A test for dependence between two point processes on the real line

Patrick Rubin-Delanchy and Nicholas A Heard

August 26, 2014

Patrick Rubin-Delanchy is a Heilbronn Research Fellow in Statistics at the University of Bristol, Bristol, UK (Email: patrick.rubin-delanchy@bristol.ac.uk) and Nicholas Heard is a Senior Lecturer in Statistics at Imperial College London, London, UK (Email: n.heard@imperial.ac.uk).

Abstract

Many scientific questions rely on determining whether two sequences of event times are associated. This article introduces a likelihood ratio test which can be parameterised in several ways to detect different forms of dependence. A common finite-sample distribution is derived, and shown to be asymptotically related to a weighted Kolmogorov-Smirnov test. Analysis leading to these results also motivates a more general tool for diagnosing dependence. The methodology is demonstrated on data generated on an email network, showing evidence of information flow using only timing information. Implementation code is available in the R package ‘mppa’.

Keywords: Point process, correlation, triggering, hypothesis test.

1 Introduction

Testing for dependence between two point processes is a long-standing statistical problem. When the two processes are on the real line, usually representing time, the points are usually interpreted as event times (or, simply, events). Scientific questions then often revolve around detecting:

(a) Triggering: the occurrence of an event in $A$ temporarily increases the rate of events in $B$.

(b) Correlation: the rate of events in $B$ is locally increased around events in $A$.

(c) Inhibition: the occurrence of an event in $A$ temporarily decreases the rate of events in $B$.

(d) Anti-correlation: the rate of events in $B$ is locally decreased around events in $A$.

Statistical methods to detect these effects have received decades of attention in the field of neurophysiology. The activity of a neuron is often recorded as a sequence of ‘spike’ times, called a neuronal spike train, which is often treated as a realisation of a point process on the real line. Comparing trains that are simultaneously generated by different neurons can shed light on how they are connected and, more generally, how information is processed in the nervous system. The literature in this field is relatively mature, for example a very highly cited paper by Perkel et al. [1967] proposed to test for interaction on the basis of
histograms of the times from $A$ to $B$ events. A number of difficulties were recognised, namely, that the proposed tests are not exact even when $A$ and $B$ are independent homogeneous Poisson processes and that the histograms can be misleading if the processes have an oscillatory component of similar frequency (causing spurious correlation when the processes are ‘phase-locked’ by chance). Since then, a number of model-based approaches were developed for this problem, notably in a series of papers by Brillinger (Brillinger et al., 1976; Brillinger and Segundo, 1979; Brillinger, 1988b, a, 1992).

In a more general context Ripley (1976, 1977) introduced the so-called $K$-function to measure second-order dependence between point processes defined on a topological space. From this work a number of articles followed, typically motivated by ecological or biological applications, adapting ideas to two- or three-dimensional settings. Lotwick and Silverman (1982) proposed to use an estimate of $K$ to test for interactions between different types of points on a plane. Berman (1986) presented some results for when $B$ is a spatial point pattern and $A$ is an arbitrary stochastic process. Back in a one-dimensional setting Doss (1989), also motivated by a neurophysiological application, proved that a one-dimensional version of the estimated (cross)-$K$ function is asymptotically normal.

The problem has now resurfaced in the analysis of network data, for example traffic generated on a computer network, messages and other connections made on social networks, mobile communications, email networks, the web, collaboration networks (e.g. in academia, music or film) and more. Such data can often be represented as a graph with point processes (e.g. communication times) occurring on every edge (e.g. a pair of computers). In being able to measure dependence between the events generated by edges or nodes, there is great potential to better understand information flow, discover new correlations and develop better network models. Recent approaches include Blundell et al. (2012), where email reciprocation is modelled using the mutually exciting point process models developed by Hawkes (1971a,b), and Perry and Wolfe (2013), developing a framework for modelling point process networks based on a version of Cox’s proportional intensity model. In this approach effects such as $A$ triggering $B$ are handled by including past events of $A$ as covariates for $B$.

This article seems to be the first to propose an exact (generalised) likelihood ratio test for association between two one-dimensional point processes. We assume we are given a model for $B$, which captures its statistical behaviour under the null hypothesis, for example encapsulating any seasonality, changepoints or drift. Then we test for a multiplicative effect on the intensity of $B$ within a certain interval following or surrounding every $A$ event. The multiplier and interval width are fit by maximum likelihood estimation, leading to a (generalised) likelihood ratio test for the corresponding effect. Depending on whether the alternative hypothesis assumes an increase or decrease in the intensity of $B$ events around $A$ events, and whether the intervals are centered on the event times or occur directly afterwards, the same procedure can be used to test for each of the effects listed in (a-d).

