needed, since then $\theta_i=\zeta_{t_i}$; shifting to time $n=t_i$ is equivalent to shifting to the i^{th} point.

The shift mappings θ_i and ζ_m extend immediately to when we have marks; for $i\in\mathbb{Z}$,

$$\theta_i\psi = \{(t_{i+j}-t_i,k_{i+j}) : j\in\mathbb{Z}\} \tag{17}$$

$$\theta_i\phi = \{0,\{(u_{i+j},k_{i+j} : j\in\mathbb{Z})\}. \tag{18}$$

For $m\in\mathbb{Z}$,

$$\zeta_m(\{(x_n,\overline{k}_n : n\in\mathbb{Z})\},\hat{J}_0) = (\{(x_{m+n},\overline{k}_{m+n} : n\in\mathbb{Z})\},x_m). \tag{19}$$

We retain the notation for the point and time shift operators, $\theta=\theta_1,\ \zeta=\zeta_1$.

3 Stationary ergodic framework: time and point stationarity

Here we focus on random marked point processes under *time* and *point* stationarity and ergodicity, and then introduce the Palm distribution, and prove inversion formulas. For a background on using ergodic theory in the context of stochastic processes, and point processes the reader is referred to [14], [8], [5], and [20].

Definition 3.1 *A random marked point process Ψ is called time-stationary if its counting sequence representation $(\{(X_n,\overline{\mathbf{K}}_n)\},\hat{J}_0)$ satisfies $\hat{J}_0=X_0$ and $\{(X_n,\overline{\mathbf{K}}_n) : n\in\mathbb{Z}\}$ is a stationary sequence. Equivalently, using the time-shift mappings:*

$$\zeta_m(\{(X_n,\overline{\mathbf{K}}_n)\},\hat{J}_0) = (\{(X_{m+n},\overline{\mathbf{K}}_{m+n})\},X_m)$$

has the same distribution as $(\{(X_n,\overline{\mathbf{K}}_n)\},\hat{J}_0)$ for all $m\in\mathbb{Z}$. It is called time-stationary and ergodic if the sequence $\{(X_n,\overline{\mathbf{K}}_n)\}$ is also ergodic (with respect to the time-shift operator: $\zeta=\zeta_1$).

We will denote a time-stationary marked point process by $\Psi^*=\{(T_j^*,K_j^*)\}$, or $\{(X_n^*,\overline{\mathbf{K}}_n^*)\}$, or $\Phi=\{T_0^*,\{U_j^*\}\}$. (Since under time-stationarity, $\hat{J}_0^*=X_0^*$, we express the counting sequence representation simply as $\{(X_n^*,\overline{\mathbf{K}}_n^*)\}$.)

The *arrival rate* of the point process is given by $\lambda=E(C^*(1))=E(X_0^*)$ because of the following (generalization of the Elementary Renewal Theorem):

Proposition 3.1 *If Ψ^* is time-stationary and ergodic, then*

$$\lim_{n\to\infty}\frac{C^*(n)}{n} = \lambda \stackrel{\text{def}}{=} E(X_0^*), \text{ wp1.} \tag{20}$$

and

$$\lim_{n\to\infty}\frac{n}{T_n^*} = \lambda, \text{ wp1.} \tag{21}$$

Proof : $C^*(n) = \sum_{i=0}^{n} X_i^*, \ n \geq 1$, so (20) is a direct application of the strong law of large numbers for stationary ergodic sequences derived from Birkoff's ergodic theorem applied to the stationary ergodic sequence $\{X_n^*\}$. Deriving (21): Observe that

$$C^*(T_n^* - 1) \leq n \leq C^*(T_n^*)$$

because $C^*(T_n^*)$ includes all the points in the batch of T_n^*, not just those in the batch that are labeled $\leq n$, and $C^*(T_n^* - 1)$ does not contain any points from the batch containing T_n^*. Dividing by T_n^* and using (20) on both the upper and lower bound yields the result since T_n^* is a subsequence of n as $T_n^* \to \infty$ and $n \to \infty$ wp1.

■

Definition 3.2 *A marked point process Ψ is called point-stationary if $\theta_i\Psi = \{(T_{i+j} - T_i, K_{i+j}) : j \in \mathbb{Z}\}$ has the same distribution as Ψ for all $i \in \mathbb{Z}$. This means that if we relabel point T_i as the origin, while retaining its mark K_i, the resulting point process has the same distribution regardless of what i we choose. Ψ is called point-stationary and ergodic if the sequence is also ergodic (with respect to the point-shift operator: $\theta = \theta_1$).*

Proposition 3.2 *A marked point process is point-stationary if and only if $P(T_0 = 1)$ and the interarrival time/mark sequence $\{(U_n, K_n) : n \in \mathbb{Z}\}$ is stationary. A marked point process is point-stationary and ergodic if and only if $P(T_0 = 1)$ and the interarrival time/mark sequence $\{(U_n, K_n) : n \in \mathbb{Z}\}$ is stationary and ergodic. (Recall that the same shift operator $\theta = \theta_1$ is used for both representations.)*

Proof : Because of the relationship (8) between interarrival times and points, the first result is immediate. The ergodicity equivalence is easily seen as follows: Ergodicity of Ψ is equivalent to

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} f(\theta_i\Psi) = E(f(\Psi)), \ wp1, \tag{22}$$

for all non-negative measurable functions f on $\mathcal{M}_K$.

Ergodicity of Φ is equivalent to

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} g(\theta_i\Phi) = E(g(\Phi)), \ wp1, \tag{23}$$

for all non-negative measurable functions g on $\mathcal{N}_K$.

But there is a one-to-one correspondence between non-negative measurable functions on $\mathcal{N}_K$ and non-negative measurable functions on $\mathcal{M}_K$: if $g = g(\phi)$ is a non-negative measurable function on $\mathcal{N}_K$, then since the mapping $\phi = \phi(\psi)$ is a homeomorphism (recall Proposition 2.2), we have that $g(\theta_i\phi) = g(\phi(\theta_i\psi)) = f(\theta_i\psi)$, where $f(\psi) = (g\circ\phi)(\psi)$ is a non-negative measurable function on $\mathcal{M}_K$. The equivalence goes the other way in the same manner. Thus (22) and (23) are equivalent.

■

Definition 3.3 *Given a random marked point process ψ, define (when it exists) a distribution $P(\psi^0 \in \cdot)$ via*

$$P(\Psi^0 \in \cdot) = \lim_{m\to\infty} \frac{1}{m}\sum_{i=1}^{m} P(\theta_i\Psi \in \cdot), \tag{24}$$

by which we mean that the convergence holds for all Borel sets $\mathrm{B} \in \mathcal{B}(\mathcal{M}_K)$, and defines a probability distribution on $\mathcal{B}(\mathcal{M}_K)$.

Theorem 3.1 *Given a time-stationary and ergodic marked point process Ψ^*, with $0 < \lambda = E(X_0^*) < \infty$ (the arrival rate), the distribution given in (24) exists, it is called the Palm distribution of Ψ^*, and is also given by*

$$P(\Psi^0 \in \cdot) = \lim_{m\to\infty} \frac{1}{m}\sum_{i=1}^{m} I\{\theta_i\Psi^* \in \cdot\},\ w.p.1, \tag{25}$$

and has representation

$$P(\Psi^0 \in \cdot) = \lambda^{-1}E\Big[\sum_{i=0}^{X_0^*-1} I\{\theta_i\Psi^* \in \cdot\}\Big], \tag{26}$$

where $\sum_{i=0}^{X_0^*-1}$ is defined to be 0 if $X_0^* = 0$. A marked point process Ψ^0 distributed as the Palm distribution is point-stationary, and is called a Palm version of Ψ^*. It satisfies $P(T_0^0 = 0) = 1$, and the interarrival time/mark sequence $\{(U_n^0, K_n^0) : n \in \mathbb{Z}\}$ is a stationary and ergodic sequence.

Proof : Taking expected values in (25) yields (24) by the bounded convergence theorem, so we will prove that (25) leads to (26). We will prove that by justifying re-writing the limit in (25) using the counting process $\{C^*(n)\}$ in lieu of m,

$$P(\Psi^0 \in \cdot) = \lim_{n\to\infty} \frac{1}{C^*(n)}\sum_{i=0}^{C^*(n)-1} I\{\theta_i\Psi^* \in \cdot\} = \lim_{n\to\infty}\Big(\frac{n}{C^*(n)}\Big)\frac{1}{n}\sum_{i=0}^{C^*(n)-1} I\{\theta_i\Psi^* \in \cdot\}. \tag{27}$$

From Proposition 3.1 and its proof we have

$$\Big(\frac{T_m^*-1}{m}\Big)\frac{1}{T_m^*-1}\sum_{i=0}^{C^*(T_m^*-1)} I\{\theta_i\Psi^* \in \cdot\} \le \frac{1}{m}\sum_{i=0}^{m} I\{\theta_i\Psi^* \in \cdot\} \le \Big(\frac{T_m^*}{m}\Big)\frac{1}{T_m^*}\sum_{i=0}^{C^*(T_m^*)} I\{\theta_i\Psi^* \in \cdot\}, \tag{28}$$

and $\lim_{n\to\infty}\frac{n}{C^*(n)} = \lambda^{-1}$, and $\lim_{m\to\infty}\frac{T_m^*}{m} = \lambda^{-1}$, wp1. Thus we see that it suffices to prove that wp1,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{C^*(n)-1} I\{\theta_i\Psi^* \in \cdot\} = E\Big[\sum_{i=0}^{X_0^*-1} I\{\theta_i\Psi^* \in \cdot\}\Big], \tag{29}$$

because if (29) does hold then it must hold along any subsequence of $n \to \infty$ including the subsequence T_m as $m \to \infty$; that is what we then can use in (28) (both the upper and lower bounds must have the same limit). We now establish (29).

