ALEA, Lat. Am. J. Probab. Math. Stat. 19, 1753–1765 (2022)
DOI: 10.30757/ALEA.v19-66

A note on the Rényi criterion for Poisson processes and their identification

Gusztáv Morvai and Benjamin Weiss

Alfréd Rényi Institute of Mathematics, 13-15 Reáltanoda utca, H-1053, Budapest, Hungary
E-mail address: morvaigusztav@gmail.com
URL: http://math.bme.hu/∼morvai

Hebrew University of Jerusalem, Institute of Mathematics,, Jerusalem 91904 Israel
E-mail address: weiss@math.huji.ac.il
URL: http://mathematics.huji.ac.il/people/benjamin-weiss

Abstract. We give a sequence of binary functions defined on the finite observations of a stationary point process which will almost surely eventually take the value POISSON if the observed process is Poisson, and NOTPOISSON otherwise. We will give explicit upper bounds on the probability of misclassification in case the process is Poisson. The construction of the discrimination procedures will be based on Rényi's characterization of the Poisson process.

1. Introduction

Many works have been devoted to the question of what can be learned about a discrete time stationary stochastic process simply on the basis of successive observations of the output of a single random sample drawn from the probability space underlying the process, cf. Bailey (1976), Ryabko (1988), Algoet (1999), Suzuki (2003), Morvai and Weiss (2005), Csiszár and Talata (2006), Györfi et al. (2002), Morvai and Weiss (2007), Györfi and Ottucsák (2007), Takahashi (2011), Györfi et al. (2012), Jones et al. (2012), Felber et al. (2013), Gallo and Leonardi (2015), Morvai and Weiss (2021) and Ryabko (2019).

One type of popular question in this area involves discriminating between two classes of processes by means of a binary function which will almost surely stabilize on the correct answer (we have borrowed the term 'discrimination procedure' from Ryabko (2010), cf. Ryabko (2019) too). On the other hand very few works have been devoted to analogous questions for continuous time processes. For discrete time processes the simplest class is that of i.i.d. random variables. For continuous time processes the simplest one is the homogeneous Poisson point processes on the line. It can be described as a random countable discrete subset of $\mathbb{R}$ with the property that the random variables that count the number of points in disjoint intervals are independent with a Poisson distribution

Received by the editors July 3rd, 2022; accepted October 13th, 2022.
2010 *Mathematics Subject Classification.* 60G55.
Key words and phrases. Point Processes, Poisson Processes.
The first author was supported partly by Alfréd Rényi Institute of Mathematics, Bolyai János Research Scholarship and OTKA grant No. K75143.

and parameter proportional to the lengths of the intervals. An alternate description is via the inter-arrival process which in the case of a Poisson process consists of independent exponentially distributed random variables with a fixed parameter.

The question we shall take up in this paper is can one devise such a binary decision procedure which will discriminate between a Poisson point process and the class of all other ergodic stationary point processes.

While a great deal of work appears in the statistical literature concerning tests to determine whether a given data set is best modeled by such processes e.g. Lewis (1965), Hora (1978), Reinmuth (1971) and Chornoboy et al. (1988), in these papers the problem is not formulated as discriminating between Poisson point processes and other stationary ergodic point processes. The emphasis is on a statistical test which will succeed when the data derives from a Poisson process without really specifying what is the general class of processes that is being considered. Here we specify precisely the general class.

We will view a point process as a random discrete subset $\omega \subset \mathbb{R}$ and then provide a sequence of discrimination procedures $DPOISSON_t(\omega \cap [0,t])$ which will eventually almost surely stabilize on POISSON if the process we are sampling is Poisson and on NOTPOISSON otherwise. We will give explicit upper bounds on the probability of misclassification for finite sample size in case the process is in fact Poisson. In principle one cannot give similar upper bounds in the other direction when the competing process is a general stationary and ergodic point process that is not Poisson. This is because one can have such processes that are not Poisson but behave like a Poisson process for a very long time. For more details on this see Remark 4.4 below.

In Morvai and Weiss (2019) we gave such a discrimination procedure which was based on a characterization of the Poisson process via the inter-arrival process which in the case of a Poisson process consists of independent exponentially distributed random variables with a fixed parameter. However no explicit bounds were given there. The discrimination procedure we will give in this paper will be based directly on the observations and a remarkable characterization of the Poisson process due to Rényi (1967).

Here is a more formal description of the setup that we are considering. The random discrete subset can be described by random variables $\ldots, R_{-1} < 0 \le R_0 < R_1 \ldots$ defined on a probability space $(\Omega, \Sigma, \mathbf{P})$ where the elements of Ω are discrete subsets of $\mathbb{R}$ and the random variables $R_i(\omega)$ are the points of ω. The σ-algebra Σ is generated by functions that count the number of points in $\omega \cap [a,b]$ for arbitrary intervals $[a,b]$. Just as in the case of discrete time stationary processes $X_n(\omega)$ where shifting the time by k to $X_{n+k}(\omega)$ is represented by the k-th iterate of a transformation T of the probability space so that $X_{n+k}(\omega) = X_n(T^k(\omega))$ there is a natural one parameter family of transformations T_t defined on the space Ω that takes the element ω to $\omega - t$. We will assume that $\mathbf{P}$ is invariant and ergodic under this flow.

The assumption of ergodicity is not really needed since we are treating pointwise phenomena. The basic theorem on the ergodic decomposition of probability preserving flows (i.e. one parameter groups of transformations T_t) implies that with probability one we observe outputs of an ergodic flow.

In addition we will only consider those point processes with the property that the expected number of points in $\omega \cap [-N, N]$ is finite for all N. It is easy to see that in this case there is a positive finite number ν called the intensity of the process such that for any Borel measurable subset B of $\mathbb{R}$ the expected number of points in B is given by $\nu|B|$ where $|B|$ represents the Lebesgue measure of B and ν is a positive constant called the **intensity** of the process.