The tests obtained are easily implemented; their $p$-values are available in an exact (albeit recursive) form and they are shown to be asymptotically related to a weighted Kolmogorov-Smirnov (K-S) test. A key insight is noticing a duality with finding a changepoint in a homogeneous Poisson process (Lemma 1), after which mathematical considerations are greatly simplified. A by-product of the lemma is a new diagnostic tool for analysing dependence between point processes.

The remainder of this article is organised as follows. In Section 2 we focus on testing for events in $A$ triggering events in $B$ under a non-homogeneous Poisson null hypothesis. We give the test and derive its finite sample $p$-value. Section 3 proposes a number of adjustments to improve power against some alternatives and extends the methodology to
detecting other forms of dependence. In Section 4 analysis is extended to more general point processes on the real line. In Section 5 an asymptotic analysis finds a connection between our test and a weighted K-S test of a set of auxiliary variables constructed during the procedure. Finally in Section 6 the test is applied to detecting reciprocation and forwarding behaviour in an email network (the Enron email dataset).

2 A fixed duration excitation model

Let $A$ and $B$ be two simple point processes on the real line observed simultaneously from the first event time of $A$ up to an observation end time $L$. Neither process is explosive, so that the observed event times of $A$ form a finite set $A = \{a_1 < \ldots < a_m\}$, $m \geq 1$, and the event times of $B$ form a finite set $B = \{b_1 < \ldots < b_n\}$ where $b_1 \geq a_1$ and $n \geq 0$. Without loss of generality, let $a_1 = 0$. We make no further assumptions on $A$, and treat its event times as given when modelling $B$.

In this section we consider the special case where $B|A$ is a non-homogeneous Poisson process. Relaxations of these conditions are considered in Section 4. Given $A$, the process $B$ is assumed to have a deterministic, bounded, Lebesgue-measurable intensity function $\lambda_B(t)$. Heuristically $\lambda_B(t)dt = E dB(t)$.

We propose the following fixed duration, proportional excitation model for $B$:

$$
\lambda_B(t) = \begin{cases} 
\lambda_1 r(t) & t - a(t) \leq \tau, \\
\lambda_2 r(t) & \text{otherwise,} 
\end{cases} \quad t \in [0, L),
$$

where $a(t)$ is the most recent event in $A$ occurring at or before $t$, $\tau > 0$, $\lambda_1 \geq \lambda_2 \geq 0$ are unknown parameters, and $r$ is a known bounded non-negative Lebesgue-measurable function satisfying $\int_0^L r(v)dv = 1$.

Model (1) leads to a test with a very straightforward interpretation: is the relative proportion of events within time $\tau$ of an event in $A$ higher than can be explained by $r$ alone? To formalise this idea, we shall test the following null hypothesis:

$$
H_0 : \lambda_1 = \lambda_2 \quad \text{versus} \quad H_1 : \lambda_1 > \lambda_2.
$$

A natural test statistic is the generalised likelihood ratio

$$
\sup \{ \ell(B; \tau, \lambda_1, \lambda_2) : \tau > 0, \lambda_1 > \lambda_2 \geq 0 \} \\
\sup \{ \ell(B; \tau, \lambda_1, \lambda_2) : \tau > 0, \lambda_1 = \lambda_2 \geq 0 \}, \quad (2)
$$

where $\ell$ is likelihood of $B$ under model (1). Note that under $H_0$, $\tau$ has no real importance and, accordingly, $\ell(B; \tau, \lambda_1, \lambda_2)$ is functionally independent of $\tau$ when $\lambda_1 = \lambda_2$.

**Remark 1.** Probably the most common choice for $r$ will be the constant $1/L$, in which case $B$ is a homogeneous Poisson process under $H_0$. If the application makes this assumption unrealistic, a more informed choice of $r$ will not only bring the null behaviour of the test closer to its nominal distribution, derived later, but may lead to a gain in power under the alternative, for example if $B$ responds to $A$ despite being relatively inactive under $H_0$. If $r$ is unknown our (heuristic) recommendation is to use an estimate (as we do in Section 6).