Recall that $C^*(n) = X_0^* + \cdots + X_n^*$, so that

$$\sum_{i=0}^{C^*(n)-1} I\{\theta_i \Psi^* \in \cdot\} = \sum_{i=0}^{n} Y_j,$$

where

$$Y_0 = \sum_{i=0}^{X_0^*-1} I\{\theta_i \Psi^* \in \cdot\},$$

and

$$Y_j = \sum_{i=X_0^*+\cdots+X_{j-1}^*}^{X_0^*+\cdots+X_j^*-1} I\{\theta_i \Psi^* \in \cdot\},\ j \geq 1.$$

But $\{Y_j : j \geq 0\}$ forms a stationary ergodic sequence (see for example [20] Proposition 2.12 on Page 44), and so from Birkhoff's ergodic Theorem

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n} Y_j = E(Y_0),\ wp1;$$

(29) is established. (That the right-hand-side of (26) defines a probability distribution is easily verified; the monotone convergence theorem handles countably infinite additivity.) That Ψ^0 must be point-stationary (e.g., $\theta\Psi^0$ has the same distribution as Ψ^0) follows since $P(\theta_1\Psi^0 \in \cdot)$ is equivalent to replacing Ψ^* by $\theta_1\Psi^*$ before taking the limit in (24) which would become

$$P(\theta_1\Psi^0 \in \cdot) = \lim_{m\to\infty} \frac{1}{m} \sum_{i=0}^{m} P(\theta_{1+i}\Psi^* \in \cdot),$$

which of course has the same limit since the difference is asymptotically negligible to (24) in the limit.

Ergodicity is proved as follows: Suppose that $\mathrm{B} \in \mathcal{B}(\mathcal{M}_K)$ is an invariant event; e.g., $\theta^{-1}B = B$, hence $\theta_i^{-1}B = B,\ i \geq 1$. Then from (25), we have

$$P(\Psi^0 \in \mathrm{B}) = \lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} I\{\theta_i\Psi^* \in \mathrm{B}\} = I\{\Psi^* \in \mathrm{B}\},\ w.p.1, \tag{30}$$

which implies that $P(\Psi^0 \in \mathrm{B}) \in \{0, 1\}$; ergodicity.

■

Because $\sum_{i=0}^{X_0^*-1}$ is defined to be 0 if $X_0^* = 0$, we can re-write (26) as

$$P(\Psi^0 \in \cdot) = \lambda^{-1} E\Big[\sum_{i=0}^{X_0^*-1} I\{\theta_i\Psi^* \in \cdot\ ;\ X_0^* > 0\}\Big]. \tag{31}$$

When Ψ^* is simple, $X_0^* \in \{0,1\}$, and thus $\{X_0^* > 0\} = \{X_0^* = 1\} = \{T_0^* = 1\}$. Therefore $\lambda = E(X_0^*) = P(X_0^* > 0) = P(T_0^* = 0)$ and the summation inside (31) reduces to

$$I\{\theta_0\Psi^* \in \cdot\,;\ T_0^* = 0\} = I\{\Psi^* \in \cdot\,;\ T_0^* = 0\}.$$

Hence (31) collapses into

$$\lambda^{-1}P(\Psi^* \in \cdot\,;\ T_0^* = 0) = \lambda^{-1}P(\Psi^* \in \cdot \mid T_0^* = 0)P(T_0^* = 0) = P(\Psi^* \in \cdot \mid T_0^* = 0).$$

Summarizing:

Corollary 3.1 *If a time-stationary ergodic marked point process Ψ^* is simple, then*

$$P(\Psi^0 \in \cdot) = P(\Psi^* \in \cdot \mid T_0^* = 0);$$

i.e., the Palm distribution is the conditional distribution of Ψ^* given there is a point at the origin.

More generally (simple or not), let $B_0^* \stackrel{\text{def}}{=} (X_0^* \mid X_0^* > 0)$, denoting a true (time-stationary) batch size in liu of X_0^*.

$$P(B_0^* = k) = \frac{P(X_0^* = k)}{P(X_0^* > 0)},\ k \geq 1. \tag{32}$$

$$E(B_0^*) = \frac{E(X_0^*)}{P(X_0^* > 0)} = \frac{\lambda}{P(X_0^* > 0)}. \tag{33}$$

The following then is immediate from Theorem 3.1:

Corollary 3.2 *For a time-stationary ergodic point process Ψ^**

$$P(\Psi^0 \in \cdot) = \{E(B_0^*)\}^{-1}E\Big[\sum_{i=0}^{B_0^*-1} I\{\theta_i\Psi^* \in \cdot\ \}\Big]. \tag{34}$$

The above generalization of Corollary 3.1 says that to obtain the Palm distribution when there are batches, you first condition on there being a batch at the origin (e.g., $X_0^* > 0$) and then average over all X_0^* shifts $\theta_i\Psi^*,\ 0 \leq i \leq X_0^* - 1$.

Theorem 3.1 generalizes in a standard way to non-negative functions:

Proposition 3.3 *For any non-negative measurable function f,*

$$E(f(\Psi^0)) = \lim_{m\to\infty} \frac{1}{m}\sum_{i=1}^{m} f(\theta_i\Psi^*),\ \ w.p.1, \tag{35}$$

and has representation

$$E(f(\Psi^0)) = \lambda^{-1}E\Big[\sum_{i=0}^{X_0^*-1} f(\theta_i\Psi^*)\Big] = \{E(B_0^*)\}^{-1}E\Big[\sum_{i=0}^{B_0^*-1} f(\theta_i\Psi^*)\Big]. \tag{36}$$

Proof : Replacing $I\{\theta_i\Psi^* \in \cdot\}$ by $f(\theta_i\Psi^*)$ in the proof of Theorem 3.1 yields

$$\lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m} f(\theta_i\Psi^*) = \lambda^{-1}E\Big[\sum_{i=0}^{X_0^*-1} f(\theta_i\Psi^*)\Big],\ wp1.$$

It thus suffices to show that

$$\lambda^{-1}E\Big[\sum_{i=0}^{X_0^*-1} f(\theta_i\Psi^*)\Big] = E(f(\Psi^0)).$$

To this end, we already know it holds true for indicator functions $I\{\theta_i\Psi^* \in \cdot\}$ from (26); hence it also holds true for simple functions. Now we approximate a non-negative measurable function f by a monotone increasing sequence of simple functions and complete the result by use of the monotone convergence theorem.

■

As an immediate consequence of Proposition 3.3, with the function $f(\psi) = U_0$, we get wp1,

$$E(U_0^0) = \lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1} U_j^* = \lim_{n\to\infty}\frac{T_n^*}{n} = \lambda^{-1},$$

where we are using Proposition 3.1 for the last equality.

We include this and more in the following:

Proposition 3.4 *The Palm version Ψ^0 of a stationary ergodic marked point process Ψ^* with $\lambda = E(X_0^*)$ satisfies*

$$\frac{1}{E(U_0^0)} = \lambda \tag{37}$$

$$\lim_{n\to\infty}\frac{T_n^0}{n} = E(U_0^0) = \lambda^{-1},\ wp1. \tag{38}$$

$$\lim_{n\to\infty}\frac{C^0(n)}{n} = \lambda,\ wp1. \tag{39}$$

Proof : We already proved the first assertion. Because $\{U_j^0\}$ is stationary and ergodic, and $T_n^0 = \sum_{j=0}^{n-1} U_j^0,\ n \ge 1$, the second assertion follows directly by the strong law of large numbers for stationary and ergodic sequences via Birkoff's ergodic theorem, with the $= \lambda^{-1}$ part coming from the first assertion.