Our discrimination procedure will deal directly with the point process and so we do not really need to use the inter-arrival times in the construction of the procedure. However they will play a certain role in the proof and so we proceed to define them.

The inter-arrival times are defined by $X_n = R_{n+1} - R_n$. Even though the event $R_0 = 0$ has zero probability, sense can be made of conditioning on this event and a measure, called the Palm measure, can be defined on Ω so that the X_n form a stationary stochastic process which is ergodic if and only if the original point process was. Indeed one verifies easily that $\mathbf{E}\{1/X_0\} < \infty$ and defines a probability on Ω by $\mathbf{P}^0 = (1/X_0 \times \mathbf{P})/\mathbf{E}\{1/X_0\}$. If we then let Σ^0 denote the σ-algebra generated by the inter-arrival times and by Ω^0 the "atoms" of this σ-algebra we get the probability space of the Palm measure. A detailed discussion of this can be found in Thorisson (2000, Chapter 8). In this duality, ergodicity of the flow corresponds exactly to the ergodicity of the discrete stationary process (cf. Thorisson 8.7). It is worth remarking that null sets are preserved under this transformation so that a typical sample for an ergodic point process gives a typical sample for the inter-arrival time process.

Our discrimination procedure will be based on examining the properties of discrete time processes that are derived from the point process in a simpler way. For any fixed increment h we can consider the $\{0,1\}$-valued random variables Y_n defined by $Y_n(\omega) = 0$ if and only if $\omega \cap [nh, (n+1)h] = \emptyset$. For a Poisson process this will be a sequence of independent and identically distributed random variables. We will use a universal procedure for determining independence for each of such processes with $h = 1/2^k$. Since in either case T_h preserves the probability, the $\{Y_n\}$ form a stationary process even if the point process is not Poisson. However, there may be values of h for which the discrete transformation T_h is not ergodic. As is well known for an ergodic flow, this can happen for at most a countable set of values of h, which however are not known to us. Rather than assuming a stronger ergodicity hypothesis we will deal with this problem in the proof. For the definition of discrimination procedure itself the possible lack of ergodcity of T_h presents no problem. To show that for non Poisson processes our discrimination procedure will eventually detect this fact we will rely on a classical criterion of Rényi which we discuss in the next section.

2. Rényi's Characterization of the Poisson Process

Let $\delta(x)$ be an increasing positive function defined for $x > 0$ such that $\lim_{x\to 0} \delta(x) = 0$, and denote by $|A|$ the Lebesgue measure on the real line. In Rényi (1967), it is proved the following.

Rényi's theorem *If a point process on the line satisfies the following two conditions:*

$$P(A \text{ is empty}) = e^{-\lambda|A|} \tag{2.1}$$

and

$$P(\text{there are more than two points in } A\ ) \le \lambda|A|\delta(\lambda|A|) \tag{2.2}$$

for all sets A which are arbitrary finite disjoint unions of intervals then the point process is a homogeneous Poisson point process with parameter λ.

Our discrimination procedure will be based on this characterization of Rényi. In the proof of this theorem Rényi used the first condition for all sets A of the form indicated but the second condition was needed only for small intervals. This is satisfied for the ergodic point processes with finite intensity that we consider and so our discussion of the discrimination procedure will be focused on the key first condition. For the sake of completeness we sketch a proof of this well known fact.

Lemma 2.1. *For an ergodic stationary point process $(\Omega, \Sigma, \mathbf{P}, T_t)$ with finite intensity where the elements of Ω are discrete subsets of $\mathbb{R}$ we have*

$$\mathbf{P}\{|\omega \cap [0,h]| \ge 2\} \le \lambda h \delta(\lambda h)$$

and lim $\delta(h) = 0$ as h tends to zero (λ is the intensity).

Proof: We will compute the probability in question using the ergodic theorem applied to a typical sample point of the process. The points in our sample can be indexed by $\ldots, R_{-1} < 0 \le R_0 < R_1 \ldots$ and the inter-arrival times defined by $X_n = R_{n+1} - R_n$ form a stationary ergodic process with respect to the Palm measure. Fix a large N and let $I = \{i \ge 0 | R_i < N\}$ and $J = \{i \in I | X_i < h\}$. By the ergodic theorem the ratio $|I|/N$ will tend to the intensity. While the ratio $|J|/|I|$ will tend to the probability with respect to the Palm measure of the event $\{X_i < h\}$. Clearly $\{|\{\omega - s\} \cap [0,h]| \geq 2\}$ for $s \in [0,N]$ can only happen when s is within an interval of length h lying to the left of a point R_i with $i \in J$. Thus an upper bound for the probability we are estimating is given by the limit of $h|J|/N = h\frac{|I|}{|N|}\frac{|J|}{|I|}$ as N tends to infinity. This yields $h\times$ intensity $\times$ Palm measure$(\{X_i < h\}) = h\lambda\times$ Palm measure$(\{X_i < h\})$. Defining $\delta(\lambda h) =$ Palm measure$(\{X_i < h\})$. Since the flow measure and the Palm measure are equivalent measures and the intensity is finite the proof of Lemma 2.1 is complete.

3. Preliminaries

Let $0 \leq R_0 < R_1 < \ldots$ be the inter-arrival times of a stationary ergodic point process. For $k = 0, 1, \ldots$ define $X_i^{(k)}$ as

$$X_i^{(k)} = \begin{cases} 0 & \text{if the interval } ((i-1)2^{-k}, i2^{-k}] \text{ is empty} \\ 1 & \text{otherwise.} \end{cases}$$

That is,

$$X_i^{(k)}(\omega) = X_1^{(k)}(T^{(i-1)2^{-k}}\omega).$$

For any fixed k, the time series $\{X_i^{(k)}\}$ will be stationary, however in general even though the real parameter flow T_t is ergodic, this does not imply that the discrete flow defined by the iterates of $T_{2^{-k}}$ is ergodic. We shall give an example of this below.