**Remark 2.** If $B$ was observed for some time before $a_1$, a period that is ignored above, it could be modelled as being generated by $\lambda_2 r(t)$ for $t < a_1$. Doing so only requires some simple modifications to the results below.
2.1 Computing the test

Let \( \rho(X) = \int_X r(v) \, dv \), where \( X \) is any set that can be constructed by a countable union of intervals. \( \rho \) can be seen as an artificial measurement of time intervals that compensates for \( B \)'s activity levels under \( H_0 \). In fact, \( \rho \) is a Borel probability measure on \([0, L)\). The likelihood of Model (1) can be written (Daley and Vere-Jones, 2003, p. 232)

\[
\ell(B|A; \tau, \lambda_1, \lambda_2) \propto \lambda_1^{K(\tau)} \exp[-\lambda_1 \rho(T(\tau))] \times \lambda_2^{n-K(\tau)} \exp[-\lambda_2(1 - \rho(T(\tau)))]
\]

where \( T(\tau) = \{ t : t - a(t) \leq \tau \} \) is the union of all intervals during which \( B \) was intensified due to the events of \( A \) and \( K(\tau) = \#\{b_i \in T(\tau)\} \) is the number of events that occurred within them.

The likelihood is maximised as follows. Let \( u_1 \leq \ldots \leq u_n \) be the order statistics of \( \rho\{T(b_i - a(b_i))\}, i = 1, \ldots, n \). \( u_k \) can be interpreted as the effective proportion of time that \( B \) was intensified if \( \tau \) is equal to the \( k \)th smallest response time. The maximum likelihood parameters \( \hat{\tau}, \hat{\lambda}_1, \hat{\lambda}_2 \) are found within the following procedure:

Procedure 1 (Computation of \( T \) given \( u_1, \ldots, u_n \)).

For \( k = 1, \ldots, n \), let

\[
\ell_k = \left(\frac{k/n}{u_k}\right)^{k/n} \left(\frac{(n-k)/n}{1-u_k}\right)^{(n-k)/n}.
\]

Then, let

\[
\hat{k} = \arg \max_{k=1,\ldots,n} \{\ell_k : u_k \leq k/n\}, \text{ and}
\]

\[
\hat{\tau} = b_{\hat{k}} - a(b_{\hat{k}}), \quad \hat{\lambda}_1 = \hat{k}/u_{\hat{k}}, \quad \hat{\lambda}_2 = (n - \hat{k})/(1-u_{\hat{k}}),
\]

Return \( T = \ell_{\hat{k}} \).

The maximum of (3) is a monotonic function of \( \ell_k \). This can be shown by a straightforward argument, given in Appendix A. Note that \( k = n \) satisfies \( u_k \leq k/n \) so that \( \hat{k} \) is always defined.

Since the number of events of \( B \) can be equally well explained under the null as under the alternative, it is natural to condition on the value of \( n \). The denominator of (2) is a multiplicative term that depends only on \( n \) and can then be ignored. Hence any monotonic function of the numerator is admissible as a test statistic, and we use \( T = \ell_{\hat{k}} \).

2.2 A simple reformulation

The following lemma establishes a duality between testing for model (1) and detecting a Poisson process change-point, and is the key observation of this article.

Lemma 1. \( u_1, u_2, \ldots \) are the event times of a Poisson process \( U(x) \) on \([0,1)\) with a change-point in its intensity,

\[
\lambda(x) = \begin{cases} 
\lambda_1 & x \leq \rho(T(\tau)), \\
\lambda_2 & \text{otherwise},
\end{cases}
\]

for \( x \in [0,1) \).

This result, although it is very simple, is new to the best of the authors’ knowledge. However similar results exist, notably Theorem 1 in Loader (1991) on finding a changepoint in the hazard rate of \( n \) independently and identically distributed (IID) survival times. Lemma 1 and its proof provide a number of fascinating insights into the testing problem.
**Testing for a Poisson process change-point.** Revisiting Procedure 1, we can see that \( T \) only depends on the data through \( \{u_1, u_2, \ldots\} \). More than that, \( T \) is the generalized likelihood ratio test for model (1) against a homogeneous Poisson null hypothesis,

\[
\sup \{ \ell(U; \tau, \lambda_1, \lambda_2) : \tau > 0, \lambda_1 > \lambda_2 \geq 0 \} \over \sup \{ \ell(U; \tau, \lambda_1, \lambda_2) : \tau > 0, \lambda_1 = \lambda_2 \geq 0 \}.
\]

A diagnostic tool. Conditional on \( n, u_1, \ldots, u_n \) are ordered Uniform random variables under the null hypothesis, whereas under the alternative they should be, roughly speaking, more concentrated towards 0. To assess dependence the \( u_i \) can be inspected without having model (1) in mind, for example using goodness-of-fit tests against uniformity or visual diagnosis tools (e.g. a plot of the empirical cumulative distribution function (CDF) of \( u_i \) compared to \( y = x \)). In fact, in Section 5, we show that \( T \) is asymptotically related to a weighted Kolmogorov-Smirnov test.