The third assertion is based on the following inequality

$$T^0_{C^0(n)-1} \le n \le T^0_{C^0(n)},$$

which implies that

$$\frac{T^0_{C^0(n)-1}}{C^0(n)} \le \frac{n}{C^0(n)} \le \frac{T^0_{C^0(n)}}{C^0(n)}.$$

Letting $n \to \infty$ while using our second assertion then yields that both the upper and lower bounds converge wp1 to λ^{-1} completing the result. ■

We now move on to deriving the probability distribution of $X_0^0 = I_0^0 + J_0^0$ which we know satisfies $P(X_0^0 > 0) = 1$ since by definition of the Palm distribution $P(T_0^0 = 0) = 1$; there is a batch at the origin. Recalling that $B_0^* = (X_0^* \mid X_0^* > 0)$ in (32) denotes a time-stationary batch size, the distribution of X_0^0 is the distribution of the batch size containing a randomly chosen point (over all points). As might be suspected, it has the stationary *spread* distribution of B_0^0 due to the inspection paradox (applied to batches) that a randomly chosen point is more likely to fall in a larger than usual batch because larger batches cover more points;

$$P(X_0^0 = k) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n} I\{T_j \text{ is in a batch of size } k\},\ k \geq 1.$$

Proposition 3.5 *The Palm version Ψ^0 of a stationary ergodic marked point process Ψ^* satisfies*

$$P(X_0^0 = k) = \frac{kP(B_0^* = k)}{E(B_0^*)},\ k \geq 1.$$

$$P(J_0^0 = k) = \frac{P(B_0^* \geq k)}{E(B_0^*)},\ k \geq 1.$$

$$P(I_0^0 = l, J_0^0 = k) = \frac{P(B_0^* = l + k)}{E(B_0^*)},\ l \geq 0,\ k \geq 1.$$

Proof : We use Proposition 3.3, with the functions $f_1(\Psi) = I\{X_0 = k\},\ f_2(\Psi) = I\{J_0 = k\}$ and $f_3(\Psi) = I\{I_0 = l,\ J_0 = k\}$. In these cases, we use

$$E(f(\Psi^0)) = \{E(B_0^*)\}^{-1} E\Big[\sum_{i=0}^{B_0^*-1} f(\theta_i \Psi^*)\Big].$$

Noting that $f_1(\theta_i\Psi^*) = I\{B_0^* = k\},\ 0 \leq i \leq B_0^* - 1$ (shifting within a batch keeps the same batch),

$$\begin{aligned} E(f_1(\Psi^0)) &= \{E(B_0^*)\}^{-1} E\Big[\sum_{j=0}^{B_0^*-1} I\{B_0^* = k\}\Big] \\ &= \{E(B_0^*)\}^{-1} E\Big[\sum_{i=0}^{k-1} I\{B_0^* = k\}\Big] \\ &= \{E(B_0^*)\}^{-1} kP(B_0^* = k). \end{aligned}$$

For dealing with f_3, let $g(\Psi^*) = (I_0^*, J_0^*)$. The labeling of the points of B_0^* is $t_0, \ldots, B_0^* - 1$, so $g(\theta_i\Psi^*) = (i, B_0^* - i),\ 0 \leq i \leq B_0^* - 1$. Thus the equality $f_3(\theta_i\Psi^*) = 1$ can only hold for at most one value of i and does so if and only if $B_0^* = l + k$ (in which case it happens for $i = l$). Thus

$$\begin{aligned} E(f_3(\Psi^0)) &= \{E(B_0^*)\}^{-1} E\Big[\sum_{i=0}^{B_0^*-1} I\{(i, B_0^* - i) = (l, k)\}\Big] \\ &= \{E(B_0^*)\}^{-1} E[I\{B_0^* = l + k\}] \\ &= \{E(B_0^*)\}^{-1} P(B_0^* = l + k). \end{aligned}$$

Similarly, for f_2, the equality $f_2(\theta_i\Psi^*) = I\{B_0^* - i = k\} = 1$ can only hold for at most one value of i within $0 \leq i \leq B_0^* - 1$, and does so if and only if $B_0^* \geq k$. Thus

$$\begin{aligned} E(f_2(\Psi^0)) &= \{E(B_0^*)\}^{-1} E\Big[\sum_{i=0}^{B_0^*-1} I\{B_0^* - i = k\}\Big] \\ &= \{E(B_0^*)\}^{-1} E[I\{B_0^* \geq k\}] \\ &= \{E(B_0^*)\}^{-1} P(B_0^* \geq k). \end{aligned}$$

■

Next we present a useful more general re-write of (26). For any time subset $A \subset \mathbb{Z}$, let $|A| = \sum_{n\in\mathbb{Z}} I\{n \in A\}$; the analog of the Lebesgue measure in continuous time.

Proposition 3.6 *For any* $0 < |A| < \infty$,

$$P(\Psi^0 \in \cdot) = \frac{E\Big[\sum_{T_j^* \in A} I\{\theta_j \Psi^* \in \cdot\}\Big]}{\lambda |A|}, \textit{i.e.}, \tag{40}$$

The Palm distribution is the expected value over all the point-shifts of points in any A $(0 < |A| < \infty)$ *of* Ψ^* *divided by the expected number of points in* A.

Proof : Because $0 < \lambda = E(X_0^*) < \infty$, note that (26) can be re-written as

$$P(\psi^0 \in \cdot) = \frac{E\Big[\sum_{T_j^* \in \{0\}} I\{\theta_j \Psi^* \in \cdot\}\Big]}{E(X_0^*)}. \tag{41}$$

Since $\{X_n^*\}$ is a stationary sequence, however, we can for any $n \in \mathbb{Z}$ also re-write the above as

$$P(\psi^0 \in \cdot) = \frac{E\Big[\sum_{T_j^* \in \{n\}} I\{\theta_j \Psi^* \in \cdot\}\Big]}{E(X_n^*)}. \tag{42}$$

For any $0 < |A| < \infty$, we have that $C^*(A) = \sum_{n\in A} X_n^*$ and hence $E(C^*(A)) = \lambda|A|$. Thus (40) follows from (42). ■

We can use (40) to derive

Proposition 3.7 *Given a time-stationary and ergodic marked point process* Ψ^*,

$$E(C^*(A \times \mathrm{K})) = \lambda |A| P(K_0^0 \in \mathrm{K}), \tag{43}$$

for all bounded $A \subset \mathbb{Z}$, *and measurable* $\mathrm{K} \subset \mathbb{K}$.

Proof : From (40) and Proposition 3.3 using $f(\psi) = I\{k_0 \in \mathrm{K}\}$ we have

$$P(K_0^0 \in \mathrm{K}) = \frac{E\Big[\sum_{T_j^* \in A} I\{K_j^* \in \mathrm{K}\}\Big]}{\lambda |A|} = \frac{E(C^*(A \times K))}{\lambda |A|}; \tag{44}$$

(43) follows. ■

Remark 3.1 We use Cesàro convergence (as in (24)) because the convergence holds by Birkoff's ergodic theorem without any further conditions, even if only stationarity without ergodicity holds. Sample-path averages converge with probability one, as well. Moreover, Birkoff's ergodic Theorem ensures that the Cesàro convergence holds *for all measurable sets* and from this it follows, in fact, that the convergence holds *uniformly over all measurable sets* and hence yields *Cesàro total variation convergence* and even *shift-coupling* (see Corollary 2.2, Theorem 2.2 and Corollary 2.3 on Pages 31-34 in [20].) In other words in our framework and with our objectives in the present paper, Cesàro convergence is the most natural one. On the other hand, if one wants to consider the much more general situation of the convergence in distribution of a sequence rmpps Ψ_n to a rmpp Ψ, as $n \to \infty$, other modes of convergence might be desired and be more useful, such as weak convergence, and would be analogous to the weak convergence of stochastic processes as in [26]. It is in that more general setting that notions of tightness and compactness play a fundamental role, and the Polish space condition is fundamental. In general, weak convergence and even stronger modes of convergence such as total variation convergence require much stricter conditions on the process even if the process is iid or regenerative (e.g., conditions such as non-lattice, aperiodic, spread-out, etc.), see Chapter VII in [1] for some examples.

Remark 3.2 While $\{X_n^*\}$ forms a stationary sequence (by definition), the same is not generally so for $\{X_n^0\}$. Recall, for example that $P(X_0^0 > 0) = 1$, while the same need not be so for the other $X_n^0,\ n \neq 0$.

3.1 Examples of stationary marked point processes

We will illustrate examples of Ψ^* and Ψ^0 by representing $\{X_n^* :\ n \in \mathbb{Z}\}$ as

$$\{X_n^*\} = \{\ldots, X_{-2}^*, X_{-1}^*, X_0^*, X_1^*, X_2^* \ldots\},$$

and $\{X_n^0\}$ as

$$\{X_n^0\} = \{\ldots, X_{-2}^0, X_{-1}^0, X_0^0, X_1^0, X_2^0 \ldots\}.$$

We will give examples when there are no marks involved. Unlike continuous time, Ψ^* can have points at the origin and this can allow for some interesting examples. Recall that since $\hat{J}_0^* = X_0^*$ by definition of time-stationarity, Ψ^* is completely determined by $\{X_n^*\}$. But in general, Ψ^0 is not completely determined by $\{X_n^0\}$ because $P(X_0^0 > 0) = 1$ and X_0^0 gets split into $X_0^0 = I_0^0 + J_0^0$. So we additionally need to determine J_0^0.

1. *Deterministic case (a).* Here we consider at first the case when $\{X_n^*\} = \{\ldots, 1, 1, 1, \ldots\}$. Then it is immediate that $\Psi^* = \Psi^0$ because

$$\{X_n^0\} = \{\ldots, 1, 1, 1, \ldots\}$$

as well, and $J_0^* = J_0^0 = 1$. This, it turns out, is the only example that can exist in which both the time and point stationary versions are identical. To see this, we know that since always $P(X_0^0 > 0) = 1$, it would have to hold too that $P(X_0^* > 0) = 1$. But if $X_0^* > 0$, then its points are always labeled $t_0, \ldots, t_{X_0^*-1}$, but when $X_0^0 > 0$ it splits X_0 into I_0 and J_0 with the I_0 points having negative labels and the J_0 points have labels ≥ 0. Whenever

$P(X_0^0 \geq 2) > 0$, it follows that $P(I_0^0 = 1,\ J_0^0 = X_0^0 - 1) > 0$, hence ruling out the condition $P(J_0^0 = X_0^0) = 1$ as would be required since $\hat{J}_0^* = X_0^*$ by definition. Our next example illustrates this difference with yet another deterministic case.