If the point process is Poisson then for all k, the random variables $\{X_i^{(k)}\}$ are independent. In addition the converse also holds, if for all k these time series are independent then the point process is Poisson. Our discrimination procedure will be based on a universal procedure for determining the independence of a time series from a single sample of the series. Our earlier paper Morvai and Weiss (2019) was based on a direct consideration of the inter-arrival times but we were unable to give there any estimates for the probability of error in the discrimination procedure as we are able to do here.

Example 3.1. Here is a simple example of a flow which illustrates both how to define point processes and why some fixed times may be non-ergodic. The space is the two dimensional torus $\mathbb{R}^2/\mathbb{Z}^2$ with Lebesgue measure and the flow is

$$T_t(u,v) = (u + ta, v + tb)$$

where the ratio b/a is irrational and the addition is taken modulo 1.

This flow preserves Lebesgue measure and is the continuous analogue of an irrational rotation of the circle. It is well known to be ergodic. You get a point process by taking some curve C in the torus which wraps around the torus so that the lines $\{(u + ta, v + tb) : t \in \mathbb{R}\}$ intersect the curve in a discrete set. Sampling a point for our process is choosing a starting point (u,v) according to Lebesgue measure. The discrete subset of the intersection points on $\mathbb{R}$ defines a point process which is stationary because T_t preserves Lebesgue measure.

For an explicit example of a curve C consider the broken line that goes from $(0,0)$ to $(1/2,1/2)$ and then goes up from $(1/2,1/2)$ to $(0,1)$ which on the torus is the same point as $(0,0)$. Now if $t_0 = 1/a$ the transformation T_{t_0} is not ergodic because all vertical lines $u =$ constant are invariant under T_{t_0}, while for $t_0 = 1/b$ the transformation T_{t_0} is not ergodic because all horizontal lines $v =$ constant are now invariant under T_{t_0}. In addition for any integer m and $t_m = \frac{1}{am}$ the transformation T_{t_m} will also not be ergodic since now for any constant c the union of the vertical lines $\bigcup_{i=1}^m \{u = c + i/m\}$ will be invariant. If we take for example $(a,b) = (1/2, \sqrt{2}/2)$ then, while the flow is ergodic the transformations defined above will not be ergodic for all k.

Despite the fact that we cannot be sure that the discrete time process $\{X_i^{(k)}\}$ is ergodic nonetheless, almost surely, when we sample $\{X_i^{(k)}(\omega)\}$ according to P we will be seeing an ergodic process. This follows immediately from the ergodic decomposition of stationary processes. This means that the universal tests we will use can be applied. Here is a more precise explanation of this. First we define **generic points**.

Definition 3.2. A point $\{x_n\}_{n=-\infty}^{\infty}$ where $x_n \in \{0,1\}$, is a generic point for an ergodic stationary process $\{Y_n\}$ if for any binary word $w_1^l \in \{0,1\}^l$ the frequency of w_1^l in $\{x_n\}_{n=-\infty}^{\infty}$ tends to the probability of $\{Y_1^l = w_1^l\}$. (We use the notation $Y_1^l = (Y_1, \ldots, Y_l)$, and $w_1^l = (w_1, \ldots, w_l)$.) More formally:

$$\frac{\#\{0 \le n < N : x_{n+1}^{n+l} = w_1^l\}}{N} \to \mathbf{P}(Y_1^l = w_1^l).$$

By the Birkhoff ergodic theorem and the ergodic decomposition (cf. Kallenberg (1997)), almost surely, points in any stationary process are generic for an ergodic process. Furthermore, the sets of generic points are disjoint for distinct ergodic components. More formally we have:

Lemma 3.3. *If (Ω, Σ, P, T_t) is an ergodic flow then almost surely, for all k, $\{X_i^{(k)}\}$ is a generic point for a stationary and ergodic process.*

Proof: For each k, $X_n^{(k)}$ defined as above on (Ω, Σ, P, S) determines a stationary process where the time shift on the sample space is given by $S = T_{2^{-k}}$. Let $\theta \in E_k$ parametrize these ergodic components. What this means is that we have a representation of P as

$$P = \int_{E_k} P_\theta d\mu(\theta)$$

where each P_θ represents an ergodic measure for S. For each θ there is a P_θ-null set $B_{k,\theta}$ such that for $\omega \notin B_{k,\theta}$ the sequence $\{X_n^{(k)}(\omega)\}$ is generic for the ergodic process defined by P_θ. Using the ergodic decomposition (cf. Kallenberg (1997)) and Fubini's theorem gives us the result that there is a P-null set B_k such that for $\omega \notin B_k$ the sequence $\{X_n^{(k)}(\omega)\}$ is generic for some ergodic process. Let $B = \bigcup_{k=1}^{\infty} B_k$, $P(B) = 0$ and for $\omega \notin B$, for all k $\{X_n^{(k)}(\omega)\}$ is generic for an ergodic process. The proof of Lemma 3.3 is complete.

The next lemma is about individual sequences.