‘Chopping and stretching’. The proof essentially relies on superposing segments of \( B \) and obtaining the intensity function of the superposed process by adding the intensities contributed by each of the relevant segments. This is possible because of the Poisson assumption. Then time is ‘stretched’ according to that intensity, leaving only the effects of the unknowns, \( \lambda_1, \lambda_2 \) and \( \tau \). This intuition invites us to consider other ways of superposing \( B \) based on \( A \), and motivates a number of tests to follow.

**Proof.** Notice that \( K(y) = \#\{b_i \in T(y)\} \) defines a point process over \([0, \max\{a_{i+1} - a_i : i = 1, \ldots, m\})\), temporarily defining \( a_{m+1} = L \). Its increments are obtained from the event times of \( B \) by superposing the segments \([a_i, a_{i+1})\), aligning to the left.

The intensity of \( K(y) \) is therefore \( h(y) = \sum\{t : a(t) = y\} \lambda_B(t) \) and its compensator is

\[
H(y) = \int_0^y h(s)ds = \begin{cases} 
\lambda_1 \rho(T(y)) & y \leq \tau, \\
\lambda_1 \rho(T(y)) + \lambda_2 (\rho(T(y)) - \rho(T(\tau))) & y > \tau.
\end{cases}
\]

If \( F \) is a continuous non-decreasing function on a sub-interval \( X \) of \( \mathbb{R} \) with image \( Y \), we define \( F^{-1}(y) = \inf\{x \in X : F(x) \geq y\} \) for \( y \in Y \). Because \( \rho \) is continuous, \( \rho(T(y)) \) is a continuous non-decreasing function \( \mu \), say, of \( y \). Then \( \mu^{-1} \) is right-continuous and non-decreasing with jumps at an at most countable set of values of \( x \) corresponding to intervals where \( \mu \) is constant (Daley and Vere-Jones 2007, p.420). Now let \( U(x) = K(\mu^{-1}(x)) \) for \( x \in [0, 1) \). The \( q \)th event time of \( U \) is

\[
\inf\{x : U(x) \geq q\} = \inf\{x : K(\mu^{-1}(x)) \geq q\} = \inf\{x : \mu^{-1}(x) \geq |b_i - a(b_i)|_{(q)}\} = \rho\{T([b_i - a(b_i)]_{(q)})\} = u_q,
\]

where \([b_i - a(b_i)]_{(q)}\) is the \( q \)th smallest response time (and the \( q \)th event time of \( K \)). Using the well-known time change theorem (Daley and Vere-Jones 2007, p.421), \( K(H^{-1}(z)) \) is a Poisson process with rate 1 for \( z \in [0, H(T)) \). Since \( U(x) = K(H^{-1}(\lambda_1 x)) \) for \( x \in [0, \rho(T(\tau))], U(x) \) is homogeneous Poisson with intensity \( \lambda_1 \) over that range. By a similar argument \( U(x) \) has intensity \( \lambda_2 \) over \( (\rho(T(\tau)), 1) \).

\[\square\]
2.3 Distribution of T

Let $F_n(x) = P(T \leq x | n)$ denote the distribution function of $T$ under the null hypothesis, conditional on $n$. To simulate $T$ from $F_n$ we can generate $n$ uniform variables independently, set their sorted values to $u_1 \leq \cdots \leq u_n$, and compute $T$ through Procedure 1. From this the moments and critical values of $T$ can be approximated by simulation.

On the other hand the p-value for an observed test statistic $t$, $P[T \geq t | n]$, can be computed exactly. To this end notice that

$$1 - P[T \geq t | n] = F_n(t) = P[u_1 \geq o_1, \ldots, u_n \geq o_n],$$

where $o_i$ is the solution for $x \in (0, i/n]$ to

$$t = \left( \frac{i/n}{x} \right)^{i/n} \left( \frac{(n-i)/n}{1-x} \right)^{(n-i)/n},$$

which is obtained numerically. Plainly speaking $o_i$ is the smallest $u_i$ can have been if we know $T = t$. Worsley (1988) has made a similar observation in the context of testing for a changepoint in the hazard rate of IID survival times. (Although the test considered there is two-sided.)

Various recursive formulas exist for computing the joint survival probability of $n$ ordered Uniform variables, although many are unsuitable for computation because they involve differences of very large numbers (so-called catastrophic cancellation). A ‘safe’ option is the $O(n^2)$ formula in Noé and Vandewiele (1968), as corrected in Noé (1972). This recursion is implemented in the R package corresponding to this article, ‘mppa’. For the reader’s convenience, it is also available in pseudo-code given below.

Algorithm 1 (The joint survival probability of ordered Uniform variables).