2. *Deterministic case (b).* Here we consider the case

$$\{X_n^*\} = \{\ldots, 2, 2, 2, \ldots\}.$$

It is immediate that

$$\{X_n^0\} = \{X_n^*\} = \{\ldots, 2, 2, 2, \ldots\}$$

because no matter what shift $\theta_i \Psi^*$ we use, the batch size covering any point is still of size 2. But Ψ^0 is not the same as Ψ^*: half of the shifts $\theta_i \Psi^*$ split the batch of size 2 at the origin into $T_0^*(i) = T_1(i) = 0$ and half split it into $T_{-1}^*(i) = T_0(i) = 0$. We have $P(J_0^0 = 1) = P(J_0^0 = 2) = 1/2$. So while Ψ^* is deterministic, Ψ^0 is not.

3. *iid case:*

$$\{X_n^*\} = \{\ldots, X_{-2}^*, X_{-1}^*, X_0^*, X_1^*, X_2^* \ldots\},$$

where $\{X_n^* : n \in \mathbb{Z}\}$ is any iid sequence of non-negative rvs with $0 < E(X_0^*) < \infty$. Then

$$\{X_n^0\} = \{\ldots, X_{-2}^*, X_{-1}^*, X_0^0, X_1^*, X_2^* \ldots\},$$

where X_0^0 and J_0^0, independent of the iid $\{X_n^0 : n \neq 0\}$, are distributed as in Proposition 3.5 by jointly constructing a copy of (I_0^0, J_0^0) and using $X_0^0 = I_0^0 + J_0^0$.

4. *Bernoulli(p) iid case:* Here we consider a simple point process (e.g., only at most one arrival in any given time slot) that is a very special but important example in applications of the above Example 3 iid case because it serves as the discrete-time analog of a Poisson process. We take $\{X_n^*\}$ as iid with a Bernoulli(p) distribution, $0 < p < 1$. $\lambda = p = E(X_0^*)$. Since $\{X_n^*\}$ is iid and the point process is simple, we can use Corollary 3.1 which instructs us to place a point at the origin ($P(X_0^0 = 1) = 1$) to get $\{X_n^0\}$; $P(T_0^0 = 0) = 1$:

$$\{X_n^0\} = \{\ldots, X_{-2}^*, X_{-1}^*, 1, X_1^*, X_2^* \ldots\},$$

and of course $J_0^0 = X_0^0 = 1$. Notice that $P(T_0^* = 0) = P(X_0^* = 1) = p$.

The interarrival times $\{U_n^0\}$ are iid with a geometric distribution with success probability p.

5. *Markov chain case:*

We start with an irreducible positive recurrent discrete-time discrete state space Markov chain $\{X_n : n \geq 0\}$ on the non-negative integers, and transition matrix $P = (P_{i,j})$ and stationary distribution $\pi = \{\pi_j : j \geq 0\}$. We assume that $0 < E_\pi(X_0) < \infty$; π has finite and non-zero mean. By starting off the chain with X_0 distributed as π, we can obtain a 1-sided stationary version $\{X_n^* : n \geq 0\}$. At this point we have two ways to obtain a 2-sided version: One is to use Kolmogorov's extension theorem which assures the existence of such an extension for any 1-sided stationary sequence. The other is to recall that since the chain is positive recurrent with stationary distribution π, we can explicitly give the transition matrix for its time reversal as

$$P^{(r)}_{i,j} = P(X^*_{-1} = j \mid X^*_0 = i) = \frac{\pi_j}{\pi_i} P_{i,j}, \ i, j \geq 0.$$

Thus starting with $\{X^*_n : n \geq 0\}$, and using X^*_0, we then can continue backwards in time to construct $\{X^*_n : n < 0\}$ by using $P^{(r)} = (P^{(r)}_{i,j})$. This yields

$$\{X^*_n\} = \{\ldots, X^*_{-2}, X^*_{-1}, X^*_0, X^*_1, X^*_2 \ldots\}.$$

Then

$$\{X^0_n\} = \{\ldots, X^0_{-2}, X^0_{-1}, X^0_0, X^0_1, X^0_2 \ldots\},$$

where X^0_0 and J^0_0 are distributed jointly as in Proposition 3.5, $\{X^0_n : n \geq 0\}$ is constructed sequentially using $P = (P_{i,j})$, and $\{X^0_n : n < 0\}$ uses $P^{(r)} = (P^{(r)}_{i,j})$, both sides starting with X^0_0.

6. *Cyclic deterministic example:* Starting with $\{X_n\} = \{\ldots, 1, 0, 2, 1, 0, 2, 1, 0, 2, \ldots\}$, we have cycles of the form $\{1, 0, 2\}$ repeating forever. This is actually a very special case of a Markov chain; $P_{1,0} = P_{0,2} = P_{2,1} = 1$, but its analysis here yields nice intuition. The time stationary version is a 1/3 mixture: $P(X^*_0 = i) = 1/3, \ i \in \{1, 0, 2\}$ which then determines the entire sequence. The idea is that 1/3 of all time begins with an X_n of size 1, 2, or 3 within a cycle.

$$\{X^*_n\} = \begin{cases} \{\ldots, 1, 0, 2, \ 1 = X^*_0 \ , 0, 2, 1, 0, 2, \ldots\} & \text{wp } 1/3 \ , \\ \{\ldots, 0, 2, 1, \ 0 = X^*_0 \ , 2, 1, 0, 2, \ldots\} & \text{wp } 1/3, \\ \{\ldots, 2, 1, 0, \ 2 = X^*_0 \ , 1, 0, 2, \ldots\} & \text{wp } 1/3. \end{cases}$$

Note that $\lambda = (1/3)(1 + 0 + 2) = 1$.

To determine $\{X^0_n\}$, we first need only consider lining up the $X_n > 0$ (the batches) to obtain $\{\ldots, 1, 2, 1, 2, \ldots\}$ and randomly select a point over all batches. 2/3 of the points sit in an $X_n = 2$ and 1/3 sit in an $X_n = 1$. Thus we obtain

$$\{X^0_n\} = \begin{cases} \{\ldots, 2, 1, 0, \ 2 = X^0_0 \ , 1, 0, 2, \ldots\} & \text{wp } 2/3 \\ \{\ldots, 1, 0, 2, \ 1 = X^0_0 \ , 0, 2, 1, 0, 2, \ldots\} & \text{wp } 1/3 \ . \end{cases}$$

Given the 2/3 case, $P((I^0_0, J^0_0) = (0, 2) = 1/2, \ P((I^0_0, J^0_0) = (1, 1) = 1/2$, while given the 1/3 case $P((I^0_0, J^0_0) = (0, 1) = 1$. Thus Ψ^0 is completely determined by the 1/3 mixture of $P(X^0_0 = 2, \ J^0_0 = 2) = P(X^0_0 = 2, \ J^0_0 = 1) = 1/3, \ P(X^0_0 = 1, \ J^0_0 = 1) = 1/3$.

This illustrates that for a cyclic deterministic point process, Ψ^0 is completely determined by the pair (X^0_0, J^0_0).

7. *Regenerative process case:* Suppose that $\{X_n\}$ is a positive recurrent regenerative process. Example 6 above is a very special case of this, and so is Example 5 (a Markov chain regenerates each time it visits a given fixed state i.) We allow general iid cycles of non-negative

random variables, $\mathcal{C}_0 = \{\{X_0, X_1, \ldots, X_{\tau_1-1}\},\ \tau_1\}$, $\mathcal{C}_1 = \{\{X_{\tau_1}, X_{\tau_1+1}, \ldots, X_{\tau_1+\tau_2-1}\},\ \tau_2\}$ and so on, where $\{\tau_m : m \geq 1\}$ forms a discrete-time renewal process with $0 < E(\tau_1) < \infty$. We attach another iid such sequence identically distributed of cycles from time past, $\{\mathcal{C}_m : m \leq -1\}$, yielding iid cycles $\{\mathcal{C}_m : m \in \mathbb{Z}\}$ and hence our two sided $\{X_n\}$. From the Renewal Reward Theorem, the arrival rate is given by

$$\lambda = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} X_m = \frac{E\left[\sum_{m=0}^{\tau_1-1} X_m\right]}{E(\tau_1)},\ wp1,$$

and we assume that $0 < \lambda < \infty$.