Lemma 3.4. *Assume that the empirical frequencies of all k-blocks in a sequence $a_1, a_2, \ldots$ where $a_i \in \{0,1\}$ converge to independence, along a subsequence N_j. That is to say for all k and $w_1^k \in \{0,1\}^k$,*

$$\frac{\#\{1 \le n \le N_j : a_{1+n}^{n+k} = w_1^k\}}{N_j} \to p^{\sum_{i=1}^k w_i}(1-p)^{k-\sum_{i=1}^k w_i}.$$

Then if we sample the k-blocks only at every other time instants, $2n$, the same empirical limits persist, i.e.

$$\frac{\#\{0 \le n \le \lfloor \frac{N_j}{2} \rfloor - 1 : a_{1+2n}^{2n+k} = w_1^k\}}{\lfloor \frac{N_j}{2} \rfloor} \to p^{\sum_{i=1}^k w_i}(1-p)^{k-\sum_{i=1}^k w_i}.$$

Proof: Fix $k \ge 1$, $w_1^k \in \{0,1\}^k$ and $\epsilon > 0$ arbitrarily. Let Z_i be independent identically distributed binary random variables with $P(Z_i = 1) = p$. Form the two first order Markov chains

$$U_i = (Z_{2i}, Z_{2i+1}, \ldots, Z_{2i+k-1})$$

and

$$V_i = (Z_{2i-1}, Z_{2i}, \ldots, Z_{2i+k-2})$$

with states space $\{0,1\}^k$. Clearly for the stationary distribution of these Markov chains is given by:

$$P(U_i = w_1^k) = P(V_i = w_1^k) = p^{\sum_{i=1}^k w_i}(1-p)^{k-\sum_{i=1}^k w_i}.$$

By the weak law of large numbers for mixing Markov chains, for $m > k/\epsilon$ sufficiently large there is a set $\mathcal{G}$ of good words with length $m+k$ so that for both $\{U_i\}$ and $\{V_i\}$ the empirical distributions of both state $(U_i = w_1^k)$ and $(V_i = w_1^k)$ are within ϵ its probability and

$$P((Z_1,\dots,Z_{m+k}) \in \mathcal{G}) > 1-\epsilon.$$

Now since all words of length $m+k$ occur in $\{a_n\}$ with the asymptotic frequency given by the distribution $Z_1,\dots,Z_{m+k}$ we know that the set of n

$$\mathcal{N} = \{n \ge 1 : a_n^{n+m+k-1} \in \mathcal{G}\}$$

has asymptotic frequency at least $(1-\epsilon)$. We now construct a sequence of disjoint intervals of lengths $(m+k)$ that cover at least $(1-\epsilon)$ of $[1,N_j]$ as follows. Let $n_1 = \min\{n \in \mathcal{N}\}$, and then inductively having chosen n_i let $n_{i+1} = \min\{n \in \mathcal{N} \bigcap [n_i+m+k, N_j]\}$. Clearly the intervals $[n_i, n_i+m+k]$ cover $\mathcal{N} \bigcap [1,N_j]$ and so these disjoint intervals cover at least $(1-\epsilon)$ of $[1,N_j]$ by $m+k$ blocks that belong to $\mathcal{G}$. Calculate the frequency of w_1^k along $2n+1$ by averaging first along the good $m+k$ blocks and then over the whole sequence. The discrepancy of the frequency and the independent distribution will be within 3ϵ for large j. Since $\epsilon > 0$ was arbitrary Lemma 3.4 is proven.

4. The Discrimination Procedure

If the point process is Poisson then for any given k, the random variables $\{X_i^{(k)}\}$ are independent. Let $\{Z_n\}$ represent a stationary binary time series, and as usual we use the notation Z_1^n to denote the initial n elements of the series $\{Z_1, Z_2 \dots Z_n\}$. We assume that there is a discrimination procedure for independence of binary time series $g_n(Z_1^n)$, which takes two values, $\{DEP, IND\}$. This means that for any stationary and ergodic non independent binary time series almost surely we will eventually have that $g_n(Z_1^n) = DEP$, while for all independent identically distributed binary time series almost surely eventually $g_n(Z_1^n) = IND$. In addition, in the latter case we have a bound for the error, $P(g_n(Z_1^n) = DEP) \le f_n$, where f_n is monotone decreasing and for a monotone sequence of integers $1 \le m_n$ increasing to ∞, and for a strictly monotone increasing sequence of integers $s_n \ge n$,

$$\sum_{n=1}^{\infty} \sum_{k=0}^{m_n-1} f_{2^k s_n} < \infty.$$

For the existence of such procedures see e.g. Section 3 (Theorem 2) in Morvai and Weiss (2011). (Cf. Ryabko and Astola (2005), Ryabko and Astola (2006), Ryabko et al. (2006), Ryabko et al. (2016) and Theorem 1 and Corollary 2 in Morvai and Weiss (2013) also.)

First we define our discrimination procedure $DPOISSON_t$ for this subsequence s_n as

$$DPOISSON_{s_n} = \begin{cases} POISSON & \text{if } g_{2^k s_n}(X_1^{(k)},\dots,X_{2^k s_n}^{(k)}) = IID \text{ for all } k=0,\dots,m_n-1 \\ NOTPOISSON & \text{otherwise.} \end{cases}$$

Now we define $DPOISSON_t$ for $t \ne s_n$ for any n as

$$DPOISSON_t = DPOISSON_{s_n} \text{ if } s_n < t < s_{n+1}.$$

Theorem 4.1. *Assume that the point process (Ω, Σ, P, T_t) is stationary and ergodic. Then if the point process is a Poisson process*

$$P(DPOISSON_t = NOTPOISSON) \le \sum_{k=0}^{m_n-1} f_{2^k s_n} \le m_n f_{s_n}$$

for $s_n \le t < s_{n+1}$ and $DPOISSON_t = POISSON$ eventually almost surely. If the point process is not a Poisson process then $DPOISSON_t = NOTPOISSON$ eventually almost surely.

Our estimate for the probability of error depends on the performance of the discrimination subprocedure for independence used by our algorithm. Next we give the results for the case when one uses the discrimination procedure g_n for independence in Morvai and Weiss (2011) (Theorem 2 in Section 3).