Let $\alpha = (o_1, \ldots, o_n, 1)$, $q(1) = 1$

for $j = 2, \ldots, n + 1$:

Set $q(j) = 0$, $q'(1) = 1$

for $i = 1, \ldots, j - 1$:

Set $q'(i + 1) = \sum_{k=0}^{i} \binom{i}{k} q(k + 1)(\alpha_j - \alpha_{j-1})^{i-k}$

Set $q = q'$

return $q'(n + 1)$

3 Adjustments and other types of association

In this section the test proposed in Section 2 is extended in two senses. First, we propose some adjustments to improve power against some more specific triggering alternatives. Second, conceptually similar methodology is developed to detect other sorts of dependence.

3.1 Time-limited $\tau$

We may be interested in limiting $\tau$ to a maximum range, $\tau_{\text{max}}$ say. This is useful if $\tau$ is expected to be small under $H_1$, in which case restricting $\hat{\tau} \leq \tau_{\text{max}}$ avoids ‘wasting’ power on testing for longer-term dependence.

The test statistic is computed as follows. Let $u_{\text{max}} = \rho\{T(\tau_{\text{max}})\}$. Modify Procedure 1 so that, if no $u_k \leq u_{\text{max}}$, the returned value is 1. Otherwise, replace $k$ with $\tilde{k} = \arg \max_{k=1,\ldots,n} \{\ell_k : u_k \leq \min(k/n, u_{\text{max}})\}$. 
To obtain a p-value of this test report \( p = 1 \) if \( T = 1 \). Otherwise compute \( o_1, \ldots, o_n \) as in (6), and calculate

\[
p = 1 - P[u_1 \geq \min(o_1, u_{\text{max}}), \ldots, u_n \geq \min(o_n, u_{\text{max}})],
\]

using Algorithm 1. Note that the p-value of this test has a distribution with a discrete component: it is exactly one whenever no events occur within \( \tau \) using Algorithm 1. Nevertheless for any \( \alpha \in [0, 1] \) we have \( P[p \leq \alpha] \leq \alpha \) under the null hypothesis, resulting in a conservative test.

3.2 Adaptive range

Consider the slightly modified alternative hypothesis

\[
\lambda_B(t) = \begin{cases} 
  \lambda_1 r(t) & \rho\{\{a(t), t-a(t)\}\} \leq \tau, \\
  \lambda_2 r(t) & \text{otherwise.}
\end{cases}
\]

(7)

In this model how long \( B \) is intensified following events of \( A \) is adapted to the null intensity of \( B \): \( B \) has longer to respond if relatively inactive according to \( r \).

Let \( b^*_i = \int_{b_i}^t r(v) dv, a^*_j = \int_{a_j}^t r(v) dv, i = 1, \ldots, n, j = 1, \ldots, m \). Then by the time change theorem \( B^* \) follows model (1) with \( A^* \{B^*\} \) replacing \( A \{B\} \), and \( r = 1, L = 1 \).

This observation provides a testing procedure for model (7). A test statistic \( T \) is computed as in Section 2.1, using as inputs \( A^* \) and \( B^* \) instead of \( A \) and \( B \), and using the Lebesgue measure for \( \rho \), that is, \( \rho(X) = \int_X 1 \, dx \). Like the original test, \( T \) has distribution \( F_n \) under \( H_0 \) conditional on \( n \).

3.3 Other types of association

The following tests simply replace the input to Procedure 1 previously \( u_1, \ldots, u_n \), by a sequence that also satisfies (4). The output is a likelihood ratio test for a different form of association, but which has the same null distribution, \( F_n \), conditional on \( n \).

**Correlation.** Relax the constraint \( a_i = 0 \). Let \( \tilde{a}(t) \) be the closest event to \( t \) in \( A \), which can now occur before or after \( t \). In (1), replacing \( t-a(t) \leq \tau \) by \(|t-\tilde{a}(t)| \leq \tau \) results in a model where the rate of events of \( B \) is locally increased around events of \( A \). A likelihood ratio test is obtained as follows. Let \( C(\tau) = \{t : |t-\tilde{a}(t)| \leq \tau\} \) and let \( v_1 \leq \ldots \leq v_n \) be the order statistics of \( \rho\{C(a_i - \tilde{a}(b_i))\} \), for \( i = 1, \ldots, n \). Compute \( T \) by inputting \( v_1, \ldots, v_n \) to Procedure 1. By a similar argument to the proof of Lemma 1 we find that \( v_1, v_2, \ldots \) are the event times of a point process following model (4). Thus \( T \) has distribution \( F_n \) under \( H_0 \), conditional on \( n \).