A time-stationary version $\{X_n^*\}$ is given by standard regenerative process theory in which the initial cycle $\mathcal{C}_0^*$ is a delayed cycle different in distribution from the original $\mathcal{C}_0$. It contains some X_n with $n \leq 0$ and some X_n with $n > 0$. It is a cycle that covers a randomly selected X_n way out in the future which is then labeled as X_0^*. From the inspection paradox applied to the cycle lengths, the cycle length τ_0^* of $\mathcal{C}_0^*$ has the spread distribution of τ_1:

$$P(\tau_0^* = k) = \frac{kP(\tau_1 = k)}{E(\tau_1)},\ k \geq 1.$$

Regenerative processes $\mathbf{X} = \{X_n : n \in \mathbb{Z}\}$ are ergodic with respect to the shift-operator $\theta = \theta_1$,

$$\theta_m \mathbf{X} = \{X_{m+n} : n \in \mathbb{Z}\} = \{X_n(m) : n \in \mathbb{Z}\},\ m \in \mathbb{Z}.$$

Letting $\mathcal{C}_0(m) = \mathcal{C}_0(\theta_m \mathbf{X})$ denote the initial cycle of $\theta_m \mathbf{X}$, the cycle containing $X_0(m)$, we have

$$P(\mathcal{C}_0^* \in \cdot) = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} I\{C_0(m) \in \cdot\} = \frac{E\left[\sum_{m=0}^{\tau_1-1} I\{C_0(m) \in \cdot\}\right]}{E(\tau_1)},\ wp1.$$

Thus starting with the iid cycles $\{\mathcal{C}_m : m \in \mathbb{Z}\}$, and independently replacing $\mathcal{C}_0$ with a copy of $\mathcal{C}_0^*$ yields time-stationary $\{X_n^*\}$, i.e.,

$$\{X_n^*\} = \{\cdots \mathcal{C}_{-2}, \mathcal{C}_{-1}, \mathcal{C}_0^*, \mathcal{C}_1, \mathcal{C}_2 \cdots\}.$$

Similarly, to obtain $\{X_n^0\}$, we need to derive the appropriate initial delay cycle $\mathcal{C}_0^0$, independent of the iid others, $\{\mathcal{C}_m : m \neq 0\}$, to obtain the desired

$$\{X_n^0\} = \{\cdots \mathcal{C}_{-2}, \mathcal{C}_{-1}, \mathcal{C}_0^0, \mathcal{C}_1, \mathcal{C}_2 \cdots\}.$$

Thus $\mathcal{C}_0^0$ represents a cycle that covers a randomly selected *point* t_j way out in the future.

3.2 Palm inversion

Recalling the time-shift operator ζ, from (15), one can retrieve back time-stationary ergodic Ψ^* from point-stationary ergodic Ψ^0 via *time* averaging (versus point averaging). Because the interarrival times $\{U_j^0\}$ form a stationary ergodic sequence, the inversion just says that the time-average is the expected value over a "cycle"(interarrival time) divided by an expected cycle length $E(U_0^0) = \lambda^{-1}$, just as in the famous renewal reward theorem in the iid case.

Theorem 3.2 (Palm inversion formula)

$$P(\Psi^* \in \cdot) = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} P(\zeta_m \Psi^0 \in \cdot) = \lambda E\Big[\sum_{m=0}^{U_0^0-1} I\{\zeta_m \Psi^0 \in \cdot\} I\{U_0^0 \geq 1\}\Big]. \tag{45}$$

$$P(\Psi^* \in \cdot) = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} I\{\zeta_m \Psi^0 \in \cdot\}, \ wp1. \tag{46}$$

Proof : We use the counting sequence representation $(X_n^0, \overline{\mathbf{K}}_n^0)$ for Ψ^0. (Since ζ_m maps $\hat{J}_0^0$ to X_m^0 for all m, by definition, we need not include it; there are no labeling issues of the points once Ψ^0 is shifted in time by ζ_m.) As used in the proof of Proposition 3.4 we have the inequality

$$T^0_{C^0(n)-1} \leq n \leq T^0_{C^0(n)},$$

which yields

$$\frac{1}{n} \sum_{m=0}^{T^0_{C^0(n)-1}} I\{\zeta_m \Psi^0 \in \cdot\} \leq \frac{1}{n} \sum_{m=0}^{n} I\{\zeta_m \Psi^0 \in \cdot\} \leq \frac{1}{n} \sum_{m=0}^{T^0_{C^0(n)}} I\{\zeta_m \Psi^0 \in \cdot\}. \tag{47}$$

We will now show that the right-hand-side of (47) (hence the left-hand side too) converges wp1 to the right-hand side of (45). For then this proves that the right-hand-side of (46) converges to the right-hand-side of (47); taking expected values then in (46) using the bounded convergence theorem then finishes the result.

To this end, recalling that $T_n^0 = \sum_{i=0}^{n-1} U_i^0, \ n \geq 1$, we can rewrite a sum over time as a sum over stationary ergodic "cycle lengths" U_i^0:

$$\sum_{m=0}^{T_n^0-1} I\{\zeta_m \Psi^0 \in \cdot\} = \sum_{i=0}^{n-1} Y_i,$$

where

$$Y_i = \sum_{m=T_i^0}^{T_{i+1}^0-1} I\{\zeta_m \Psi^0 \in \cdot\} I\{U_i^0 \geq 1\}, \ i \geq 0.$$

Since Ψ^0 is point-stationary and ergodic with respect to the point-shifts $\theta_i, \ i \geq 1$, the $\{Y_i : i \geq 0\}$ form a stationary ergodic sequence.

Thus from Birkoff's ergodic theorem,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{T_n^0-1} I\{\zeta_m \Psi^0 \in \cdot\} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} Y_i \tag{48}$$

$$= E(Y_0) \tag{49}$$

$$= E\Big[\sum_{m=0}^{U_0^0-1} I\{\zeta_n \Psi^0 \in \cdot\} I\{U_0^0 \geq 1\}\Big], \ wp1. \tag{50}$$

The limit in (48) must hold over any subsequence of T_n^0 such as $T_{C^0(n)}$, that is, we can replace n by $C^0(n)$; that is what we now use on the right-hand-side of (47):

$$\lim_{n\to\infty}\frac{1}{n}\sum_{m=0}^{T^0_{C^0(n)}} I\{\zeta_m\Psi^0\in\cdot\} = \lim_{n\to\infty}\Big(\frac{C^0(n)}{n}\Big)\frac{1}{C^0(n)}\sum_{m=0}^{T^0_{C^0(n)}} I\{\zeta_m\Psi^0\in\cdot\} \tag{51}$$

$$= \lambda E\Big[\sum_{m=0}^{U_0^0-1} I\{\zeta_n\Psi^0\in\cdot\}I\{U_0^0\ge 1\}\Big], wp1, \tag{52}$$

where we use the fact that $\frac{C^0(n)}{n}\to\lambda$, wp1, from Proposition 3.4.

■

Remark 3.3 The U_0^0 in the Palm inversion formula in Proposition 45 is taken from Ψ^0, not from $\zeta_0\Psi^0$: we are crucially using the fact that for Ψ^0, the interarrival time sequence $\{U_n^0\}$ forms a stationary ergodic sequence; the interarrival times are no longer stationary for $\zeta_0\Psi^0$. (They remain stationary under the *point shifts*, $\theta_i\Psi^0$, not the time shifts $\zeta_m\Psi^0$.) We are breaking up time into stationary ergodic cycles of time via $T_n^0=\sum_{i=0}^{n-1}U_i^0,\ n\ge 1$.

3.2.1 Applications of the Palm inversion formula

Here we give several examples illustrating how the Palm inversion formula works. We re-visit examples from Section 3.1.

1. We consider the cyclic deterministic Example 6 in Section 3.1, with cycles $\{1,0,2\}$. We have

$$\{X_n^0\}=\begin{cases}\{\dots,2,1,0,\ 2=X_0^0\ ,1,0,2,\dots\} & \text{wp } 2/3\\ \{\dots,1,0,2,\ 1=X_0^0\ ,0,2,1,0,2,\dots\} & \text{wp } 1/3\ .\end{cases}$$

We will show how the inversion formula yields $P(X_0^*=1)=P(X_0^*=0)=P(X_0^*=2)=1/3$, hence giving us $\{X_n^*\}$.

Since $\lambda=E(X_0^*)=1$ we must compute for $i\in\{0,1,2\}$,

$$P(X_0^*=i)=E\Big[\sum_{m=0}^{U_0^0-1} I\{X_m^0=i\}I\{U_0^0\ge 1\}\Big]. \tag{53}$$

Recalling that $P(X_0^0=2,\ J_0^0=2)=P(X_0^0=2,\ J_0^0=1)=1/3,\ P(X_0^0=1,\ J_0^0=1)=P(X_0^0=1)=1/3$, we see that $\{U_0^0\ge 1\}$ can happen only in two (disjoint) ways:

(a) $\{X_0^0=2,\ J_0^0=1\}=\{X_0^0=2,\ U_0^0=T_1^0=1\}$, in which case $U_0^0-1=0$ and thus only $m=0$ is counted in (53) yielding

$$P(X_0^*=i)=P(X_0^0=i,X_0^0=2,J_0^0=1),$$

or

(b) $\{X_0^0 = 1,\ J_0^0 = 1\} = \{X_0^0 = 1,\ U_0^0 = T_1^0 = 2\}$ in which case $U_0^0 - 1 = 1$ and thus $m = 0$ and $m = 1$ are counted in (53) yielding

$$P(X_0^* = i) = P(X_0^0 = i, X_0^0 = 1, J_0^0 = 1) + P(X_1^0 = i, X_0^0 = 1, J_0^0 = 1).$$

For $i = 2$, only (a) above yields a non-zero probability, $P(X_0^* = 2) = P(X_0^0 = 2, J_0^0 = 1) = 1/3$. For $i = 1$, or $i = 0$, only (b) above yields a non-zero probability each using only one of the sum, $P(X_0^* = 0) = P(X_1^0 = 0, X_0^0 = 1, J_0^0 = 1) = P(X_0^0 = 1) = 1/3$
$P(X_0^* = 1) = P(X_0^0 = 1, X_0^0 = 1, J_0^0 = 1) = P(X_0^0 = 1) = 1/3$

2. Our second example: the iid case, Example 3 in Section 3.1.

$$\{X_n^*\} = \{\ldots, X_{-2}^*, X_{-1}^*, X_0^*, X_1^*, X_2^* \ldots\},$$

where $\{X_n^* : n \in \mathbb{Z}\}$ is any iid sequence of non-negative rvs with $0 < E(X_0^*) < \infty$. Then

$$\{X_n^0\} = \{\ldots, X_{-2}^*, X_{-1}^*, X_0^0, X_1^*, X_2^* \ldots\},$$

where X_0^0 and J_0^0, independent of the iid $\{X_n^0 : n \neq 0\}$, are distributed as in Proposition 3.5 by jointly constructing a copy of (I_0^0, J_0^0) and using $X_0^0 = I_0^0 + J_0^0$.