Corollary 4.2. *Use the discrimination procedure g_n for independence in Morvai and Weiss (2011) (Theorem 2 in Section 3 with arbitrary paramethers $0 < \gamma < 1$ and $0 < \beta < \frac{1-\gamma}{2}$). Then*

$$f_n = \begin{cases} 1 & \text{if } 1 \le n \le \left(\frac{10}{1-2\beta-\gamma}\right)^{\frac{1}{1-2\beta-\gamma}} \\ \min\left(28n^4 e^{\frac{-n^{1-2\beta-\gamma}}{2}}, 1\right) & \text{otherwise} \end{cases}$$

and f_n is monotone decreasing. Choose $m_n = s_n = n$. Assume that the point process (Ω, Σ, P, T_t) is stationary and ergodic. If the point process is a Poisson process then

$$P(DPOISSON_t = NOTPOISSON) \le 28(t-1)^5 e^{\frac{-(t-1)^{1-2\beta-\gamma}}{2}}$$

for $1 + \left(\frac{10}{1-2\beta-\gamma}\right)^{\frac{1}{1-2\beta-\gamma}} < t$ and $DPOISSON_t = POISSON$ eventually almost surely. If the point process is not a Poisson process then $DPOISSON_t = NOTPOISSON$ eventually almost surely. Particularly, with the choice of $\gamma = \beta = 0.01$ we get

$$f_n = \begin{cases} 1 & \text{if } 1 \le n \le 10 \\ \min\left(28n^4 e^{\frac{-n^{0.97}}{2}}, 1\right) & \text{otherwise.} \end{cases}$$

Now if the point process is a Poisson process then

$$P(DPOISSON_t = NOTPOISSON) \le 28(t-1)^5 e^{\frac{-(t-1)^{0.97}}{2}}$$

for $12 < t$ and $DPOISSON_t = POISSON$ eventually almost surely. If the point process is not a Poisson process then $DPOISSON_t = NOTPOISSON$ eventually almost surely.

Remark 4.3. We should point out that we will actually demonstrate in the course of the proof a "single orbit" version of Rényi's theorem which may be formulated as follows. If we have a single point ω which is generic for some ergodic point process (Ω, Σ, P, T_t) of finite intensity, and if in addition for all k the sequences defined by $X_i^k(\omega)$ are generic for independent processes (not related a priori) then the point process (Ω, Σ, P, T_t) is a Poisson process with parameter λ determined by the formula $P\{X_1^0 = 0\} = e^{-\lambda}$. This is a further result in the spirit of "Single Orbit Dynamics".

Remark 4.4. In the literature of statistical tests it is required to give estimates for both type I and type II errors. In Theorem 4.1 and Corollary 4.2 we give an estimate for the probability of misclassification only in case the process is a Poisson point process. In principle one cannot give similar upper bounds in the other direction when the competing process is a general stationary and ergodic point process that is not Poisson. This is because one can have such processes that are not Poisson but behave like a Poisson process for a very long time. To see this just consider point processes with independent inter-arrival times but with distributions that are very close to an exponential distribution. This is why we do not call our scheme a test but a discrimination procedure.

Remark 4.5. We note that this discrimination procedure can be extended to $\mathbb{R}^d$. The point is that Renyi's theorem is quite general. (Cf. Theorem 10.9 in Kallenberg (1997).) For $\mathbb{R}^d$ one has to modify the discrimination procedure for independence to random fields X_k where $k = (k_1, k_2, ...k_d)$ is a multi index and the finite observations are now for $|k| < n$. Then using the same arguments we can get a discrimination procedure with an error bound for Poisson point processes in higher dimensions.

Remark 4.6. Note that the parameters $0 < \alpha$, and $0 < \beta$ in Corollary 4.2 can be chosen arbitrarily small. If the process is a Poisson point process, the smaller these parameters are, the faster the probability of misclassification will tend to zero (but the rate will still remain subexponential). We do not know if, possibly with a different discrimination procedure, one can achieve exponential error rate instead of our subexponential one in Corollary 4.2.

Remark 4.7. It seems to us unlikely that one could find a way to abstractly define a "best discrimination procedure", in the sense of achieving a smallest probability of misclassification. Given a discrimination procedure which eventually classifies the process correctly, one can define a second discrimination procedure which, for a fixed period of time, will classify the observed process as Poisson (regardless of the observed data) and then will agree with the first discrimination procedure. In this way, for the second classification procedure, the probability of misclassification given the process is Poisson is zero during this time period and the asymptotic behaviour of both discrimination procedures are the same. Since this time period can be chosen arbitrarily long, the first discrimination procedure can not be considered better than the second.

Remark 4.8. Our earlier paper Morvai and Weiss (2019) was based on a direct consideration of the inter-arrival times which are real valued random variables. The discrimination procedure in Morvai and Weiss (2019) first examines if these real valued random variables are independent or not. If the answer is yes then the discrimination procedure determines if the distribution of these inter-arrival times is exponential or not. If both answers are yes then the discrimination procedure says that the point process is Poisson. In this paper, our discrimination procedure deals with a simpler situation in which the random variables take only two values. We were unable to give any estimates for the probability of error in Morvai and Weiss (2019) whereas in this paper, in Theorem 4.1 and Corollary 4.2, we do give estimates for the probability of misclassification in case the process is indeed a Poisson point process.

Proof of Theorem 4.1:

If the point process is Poisson then for each k, the process $\{X_i^{(k)}\}$ is a sequence of independent binary random variables. By assumption,

$$P(g_{s_n 2^k}(X_1^{(k)}, \dots, X_{2^k s_n}^{(k)}) = DEP \quad \text{for some } 0 \le k \le m_n - 1)$$
$$\le \sum_{k=0}^{m_n-1} P(g_{s_n 2^k}(X_1^{(k)}, \dots, X_{2^k s_n}^{(k)}) = DEP) \le \sum_{k=0}^{m_n-1} f_{2^k s_n}$$

and since by assumption

$$\sum_{n=1}^{\infty} \sum_{k=0}^{m_n-1} f_{2^k s_n} < \infty$$

by the Borel-Cantelli lemma, eventually almost surely,

$$g_{s_n 2^k}(X_1^{(k)}, \dots, X_{2^k s_n}^{(k)}) = IND \, \text{for all } 0 \le k \le m_n - 1$$

thus

$$DPOISSON_{s_n} = POISSON$$

eventually almost surely. Therefore, by definition,

$$DPOISSON_t = POISSON$$

eventually almost surely.