**Inhibition.** In model (1), constraining \( \lambda_1 \leq \lambda_2 \) and replacing \( t-a(t) \leq \tau \) by \( t-a(t) < \tau \) (to simplify analysis), we obtain a model for inhibition. By calculations analogous to those of Section 2 we find that the likelihood ratio test \( T \) for this model can be computed by simply inputting \( x_i = 1 - u_{n-i+1}, i = 1, \ldots, n \) in Procedure 1. As before \( T \) has distribution \( F_n \) under \( H_0 \), conditional on \( n \).

**Anti-correlation.** Putting \( \lambda_1 \leq \lambda_2 \) and replacing \( t-a(t) \leq \tau \) by \(|t-\tilde{a}(t)| < \tau \) we have a model where the rate of events of \( B \) are decreased around events of \( A \). To test for this, use \( y_i = 1 - v_{n-i+1}, i = 1, \ldots, n \) in Procedure 1.
Role reversal. For each of these forms of association analysis could conceivably be turned around by testing A given B. This only needs reversing time and swapping A and B in Section 2 (but note that an estimate of the null intensity of A is required). These role and time reversals result in different test statistics from those obtained by testing B, even though they target a similar type of association. It is an interesting question what circumstances could lead us to choose this approach.

4 Extension to more general point processes

In this section we consider relaxations of the non-homogeneous Poisson assumption in Section 2. Lemma 1 does not easily translate to a general point process if r in (1) is simply replaced by a conditional intensity (formally defined in Section 4.2). This also invalidates the results that follow, including the null distribution of the test. The difficulty is illustrated in a simple example in Appendix B where B is a point process with just one point uniformly distributed on [0, T), and A = {0, T/2}.

4.1 Independence conditional on n

A simple generalisation is possible if we can replace the intensity by a density over the observation region. Suppose that B can be generated by drawing n from some distribution and then placing the n event times independently according to some probability measure over [0, L). The non-homogeneous Poisson process is an example, as is the single-point point process mentioned above. Further remarks on the construction and properties of such processes are given in Daley and Vere-Jones (2003, Chapter 5). Analogously to model (1), suppose that the measure has a density

\[ d_B(t) \propto \begin{cases} 
\lambda_1 r(t) & t - a(t) \leq \tau, \\
\lambda_2 r(t) & \text{otherwise}, 
\end{cases} \]

(8)

where r is as before a bounded non-negative Lebesgue measurable function satisfying

\[ \int_0^L r(v) dv = 1, \]

thereby defining \( \rho \). If n has a Poisson distribution this model reduces to (1). The hypothesis test \( H_0 : \lambda_1 = \lambda_2 \) versus \( H_1 : \lambda_1 > \lambda_2 \) can be evaluated through a similar generalised likelihood ratio test: we first compute \( u_1, \ldots, u_n \) as the order statistics of \( \rho(T(b_i - a(b_i))) \), \( i = 1, \ldots, n \) and then T using Procedure 1. By straightforward modifications to the proof of Lemma 1 we find:

Lemma 2. Under model (8), conditional on n, \( \{u_i\} \) is a set of IID random variables with support on [0, 1) and density

\[ d(x) \propto \begin{cases} 
\lambda_1 & x \leq \rho(T(\tau)), \\
\lambda_2 & \text{otherwise}. 
\end{cases} \]

(9)

From this we establish that T given n also has null distribution \( F_n \) under model (8).

4.2 Random time change

A much wider class of point processes can be conceived by allowing the intensity at time t, previously a deterministic quantity, to be dependent on the past. Formally, information available at time t is captured by an increasing family of \( \sigma \)-algebras \( \mathcal{F} = \{ \mathcal{F}_s : 0 \leq s < \infty \} \) with \( \mathcal{F}_s \subseteq \mathcal{F}_t \) for \( 0 \leq s \leq t < \infty \). If B is adapted to a right-continuous history \( \mathcal{F} \), by the
Doob-Meyer decomposition there exists a right-continuous predictable process $\Lambda(t)$ such that for every $i \in \mathbb{N}$ the stopped process $\{B(t \wedge b_i) - \Lambda(t \wedge b_i)\}$ is an $\mathcal{F}$-martingale. This process is called the compensator for $B$. Under mild regularity conditions the conditional intensity $\lambda(t)$ of $B$ can then be defined via $\Lambda(t) = \int_0^t \lambda(x)dx$ (Daley and Vere-Jones, 2007, p.358, p.367, p.390).