Recalling $B_0^* \stackrel{\text{def}}{=} (X_0^* \mid X_0^* > 0)$, denoting a true (time-stationary) batch size,

$$P(B_0^* = i) = \frac{P(X_0^* = i)}{P(X_0^* > 0)},\ i \geq 1, \tag{54}$$

and

$$E(B_0^*) = \frac{E(X_0^*)}{P(X_0^* > 0)} = \frac{\lambda}{P(X_0^* > 0)}. \tag{55}$$

we deduce that

$$P(X_0^* = i) = \frac{\lambda P(B_0^* = i)}{E(B_0^*)},\ i \geq 1. \tag{56}$$

We will prove that the Palm inversion formula yields (56) for $i \geq 1$ and yields $P(X_0^* = 0) = P(X_0^* = 0)$ too, thus showing how the Palm inversion formula indeed retrieves Ψ^* from Ψ^0.

We will use the Palm inversion formula via

$$P(X_0^* = i) = \lambda E\Big[\sum_{m=0}^{U_0^0 - 1} I\{X_m^0 = i\} I\{U_0^0 \geq 1\}\Big] = \lambda \sum_{l=1}^{\infty} E\Big[\sum_{m=0}^{l-1} I\{X_m^0 = i\} I\{U_0^0 = l\}\Big]. \tag{57}$$

As seen in our previous example, $\{U_0^0 \geq 1\} = \{J_0^0 = 1\}$; the interarrival time $U_0^0 = T_1^0$ is positive only if $T_0^0 = 0$ is the last point in the batch X_0^0 at the origin. Note that if $U_0^0 = l \geq 2$, then $X_m^0 = 0,\ 1 \leq m \leq l - 1$. Thus for $l \geq 1$ and any $i \geq 1$,

$$\Big[\sum_{m=0}^{l-1} I\{X_m^0 = i\} I\{U_0^0 = l\}\Big] = I\{X_0^0 = i,\ U_0^0 = l\},$$

which implies from (57) that

$$\lambda E\Big[\sum_{m=0}^{U_0^0-1} I\{X_m^0 = i\}I\{U_0^0 \geq 1\}\Big] = \lambda \sum_{l=1}^{\infty} P(X_0^0 = i,\ U_0^0 = l). \tag{58}$$

Note that for $l \geq 2,\ i \geq 1$,

$$\{X_0^0 = i,\ U_0^0 = l\} = \{(I_0^0 = i-1, J_0^0 = 1),\ X_1^0 = 0, \ldots, X_{l-1}^0 = 0,\ X_l^0 > 0\}.$$

For $l = 1,\ i \geq 1$,

$$\{X_0^0 = i,\ U_0^0 = 1\} = \{(I_0^0 = i-1, J_0^0 = 1,\ X_1^0 > 0\}.$$

Thus by the iid $\{X_n^0 : n \geq 1\}$ all distributed as X_0^*, and, independently, the biased X_0^0, we have

$$P(X_0^0 = i,\ U_0^0 = l) = \frac{P(B_0^* = i)}{E(B_0^*)} P(X_0^* = 0)^{l-1} P(X_0^* > 0),\ l \geq 1,\ i \geq 1,$$

where we are using from Proposition 3.5,

$$P(I_0^0 = l, J_0^0 = k) = \frac{P(B_0^* = l+k)}{E(B_0^*)},\ l \geq 0,\ k \geq 1.$$

Thus from (58) we have

$$P(X_0^* = i) = \frac{\lambda P(B_0^* = i)}{E(B_0^*)},\ i \geq 1,$$

which indeed is correct from (56) above.

For $i = 0$, we again use (57) and simply observe that since $P(X_0^0 = 0) = 0$ and $P(X_m^0 = 0) = P(X_0^* = 0),\ m \geq 1$, and X_m^0 is independent of $U_0^0 = T_1^0,\ m \geq 1$, we have

$$P(X_m^0 = 0,\ U_0^0 = l) = P(X_0^* = 0)P(U_0^0 = l),\ l \geq 1,\ m \geq 1$$

and hence (57) reduces to ($m = 0$ can't be counted since $P(X_0^0 = 0) = 0$, so $l = 0$ takes care of that)

$$\begin{aligned} P(X_0^* = 0) &= \lambda P(X_0^* = 0) \sum_{l=0}^{\infty} l P(U_0^0 = l) \\ &= \lambda P(X_0^* = 0) E(U_0^0) \\ &= P(X_0^* = 0), \end{aligned}$$

where we are using the fact that $E(U_0^0) = \lambda^{-1}$ from Proposition 3.4.

3. Markov chain Example 5 in Section 3.1.

We will need the transition matrix $P = (P_{i,j}),\ i,j \geq 0$, and follow along in the spirit of our previous example, replacing step by step independence with step by step conditional independence (e.g., Markov property). For $i \geq 1$, and $l = 1$

$$P(X_0^0 = i,\ U_0^0 = 1) = \frac{P(B_0^* = i)}{E(B_0^*)}(1 - P_{i,0}).$$

For $i \geq 1$, and $l \geq 2$

$$P(X_0^0 = i,\ U_0^0 = l) = \frac{P(B_0^* = i)}{E(B_0^*)} P_{i,0} P_{0,0}^{l-2}(1 - P_{0,0}).$$

$$\sum_{l=2}^{\infty} P_{0,0}^{l-2}(1 - P_{0,0}) = 1.$$

Thus the final answer summed up from $l = 1$ to ∞ is:

$$\frac{P(B_0^* = i)}{E(B_0^*)}(1 - P_{i,0}) + \frac{P(B_0^* = i)}{E(B_0^*)} P_{i,0} = \frac{P(B_0^* = i)}{E(B_0^*)}.$$

Thus multiplying by λ gets us back to $P(X_0^* = i)$ just as for the iid case via the use of (56).

For the $P(X_0^* = 0)$ computation, we will join in $\{X_0^0 = i\}$ for $i \geq 1$ and then sum up over $i \geq 1$ at the end. Recalling that the X_0^* has the stationary distribution satisfying $\pi = \pi P$, we have that

$$\pi_i = \frac{\lambda P(B_0^* = i)}{E(B_0^*)},\ i \geq 1.$$

We now want to retrieve $\pi_0 = P(X_0^* = 0)$. For any $1 \leq m \leq l-1$, and $i \geq 1$, $l \geq 2$,

$$P(X_m^0 = 0,\ X_0^0 = i,\ U_0^0 = l) = P(X_0^0 = i,\ U_0^0 = l) = \frac{P(B_0^* = i)}{E(B_0^*)} P_{i,0} P_{0,0}^{l-2}(1 - P_{0,0}).$$

Thus summing up to $l - 1$ yields

$$\frac{P(B_0^* = i)}{E(B_0^*)} P_{i,0}(l-1) P_{0,0}^{l-2}(1 - P_{0,0}).$$

Summing up $(l-1)P_{0,0}^{l-2}(1-P_{0,0})$ over l then yields the mean of the geometric distribution, $(1 - P_{0,0})^{-1}$. Thus, the Palm inversion formula yields

$$P(X_0^* = 0,\ X_0^0 = i) = \lambda \frac{P(B_0^* = i)}{E(B_0^*)} P_{i,0}(1 - P_{0,0})^{-1} = \pi_i P_{i,0}(1 - P_{0,0})^{-1},\ i \geq 1. \quad (59)$$

But, from $\pi = \pi P$, we have

$$\pi_0 = \sum_{i=0}^{\infty} \pi_i P_{i,0},$$

and hence

$$\sum_{i=1}^{\infty} \pi_i P_{i,0} = \pi_0 - \pi_0 P_{0,0} = \pi_0(1 - P_{0,0}).$$

Thus summing up (59) over $i \geq 1$ yields $P(X_0^* = 0) = \pi_0$ as was to be shown.

4 Campbell's Theorem

Campbell's Theorem extends rather obvious relations for product sets to arbitrary sets using the monotone class theorem in measure/integration theory. Applications to queueing theory such as Little's Law become direct applications. We cover that here, starting with the most general form, and then moving on to cover cases when the marked point process is endowed with some form of stationarity.

For any non-negative measurable function $f = f(n, k)$, $f : \mathbb{Z} \times \mathbb{K} \longrightarrow \mathbb{R}_+$, and any marked point process ψ, define

$$\psi(f) = \sum_{j=-\infty}^{\infty} f(t_j, k_j).$$

Proposition 4.1 (Campbell's Theorem, general case) *If Ψ is a random marked point process, then for any non-negative measurable function $f = f(n, k)$,*

$$E(\Psi(f)) = \int_{\mathbb{Z}\times\mathbb{K}} f(b) v(db),$$

where v is the intensity measure, $v(B) = E(C(B)),\ B \in \mathcal{B}(\mathbb{Z} \times \mathbb{K})$.