By assumption, $g_n(Z_1^n) = DEP$ eventually almost surely for all stationary and ergodic non independent binary processes and $g_n(Z_1^n) = IND$ eventually almost surely for all independent identically distributed binary processes. The ergodic decomposition (cf. Kallenberg (1997)) easily shows, by an argument similar to the proof of Lemma 3.3 that the same conclusion will hold when the procedure is applied to a stationary non ergodic process. There are only two possibilities. The first is that for some k, the point ω is generic for a non independent process. Then eventually, $g_{s_n 2^k}(X_1^{(k)}(\omega), \ldots, X_{2^k s_n}^{(k)}(\omega)) = DEP$. In this case clearly our discrimination procedure will eventually say NOTPOISSON. The second possibility is that for all k, $\{X_i^{(k)}(\omega)\}$ is a generic point of an independent binary process. In this case we would at first like to see that our discrimination procedure will eventually stabilize on POISSON. This will give us the fact that our procedure eventually stabilizes and it will then suffice to show that if it eventually stabilizes on POISSON then the process is Poisson. Using the ergodic decomposition (cf. Kallenberg (1997)) as we did before we can define a set $B(n,k)$ which consists of those $\omega\prime$ such that $\{X_i^{(k)}(\omega\prime)\}$ is a generic point of some independent ergodic process and $g_{s_n 2^k}(X_1^{(k)}(\omega\prime), \ldots, X_{2^k s_n}^{(k)}(\omega\prime)) = DEP$. Furthermore by Fubini's theorem again we will have,

$$P\left(B(n,k)\right) \le f_{2^k s_n}$$

and in turn,

$$P\left(\bigcup_{0\le k\le m_n-1} B(n,k)\right) \le \sum_{k=0}^{m_n-1} f_{2^k s_n}.$$

Since the upper bound is summable, by the Borel-Cantelli lemma eventually ω will not be in that set and we conclude that in case for all k the sequences $\{X_i^{(k)}(\omega)\}$ are generic for independent processes our discrimination procedure will eventually stabilize on POISSON.

We must now show that in this case the original process is indeed a Poisson process. For this we will apply Rényi's Theorem. By Lemma 2.1 Rényi's second condition (2.2) holds and so it remains to verify the first. To do this we will assume, as we may, that the point ω is generic for our real flow in the sense that for all finite disjoint unions of intervals, A, the probability of the event $\omega \cap A = \emptyset$ can be computed using the ergodic theorem on the indicator of that event along the orbit of ω. Under this assumption and our preceding one that for all k the sequence $\{X_i^{(k)}(\omega)\}$ is generic for an independent process we proceed to verify Rényi's first condition.

For each k, there is a probability distribution

$$\begin{aligned} P_{\omega,k}(w_1^l) &= \lim_{n\to\infty} \frac{\#\{1 \le i \le 2^k s_n - l : (X_{1+i}^{(k)}, \ldots, X_{l+i}^{(k)}) = w_1^l\}}{2^k s_n - l} \\ &= P_{\omega,k}(1)^{\sum_{i=1}^l w_i} P_{\omega,k}(0)^{\sum_{i=1}^l (1-w_i)} \end{aligned}$$

where $w_1^l \in \{0,1\}^l$.

Define

$$\lambda_0 = -\ln(P_{\omega,0}(0)) \;\text{ that is }\; P_{\omega,0}(0) = e^{-\lambda_0}.$$

Observe that

$$X_1^{(k)} = 0 \text{ if and only if } (X_1^{(k+1)}, X_2^{(k+1)}) = (0,0)$$

$$X_2^{(k)} = 0 \text{ if and only if } (X_3^{(k+1)}, X_4^{(k+1)}) = (0,0)$$

and

$$X_i^{(k)} = 0 \text{ if and only if } (X_{2i-1}^{(k+1)}, X_{2i}^{(k+1)}) = (0,0).$$

In order to connect the unknown parameters of the k and $k+1$ processes we use Lemma 3.4 which gives us that,

$$P_{\omega,k}(0) = (P_{\omega,k+1}(0))^2.$$

At this point a simple induction will show that:

$$P_{\omega,k}(0) = (P_{\omega,k-1}(0))^{\frac{1}{2}} = (P_{\omega,k-2}(0))^{\frac{1}{4}} = \cdots = (P_{\omega,0}(0))^{2^{-k}}.$$

Thus

$$P_{\omega,k}(0) = e^{-\lambda_0 2^{-k}}.$$

If A is a finite union of dyadic intervals then for large enough K, the event "A is empty" is measurable with respect to all the k-grids for $k \geq K$. In other words one can determine the event "A is empty" in terms of the $\{X_i^{(k)}\}$, and the frequency of dyadic positions in the sequence where that event occurs in $X_1^{(k)}, \dots, X_{2^k s_n}^{(k)}$ tends to $e^{-\lambda_0 |A|}$ as n tends to ∞, where $|A|$ denotes the Lebesgue measure of the set A. That is, for $k \geq K$,

$$P_{\omega,k}(A \text{ is empty}) = e^{-\lambda_0 |A|}. \tag{4.1}$$

Now our assumption that ω is generic for the flow implies:

$$\frac{1}{T}\int_0^T I_{\{A \text{ is empty}\}}(T_t\omega)dt \to P(A \text{ is empty}).$$