Note the following result, more general than that needed in the proof of Lemma 1, known as the random time change theorem (Daley and Vere-Jones, 2007, p.421): if $B$ is non-terminating (there are infinite events as $t \to \infty$) and $\Lambda_0$ is continuous then the process $\bar{B}(t) = B(\Lambda_0^{-1}(t)), t \in [0, \infty)$ is a homogeneous Poisson process with unit rate. (Recall our definition $F^{-1}(y) = \inf \{x : F(x) \geq y\}$ for a non-decreasing $F$.)

Suppose then that under the null hypothesis $B$ has a conditional intensity $\lambda_0$. The random time change theorem allows construction of a process $\bar{B} : t \in [0, \Lambda_0(L))$, which, under the null hypothesis, is a stopped unit-rate homogeneous Poisson process. To compare like for like, let $\bar{A}(t) = A(\Lambda_0^{-1}(t))$.

Since we cannot get exact distributional results when we replace $r$ in (1) by $\lambda_0$ a natural alternative might be to test for $\bar{A}$ triggering $\bar{B}$ via the model

$$\lambda_{\bar{B}}(t) = \begin{cases} \lambda_1/\Lambda_0(L) & t - \bar{a}(t) \leq \tau, \\ \lambda_2/\Lambda_0(L) & t - \bar{a}(t) > \tau, \end{cases} \quad t \in [0, \Lambda_0(L)), \quad (10)$$

where $\bar{a}(t)$ is the most recent event in $\bar{A}$ occurring at or before $t$. A test $\tilde{T}$ for this model is obtained by computing the corresponding $u_i$ and then implementing Procedure 1 replacing $b_i$ by $\Lambda_0(b_i), a_j$ by $\Lambda_0(a_j), L$ by $\Lambda_0(L)$ and finally $r(t)$ by the constant $1/\Lambda_0(L)$.

Although conceptually similar to the test described in Section 3.2 where $\tau$ was adapted the null intensity of $B$, the dependence of the stopping-time $\Lambda_0(L)$ on $B$ under $H_0$ makes inference more complicated. In particular Lemma 1 and the $p$-value computed in Section 2.3 no longer hold exactly, but may be sufficiently close approximations for practical use. Progress is possible if we allow $L$ to be random, but this seems contrived.

5 Asymptotic distribution

In this section we demonstrate an asymptotic connection between our test and a weighted Kolmogorov-Smirnov (K-S) test. An especially surprising aspect of this result is that the weight function needed is exactly that which is most often used for K-S tests, e.g. in Chicheportiche and Bouchaud (2012). Note that classical asymptotic results about the likelihood ratio test do not hold here because of the discontinuity of the likelihood function at $\tau$ under $H_1$ (Müller and Wang 1994, p.227).

Let $y_1, \ldots, y_n$ be independent replicates of an absolutely continuous random variable $Y$ with support $\mathbb{Y} \subseteq \mathbb{R}$ and CDF $F$, and consider their empirical CDF

$$\hat{F}(y) = \frac{1}{n} \sum_{i=1}^N \mathbb{I}[y_i \leq y].$$

A generalized K-S test (Anderson and Darling 1952) can be used to test whether $\hat{F}$ is consistent with a hypothesized distribution $F$, 

$$G = \sup \left\{ \sqrt{n}(\hat{F}(y) - F(y))\sqrt{\phi(F(y))} : y \in \mathbb{Y} \right\},$$
Figure 1: Empirical CDFs based on 1000 replicates of $T_{[γ_1, γ_2]}$ (solid line) and the weighted upper K-S test (dashed line), with $n = 1000$ and $[γ_1, γ_2] = [.01, .99]$.

for some weight function $φ(x) ≥ 0, x ∈ [0, 1]$. We need a one-sided, interval-restricted version of the above,

$$G_{[γ_1, γ_2]}^+ = \max \left\{ \sqrt{n} \left| \frac{\hat{F}(y_i) - F(y_i)}{\sqrt{φ(F(y_i))}} \right| : F(y_i) ≤ \hat{F}(y_i); γ_1 ≤ F(y_i) ≤ γ_2 \right\} \quad (11)$$

for $0 < γ_1 < γ_2 < 1$, defining $G_{[γ_1, γ_2]}^+$ to be 0 if the set above is empty. In the theory of K-S tests, it is often a consequence of $\hat{F}$ being right-continuous and $φ$ satisfying a specific property given in Anderson and Darling [1952, Eq 2.8] that attention can be restricted to the value of $F$ at the observed points. This is not the case here: the presence of $γ_1$ and $γ_2$ forces us to define $G_{[γ_1, γ_2]}^+$ as (11).