Proof : Let $B \in \mathcal{B}(\mathbb{Z} \times \mathbb{K})$, and let $f(n, k) = I\{(n, k) \in B\}$. Then $\Psi(f) = C(B)$, and

$$E(C(B)) = v(B) = \int_B v(db) = \int_{\mathbb{Z}\times\mathbb{K}} f(b) v(db).$$

So the result holds for simple functions of the form $f(n, k) = \sum_{i=1}^{l} a_i I\{(n, k) \in B_i\}$, where the B_i are disjoint Borel sets, and the $a_i \geq 0$. Then from standard integration theory we can construct a monotone increasing sequence f_m of such simple functions such that $f_m \to f$ point wise as $m \to \infty$ and use the monotone convergence theorem.

■

When the marked point process is time-stationary, we get a much stronger result:

Proposition 4.2 (Campbell's Theorem, stationary case) *If Ψ^* is a time-stationary and ergodic marked point process, then for any non-negative measurable function $f = f(n, k)$,*

$$E(\Psi^*(f)) = \lambda E\Big[\sum_{n=-\infty}^{\infty} f(n, K_0^0)\Big] = \lambda \sum_{n=-\infty}^{\infty} E(f(n, K_0^0)).$$

Proof : That the last equality holds is standard since f is assumed non-negative, so Fubini's theorem (in the special form of Tonelli's Theorem) can be used. So we need to prove the first equality. For any indicator function of the form $f(n, k) = I\{n \in A,\ k \in \mathrm{K}\}$, with $|A| < \infty$ we have $E(\Psi^*(f)) = E(C^*(A \times \mathrm{K})) = \lambda|A|P(K_0^0 \in \mathrm{K})$ from Proposition 3.7. Also, it is immediate that for an f of this kind

$$\lambda E\Big[\sum_{n=-\infty}^{\infty} f(n, K_0^0)\Big] = \lambda E(\sum_{n\in A} I\{K_0^0 \in \mathrm{K}\}) = \lambda|A|P(K_0^0 \in \mathrm{K}).$$

So the result holds for such indicator functions. Thus it is immediate that the result will hold more generally for simple functions of the form $f(n,k) = \sum_{i=1}^{l} a_i f_i(n,k)$ where $f_i(n,k) = I\{n \in A_i,\ k \in \mathrm{K}_i\}$, the $a_i \geq 0$ are constants, and the l pairs (A_i, K_i) are disjoint. Then, we can approximate a general f (such as $f(n,k) = I\{(n,k) \in B\},\ B \in \mathcal{B}(\mathbb{Z} \times \mathbb{K})$) point-wise by a monotone increasing sequence of such non-negative simple functions $f_m \to f$ as $m \to \infty$ and use the monotone convergence theorem. ■

A classic example utilizing Campbell's Theorem is a proof of *Little's Law* ($l = \lambda w$) in a stationary ergodic setting. In this case $\Psi^* = \{(T_j^*, W_j^*)\}$, where T_j^* is the j^{th} customer's arrival time into a queueing system and $W_j^* \in \mathbb{R}_+$ (the j^{th} mark) denotes their sojourn time (total time spent in the system), and we are assuming the existence of such a time-stationary version. The Palm version $\Psi^0 = \{(T_j^0, W_j^0)\}$ represents stationarity from the view of arriving customers. It is important to understand that the existence of stationary versions depend highly on the queueing model in question, and proving the existence of such stationarity is not trivial in general. A time-stationary version of $L(n) =$ the number of customers in the system at time $n \in \mathbb{Z}$ is given by

$$L^*(n) = \sum_{T_j^* \leq n} I\{W_j^* > n - T_j^*\},\ n \in \mathbb{Z}.$$

Since it is time-stationary, we can and will focus on $L^*(0)$,

$$L^*(0) = \sum_{T_j^* \leq 0} I\{W_j^* > |T_j^*|\},\ n \in \mathbb{Z}.$$

In continuous time, and under the assumption of non-batches, this kind of proof using a Campbell's Theorem can be found in Franken et al [10]. Also see [2] and [20], for various continuous-time queueing applications of stationary marked point process theory. Perhaps what is new below is that we are allowing batches and are in discrete time:

Proposition 4.3 (Little's Law) *Suppose for a queueing system that there exists a time-stationary ergodic version $\Psi^* = \{(T_j^*, W_j^*) : j \in \mathbb{Z}\}$. (We are assuming as always that $0 < \lambda = E(X_0^*) < \infty$.) If $E(W_0^0) < \infty$, then $E(L^*(0)) < \infty$ and $E(L^*(0)) = \lambda E(W_0^0)$.*

Proof : Defining Borel set $B = \{(n,w) \in \mathbb{Z} \times \mathbb{R}_+ : n \leq 0,\ w > |n|\}$, and $f(n,w) = I\{(n,w) \in B\}$, we see that $L^*(0) = \Psi^*(f)$. Applying Campbell's theorem yields

$$E(L^*(0)) = \lambda \sum_{n \leq 0} P(W_0^0 > |n|) = \lambda \sum_{n=0}^{\infty} P(W_0^0 > n) = \lambda E(W_0^0).$$

■

We now move on to a form of Campbell's Theorem that is in between the above two cases: the case of a *periodic stationary* marked point process. Here is the setup: A marked point process $\{(X_n, \overline{\mathbf{K}}_n)\} : n \in \mathbb{Z}\}$ has the property that for a fixed integer $d \geq 2$ (the *period*), $\Psi_l^* \stackrel{\text{def}}{=} \{(X_{md+l}, \overline{\mathbf{K}}_{md+l}) : m \in \mathbb{Z}\}$ forms a time-stationary marked point process for each $0 \leq l \leq d-1$. $\mathcal{C}_m \stackrel{\text{def}}{=} \{(X_{md+l}, \overline{\mathbf{K}}_{md+l}) : 0 \leq l \leq d-1\},\ m \in \mathbb{Z}$, is called the m^{th} *cycle* and it is assumed that $\{\mathcal{C}_m : m \in \mathbb{Z}\}$ forms a *stationary and ergodic sequence*. In particular, each cycle has the same distribution as the initial one $\mathcal{C}_0 = \{(X_l, \overline{\mathbf{K}}_l) : 0 \leq l \leq d-1\}$.

The marked point process will be referred to as a *periodic stationary ergodic marked point process*. (If $d = 1$ we are back to a time-stationary and ergodic point process.)

We let Ψ_l^0 denote a Palm version of Ψ_l^* and to simplify notation, we let P_l^0 and E_l^0 denote the distribution and expected value under the distribution of Ψ_l^0. We define $\lambda_l = E(X_l)$, and we assume that $0 < \lambda_l < \infty,\ 0 \leq l \leq d-1$. Because of the periodicity, $\lambda_n = E(X_n) = \lambda_l$, and $P_n^0 \stackrel{\text{def}}{=} P_l^0$, if $n \in \{md + l : m \in \mathbb{Z}\},\ 0 \leq l \leq d-1$.

Proposition 4.4 (Campbell's Theorem, periodic stationary case) *If Ψ is a periodic stationary ergodic marked point process with period d, then for any non-negative measurable function $f = f(n, k)$,*

$$E(\Psi(f)) = \sum_{n=-\infty}^{\infty} \lambda_n E_n^0(f(n, K_0)).$$

Proof : Recall from Proposition 3.7 and from the proof of Campbell's Theorem, the stationary case, that for each $0 \leq l \leq d-1$, $E(\Psi_l^*(f)) = \lambda_l |A| P_l^0(K_0 \in \mathrm{K})$, for any f of the form $f(n, k) = I\{n \in A,\ k \in \mathrm{K}\}$, with $|A| < \infty$. For any subset $A \subseteq \mathbb{Z}$, let $A_l = A \cap \{md + l : m \in \mathbb{Z}\},\ \ 0 \leq l \leq d-1$. The A_l are disjoint and $A = \cup_{l=0}^{d-1} A_l$.

For any $A \subseteq \mathbb{Z}$ and any measurable $\mathrm{K} \subseteq \mathbf{K}$ it thus follows that for f of the kind $f(n, k) = I\{n \in A,\ k \in \mathrm{K}\}$, with $|A| < \infty$,

$$E(\Psi(f)) = \sum_{l=0}^{d-1} \lambda_l |A_l| P_l^0(K_0 \in K) = \sum_{n=-\infty}^{\infty} \lambda_n E_n^0(f(n, K_0)).$$

The proof is then completed by moving on to simple functions and the monotone convergence theorem as in the proof of Campbell's Theorem, the stationary case. ■

As an application of Proposition 4.4, we now will directly derive the stochastic discrete-time Periodic Little's Law (PLL) of Whitt and Zhang in [23], Theorem 3. They first derive a sample-path PLL (Theorem 1), and then give a stochastic version (Theorem 3) by using the sample-path version (almost surely). In continuous-time, there is a general stochastic version of a PLL for the case when the arrival process is simple (no batches) and has a periodic rate, such as Theorem 4 in [23], which utilizes methods from [19] which dealt with special models with iid service times and a periodic non-stationary Poisson arrival process; Palm distributions are used.