By assumption,

$$A = \bigcup_{i=1}^{c} (a_i 2^{-K}, b_i 2^{-K}]$$

for some $a_1, \dots, a_c$ and $b_1, \dots, b_c$. Let $k > K$ be arbitrary. Now for a fix ω consider the set defined as

$$\{t \in [0, s_n] : T_t\omega \in \{A \;\text{ is not empty}\}\}.$$

But this set equals

$$\bigcup_{i=0}^{r} \{R_i - a : \quad a \in A\}$$

where $r = \max\{i : R_i \leq s_n\}$. (Recall that the R_i's represent the places of the marks of the point process.) Therefore the set

$$\{t \in [0, s_n] : T_t\omega \in \{A \;\text{ is not empty}\}\}$$

is the union of not more than rc intervals. Therefore the set

$$\{t \in [0, s_n] : T_t\omega \in \{A \;\text{ is empty}\}\}$$

is also a union of not more that rc pieces of intervals. If $(u, v]$ is one of these intervals then

$$(u,v] \subseteq (\lfloor 2^k u \rfloor 2^{-k}, (\lfloor 2^k u \rfloor + 1)2^{-k}] \bigcup \left(\bigcup_{\substack{1 \leq i \leq s_n 2^k: \\ T^{i2^{-k}}\omega \in \{A \text{ is empty}\}}} (i2^{-k}, (i+1)2^{-k}] \right).$$

Therefore, if for $i = 1, 2, \ldots, z$, $(u_i, v_i]$ denotes these intervals, where $z \leq cr$,

$$\bigcup_{i=1}^{z}(u_i, v_i] \subseteq \left(\bigcup_{i=1}^{z}(\lfloor 2^k u_i \rfloor 2^{-k}, (\lfloor 2^k u_i \rfloor + 1)2^{-k}]\right) \bigcup \left(\bigcup_{\substack{1 \leq i \leq s_n 2^k: \\ T^{i2^{-k}}\omega \in \{A \text{ is empty}\}}} (i2^{-k}, (i+1)2^{-k}]\right).$$

It follows that

$$\frac{1}{s_n}\int_0^{s_n} I_{\{A \text{ is empty}\}}(T_t\omega)dt \leq \frac{c}{2^k}\frac{r}{s_n} + 2^{-k}\frac{|\{1 \leq i \leq s_n 2^k : T^{i2^{-k}}\omega \in \{A \text{ is empty}\}\}|}{s_n}.$$

But

$$\frac{r}{s_n} \to \mu$$

almost surely and

$$\frac{|\{1 \leq i \leq s_n 2^k : T^{i2^{-k}}\omega \in \{A \text{ is empty}\}\}|}{s_n 2^k} \to P_{\omega,k}(A \text{ is empty}) = e^{-\lambda_0 |A|}.$$

Therefore

$$P(A \text{ is empty}) \leq c\mu 2^{-k} + e^{-\lambda_0|A|}.$$

Now since k was arbitrary,

$$P(A \text{ is empty}) \leq e^{-\lambda_0|A|}. \tag{4.2}$$

Now we will prove that $P(A \text{ is empty}) \geq e^{-\lambda_0|A|}$. Define

$$\hat{A}_k = \bigcup_{i=1}^{c}(a_i 2^{-K}, b_i 2^{-K} - 2^{-k}].$$

By ergodicity,

$$\frac{1}{s_n}\int_0^{s_n} I_{\{\hat{A}_k \text{ is empty}\}}(T_t\omega)dt \to P(\hat{A}_k \text{ is empty}).$$

By (4.1), we know that

$$P_{\omega,k}(A \text{ is empty}) = e^{-\lambda_0|A|}.$$

When counting those i's in $X_i^{(k)}(\omega)$ up to $2^k s_n$ which contribute to the calculation of $P_{\omega,k}(A \text{ is empty})$ notice that each such interval $(i2^{-k},\ (i+1)\ 2^{-k}]$ is contained in the set of t for which $I_{\{\hat{A}_k \text{ is empty}\}}(T_t\omega) = 1$. Therefore,

$$P(\hat{A}_k \text{ is empty}) \geq P_{\omega,k}(A \text{ is empty}) = e^{-\lambda_0|A|}.$$

Now

$$\{\hat{A}_k \text{ is not empty}\} \subseteq \{A \text{ is not empty}\}$$

and

$$\{A \text{ is not empty}\} = \{\hat{A}_k \text{ is not empty}\} \bigcup (\bigcup_{i=1}^{c}(b_i 2^{-K} - 2^{-k}, b_i 2^{-K}]).$$

But

$$\begin{aligned} P(A \text{ is not empty}) &\leq P(\hat{A}_k \text{ is not empty}) + cP((0, 2^{-k}] \text{ is not empty}) \\ &\leq P(\hat{A}_k \text{ is not empty}) + c2^{-k}\mu. \end{aligned}$$

Therefore,

$$0 \leq P(A \text{ is not empty}) - P(\hat{A}_k \text{ is not empty}) \leq c2^{-k}\mu$$

and

$$|P(A \text{ is not empty}) - P(\hat{A}_k \text{ is not empty})| \leq c2^{-k}\mu.$$

Thus

$$|P(A \text{ is empty}) - P(\hat{A}_k \text{ is empty})| \le c2^{-k}\mu.$$

Now

$$P(A \text{ is empty}) \ge P(\hat{A}_k \text{ is empty}) - c2^{-k}\mu \ge e^{-\lambda_0|A|} - c2^{-k}.$$

Since k was arbitrary,

$$P(A \text{ is empty}) \ge e^{-\lambda_0|A|}.$$

By (4.2),

$$P(A \text{ is empty}) = e^{-\lambda_0|A|}.$$

Since any finite union of intervals can be approximated by finite union of dyadic intervals we have proved the first condition of Rényi (2.1). By Rényi's characterization this implies that the point process is Poisson. The proof of Theorem 4.1 is complete.

References

Algoet, P. Universal schemes for learning the best nonlinear predictor given the infinite past and side information. *IEEE Trans. Inform. Theory*, **45** (4), 1165–1185 (1999). MR1686250.