As our primitive, we start with a periodic stationary marked point process $\{(X_n, \overline{\mathbf{K}}_n)\} : n \in \mathbb{Z}\}$, with period d, in which the $\overline{\mathbf{K}}_n$ are a list of the sojourn times $\{W_j\}$ of the X_n customer arrivals at time n.

Using the Palm distribution P_l^0, $P_l^0(W_0 \in \cdot)$, denotes the stationary distribution for sojourn time over all customers who arrive in a time slot l. (Under P_l^0, W_0 is the sojourn time of a randomly chosen customer from a batch in a time slot l.)

If $d = 1$, then it would simply be the stationary distribution of sojourn time over all customers, and we could use Proposition 4.3. But we want to handle the case when $d \geq 2$.

The quantity $\lambda(c) \stackrel{\text{def}}{=} \sum_{l=0}^{d-1} \lambda_s$ is the total arrival rate per cycle. We assume for each l that $P_l^0(W_0 \in \cdot)$ defines a proper distribution and has finite and non-zero first moment, $0 < E_l^0(W_0) < \infty$. (This also ensures that $E^0(W_0) \stackrel{\text{def}}{=} \sum_{l=0}^{d-1} \frac{\lambda_l}{\lambda_c} E_l^0(W_0) < \infty$; it is the average sojourn time over *all* customers.)

For each $0 \leq l \leq d-1$ the total number of customers in the system at time l is given by

$$L_l = \sum_{j=-\infty}^{l} I\{T_j \leq l,\ W_j > l - T_j\} = \Psi(f_l),$$

where $f_l(n, w) = I\{n \leq l,\ w > l - n\},\ n \in \mathbb{Z}$.

Proposition 4.5 (Periodic Little's Law) *Assuming a periodic stationary (and ergodic) marked point process for the queueing model, it holds for each* $0 \leq l \leq d-1$ *that*

$$E(L_l) = \sum_{n=-\infty}^{l} \lambda_n P_n^0(W_0 > l - n) < \infty.$$

Proof : Direct application of Proposition 4.4 as in the proof of Proposition 4.3 using the function $f_l(n, w) = I\{n \leq l,\ w > l - n\},\ n \in \mathbb{Z}$. Finiteness follows since for any $n \in \mathbb{Z}$, there are bounds $\lambda_n \leq \lambda(c)$ and $P_n^0(W_0 > l - n) \leq M(|n|) = \sum_{l=0}^{d-1} P_l^0(W_0 > |n|)$. But

$$\sum_{n=-\infty}^{0} M(|n|) = \sum_{l=0}^{d-1} E_l^0(W_0) < \infty,$$

because we assumed that $E_l^0(W_0) < \infty$ for all $0 \leq l \leq d-1$. ■

Remark 4.1 Inherent in our queueing applications (Little's Law, Periodic Little's Law) is the assumption that within any time slot, arrivals that occur are counted before any departures occur, and that the number of customers in the system is counted after the arrivals but before the departures. This is due to the discrete-time framework here; in continuous-time, the set of times at which an arrival and departure both occur simultaneously forms a set of Lebesgue measure 0 hence has no effect on such results.

Remark 4.2 Analogous to what we did above for $l = \lambda w$, one can also derive a stationary version of $H = \lambda G$ and a periodic stationary version of $H = \lambda G$ in discrete time. In fact $H = \lambda G$ can be considered equivalent to Campbell's Theorem; in continuous time, see for example Page 155 in [20], and [25].

5 Non-ergodicity

If a stationary marked point process Ψ^* is not ergodic, then issues arise: The distribution defined by

$$P(\Psi^0 \in \cdot) = \lim_{m \to \infty} \frac{1}{m} \sum_{j=0}^{m} P(\theta_j \Psi^* \in \cdot), \tag{60}$$

still exists, is point-stationary and of course still has **the natural (in applications) nice interpretation as "the distribution as seen by a randomly chosen point way out in the future"**, an empirical average over what all arrivals see, but it is no longer the same as

$$Q(\cdot) \stackrel{\text{def}}{=} \lambda^{-1} E\Big[\sum_{j=0}^{X_0^*-1} I\{\theta_j \Psi^* \in \cdot\}\Big], \tag{61}$$

which also exists, and is point-stationary, where $\lambda \stackrel{\text{def}}{=} E(X_0^*) = \{E^Q(U_0)\}^{-1}$. It is this distribution Q, in the literature, that is called the Palm distribution of Ψ^* (whether stationary and ergodic or just stationary but non-ergodic). Under ergodicity, they are the same; but otherwise in general they are quite different. This motivates right from the start a serious question: *If you are given a "stationary" but non-ergodic marked point process, how was it derived*?

As a very simple example, let us move to continuous time and consider a point process $\Psi = \{T_n\}$ that is a mixture of two Poisson processes hence non-ergodic: with probability $1/2$ it is a Poisson process at rate 1, and with probability $1/2$ it is a Poisson process at rate 2. With no point at the origin it is time-stationary, with a point at the origin it is point-stationary which would give us $P(\Psi^0 \in \cdot)$. Both of those versions are the natural way to think about what is the time-stationary versus the point-stationary versions. Under the time-stationary version, $\lambda = (1/2)(1+2) = \frac{3}{2}$ as should be. Moreover using the point-stationary version Ψ^0 we get $E(U_0^0) = (1/2)(1+1/2) = 3/4$, again as should be. But using the Q distribution it holds that $E^Q(U_0) = 2/3$. Going a bit further, one can show that under Q, *the point process is a* $1/3$, $2/3$ *mixture* of a Poisson process at rate 1 and a Poisson process at rate 2.

To handle non-ergodicity along the lines of our approach, ones needs to condition on the invariant $\sigma-$ field as in Birkoff's ergodic theorem applied to a stationary but non-ergodic process. For details, the reader can consult [20] where that is exactly what is done in continuous time. Q ignores the invariant $\sigma-$ field and treats the process *as if it were ergodic.*

References

[1] S. Asmussen (2003). *Applied Probability and Queues* (2nd Ed.). Springer-Verlag, New York.

[2] F. Baccelli and P. Bremaud (1994). *Elements of Queueing Theory.* Springer-Verlag, Berlin.

[3] L. Billingsley (1968). *Probability* . Addison-Wesley, New York.

[4] A. Brandt, P. Franken, and B. Lisek (1990). Stationary Stochastic Models. John Wiley and Sons, New York.

[5] L. Breiman (1968). *Probability.* Addison-Wesley. Reprinted by SIAM 1998.

[6] H. Bruneel and B. G. Kim (2012). *Discrete-Time Models for Communication Systems Including ATM.* Springer, New York.

[7] H. Daduna and F. Shaßberger (1982). Networks of Queues in Discrete Time. *Zeitschrift fur Operations Research*, **27**, 139-175.

[8] R. Durrett (1991). *Probability: Theory And Examples.* Wadsworth and Brooks/Cole, Pacific Grove, California

[9] P. Gao, S. Wiitevrongel, and H. Bruneel (2004). Discrete-time multiserver queues with geometric service times. *Computers & Operations Research* **31**, 81-99.

[10] P. Franken, D. König, U. Arndt, and V. Schmidt (1981). Queues and Point Processes. Akademie-Verlag, Berlin.

[11] P. Glynn and K. Sigman (1992). Uniform Cesaro limit theorems for synchronous processes with applications to queues. *Stochastic Processes and their Applications*, **40**, 29-43.

[12] H. Halfin (1983). Batch Delays Versus Customer Delays. *The Bell System Technical Journal*, **62**, 2011-2015.

[13] O. Kallenberg (2017). *Random Measures, Theory and Applications*, Springer, New York.

[14] M. Loève (1978). *Probability Theory II*, 4th Edition. Springer-Verlag, New York.

[15] M. Miyazawa (1990). Derivation of Little's and related formulas by rate conversation law with multiplicity. Preprint Science Univ. of Tokyo.

[16] M. Miyazawa and Y. Takahashi (1992). Rate conservation principle for discrete-time queues. *Queueing Systems*, **12**, 215-230.

[17] J. Munkres (1975). *Topology a first course*, Princeton-Hall, New Jersey.

[18] G. Nieuwenhuis (1989). Equivalence of functional limit theorems for stationary point processes and their Palm distributions. *Probability Theory and Related Fields*, **81**, 593-608.

[19] T. Rolski (1989). Relationships between characteristics in periodic Poisson queues. *Queueing Systems*, **4**, 17-26.

[20] K. Sigman (1995). *Stationary Marked Point Processes: An Intuitive Approach*, Chapman and Hall, New York.

[21] W. Whitt (1983). Comparing batch delays and customer delays. *The Bell System Technical Journal*, **62**(7), 2001-2009.

[22] W. Whitt and X. Zhang (2017). A Data-Driven Model of an Emergency Department. *Operations Research for Health Care*, **12**, 1-15.

[23] W. Whitt and X. Zhang (2018). Periodic Little's Law. *Operations Research*, To Appear.

[24] W. Whitt (1985). The Renewal-Process Stationary-Excess Operator. *Journal of Applied Probability*, **22**, 156-167.

[25] W. Whitt (1992). $H = \lambda G$ and the Palm transformaton *Advances of Applied Probability*, **24**, 755-758.

[26] W. Whitt (2002). *Stochastic Process Limits*, Springer-Verlag, New York.

[27] R.W. Wolff (1989). *Stochastic Modeling and the Theory of Queues.* Prentice Hall, New Jersey.