Bailey, D. H. *Sequential Schemes for Classifying and Predicting Ergodic Processes*. Ph.D. thesis, Stanford University (1976). MR2626644.

Chornoboy, E. S., Schramm, L. P., and Karr, A. F. Maximum likelihood identification of neural point process systems. *Biol. Cybernet.*, **59** (4-5), 265–275 (1988). MR961117.

Csiszár, I. and Talata, Z. Context tree estimation for not necessarily finite memory processes, via BIC and MDL. *IEEE Trans. Inform. Theory*, **52** (3), 1007–1016 (2006). MR2238067.

Felber, T., Jones, D., Kohler, M., and Walk, H. Weakly universally consistent static forecasting of stationary and ergodic time series via local averaging and least squares estimates. *J. Statist. Plann. Inference*, **143** (10), 1689–1707 (2013). MR3082227.

Gallo, S. and Leonardi, F. Nonparametric statistical inference for the context tree of a stationary ergodic process. *Electron. J. Stat.*, **9** (2), 2076–2098 (2015). MR3397402.

Györfi, L., Kohler, M., Krzyżak, A., and Walk, H. *A distribution-free theory of nonparametric regression*. Springer Series in Statistics. Springer-Verlag, New York (2002). ISBN 0-387-95441-4. MR1920390.

Györfi, L. and Ottucsák, G. Sequential prediction of unbounded stationary time series. *IEEE Trans. Inform. Theory*, **53** (5), 1866–1872 (2007). MR2317147.

Györfi, L., Ottucsák, G., and Walk, H. *Machine Learning for Financial Engineering*, volume 8 of *Advances in Computer Science and Engineering*. World Scientific (2012). DOI: 10.1142/p818.

Hora, S. C. A screening test for the Poisson process. *Decis. Sci.*, **9** (3), 414–420 (1978). DOI: 10.1111/j.1540-5915.1978.tb00730.x.

Jones, D., Kohler, M., and Walk, H. Weakly universally consistent forecasting of stationary and ergodic time series. *IEEE Trans. Inform. Theory*, **58** (2), 1191–1202 (2012). MR2918019.

Kallenberg, O. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York (1997). ISBN 0-387-94957-7. MR1464694.

Lewis, P. A. W. Some results on tests for Poisson processes. *Biometrika*, **52**, 67–77 (1965). MR207107.

Morvai, G. and Weiss, B. On classifying processes. *Bernoulli*, **11** (3), 523–532 (2005). MR2146893.

Morvai, G. and Weiss, B. On sequential estimation and prediction for discrete time series. *Stoch. Dyn.*, **7** (4), 417–437 (2007). MR2378577.

Morvai, G. and Weiss, B. Testing stationary processes for independence. *Ann. Inst. Henri Poincaré Probab. Stat.*, **47** (4), 1219–1225 (2011). MR2884232.

Morvai, G. and Weiss, B. Universal tests for memory words. *IEEE Trans. Inform. Theory*, **59** (10), 6873–6879 (2013). MR3106870.

Morvai, G. and Weiss, B. A note on discriminating Poisson processes from other point processes with stationary inter arrival times. *Kybernetika (Prague)*, **55** (5), 802–808 (2019). MR4055577.

Morvai, G. and Weiss, B. On universal algorithms for classifying and predicting stationary processes. *Probab. Surv.*, **18**, 77–131 (2021). MR4255241.

Reinmuth, J. E. A Test for the Detection of a Poisson Process. *Decis. Sci.*, **2** (3), 260–263 (1971). DOI: 10.1111/j.1540-5915.1971.tb01461.x.

Rényi, A. Remarks on the Poisson process. *Studia Sci. Math. Hungar.*, **2**, 119–123 (1967). MR212861.

Ryabko, B. and Astola, J. Application of data compression methods to hypothesis testing for ergodic and stationary processes. In *2005 International Conference on Analysis of Algorithms*, Discrete Math. Theor. Comput. Sci. Proc., AD, pp. 399–407. Assoc. Discrete Math. Theor. Comput. Sci., Nancy (2005). MR2193138.

Ryabko, B. and Astola, J. Universal codes as a basis for time series testing. *Stat. Methodol.*, **3** (4), 375–397 (2006). MR2252392.

Ryabko, B., Astola, J., and Gammerman, A. Application of Kolmogorov complexity and universal codes to identity testing and nonparametric testing of serial independence for time series. *Theoret. Comput. Sci.*, **359** (1-3), 440–448 (2006). MR2252209.

Ryabko, B., Astola, J., and Malyutov, M. *Compression-based methods of statistical analysis and prediction of time series.* Springer, Cham (2016). ISBN 978-3-319-32251-3; 978-3-319-32253-7. MR3495276.

Ryabko, B. Y. Prediction of random sequences and universal coding. *Problemy Peredachi Informatsii*, **24** (2), 3–14 (1988). MR955983.

Ryabko, D. Discrimination between B-processes is impossible. *J. Theoret. Probab.*, **23** (2), 565–575 (2010). MR2644876.

Ryabko, D. *Asymptotic nonparametric statistical analysis of stationary time series.* SpringerBriefs in Computer Science. Springer, Cham (2019). ISBN 978-3-030-12563-9; 978-3-030-12564-6. MR4290872.

Suzuki, J. Universal prediction and universal coding. *Syst. Comput. Japan*, **34** (6), 1–11 (2003). DOI: 10.1002/scj.10357.

Takahashi, H. Computational limits to nonparametric estimation for ergodic processes. *IEEE Trans. Inform. Theory*, **57** (10), 6995–6999 (2011). MR2882275.

Thorisson, H. *Coupling, stationarity, and regeneration.* Probability and its Applications (New York). Springer-Verlag, New York (2000). ISBN 0-387-98779-7. MR1741181.