Analogously, let

$$T_{[γ_1, γ_2]} = \max \{ \ell_i : u_i ≤ i/n; γ_1 ≤ u_i ≤ γ_2 \}$$

now setting $T_{[γ_1, γ_2]} = 1$ if the set above is empty.

We next loosely demonstrate the following: if $λ_1 = λ_2$ then as $n → ∞$, $\sqrt{2n}(T_{[γ_1, γ_2]} - 1)^{1/2}$ given $n$ converges in distribution to the limiting distribution as $n → ∞$ of $G_{[γ_1, γ_2]}^+$ with $φ(p) = [p(1 - p)]^{-1}$. Similar results can be found in Matthews et al. [1985, Section 4], Akman and Raftery [1986, Section 3] and Loader [1991, Appendix], however we have not found a statement that is equivalent.

Let $n, i → ∞$ with $i/n = p$. Then if $p ∈ [γ_1, γ_2]$, $u_i$ is asymptotically normal with mean $p$ and variance proportional to $1/n$ and therefore $u_i = p + O_p(1/√n)$. Viewed as a function of $u_i$, $\ell_i$ has first and second derivatives 0 and $p - 1/(p(1 - p))$. Therefore by Taylor expansion we find $\ell_i = 1 + (p - u_i)^2/[2p(1 - p)] + o_p(1/n)$ [Davis and Brockwell 1991, Prop. 6.1.5]. Hence

$$\sqrt{2n}(\ell_i - 1)^{1/2} = \sqrt{n}|p - u_i|/\sqrt{p(1 - p)} + o_p(1).$$

Let $S = \{ i : u_i ≤ i/n; γ_1 ≤ u_i ≤ γ_2 \}$. Heuristically ignoring the influence of $o_p(1)$ terms
(of which there are a growing number with \( n \)), assume
\[
\sqrt{2n} |T_{[\gamma_1, \gamma_2]} - 1|^{1/2} = \max \left\{ \sqrt{2n} |\ell_i - 1|^{1/2} : i \in S \right\}
\]

\[\to d \max \left\{ \sqrt{n} (p - u_i) / \sqrt{p(1 - p)} : i \in S \right\},\]

defining the RHS to be zero if \( S \) is empty. The absolute value was removed because \( i \in S \) guarantees \( u_i \leq p \). Now, use \( p = \hat{F}(y_i) \) and \( F(y_i) - u_i \) in the numerator and \( \hat{F}(y_i) \to p \) in the denominator to find that the RHS above is also the limiting distribution of \( G^+_{[\gamma_1, \gamma_2]} \).

To prove this more rigorously we would need a better understanding of the joint behaviour of the \( o_p(1) \) terms (if arbitrary random variables \( x_1, \ldots, x_m \) are each individually \( o_p(1) \) then \( \max(x_i) \) is not \( o_p(1) \) in general if \( m \) increases with \( n \)), perhaps reasoning along the lines of [Donsker, 1952] (on replacing a sequence or order statistics of a Uniform random variable by its limiting process in the asymptotic theory of K-S tests).

By simulation it seems that result does not hold in the limit \( \gamma_1 \to 0, \gamma_2 \to 1 \). This is not necessarily surprising — for instance it is noted in [Chicheportiche and Bouchaud, 2012] that with \( \gamma_1 = 1/(n + 1) \) and \( \gamma_2 = n/(n + 1) \) the asymptotic distribution of the two-sided version of \( G_{[\gamma_1, \gamma_2]} \) still depends on \( n \), and in fact the asymptotic theory of weighted K-S tests generally relies on \( \phi \) being bounded over the unit interval (Anderson and Darling, 1952, p.196), which \( [p(1 - p)]^{-1} \) is not.

6 Example: information flow in the Enron email corpus

The Enron email corpus is a dataset that comprises emails sent and received by about 150 senior executives at the Enron Corporation, over the period 1998 to 2002. Although it is well-known to suffer from various integrity problems, it makes an attractive real data example because it is publically available and many contemporary readers will be familiar with emailing behaviour. The dataset we analyse was downloaded from [http://bailando.sims.berkeley.edu/enron_email.html](http://bailando.sims.berkeley.edu/enron_email.html) and reprocessed for our application. Only emails sent during the year 2001 were retained, because the record appears to be cleanest for that year. Some further effort was then needed to obtain reliable data. For example, many different email addresses can correspond to the same identity since an individual, John Smith say, could appear as any of john.smith, x..smith, jsmith @ either enron.com or ect.enron.com and more.

Following [Perry and Wolfe, 2013] we discarded emails sent to more than 5 recipients, a subjectively chosen threshold that allows us to focus on inter-personal communications rather than company-wide announcements.

Results will be presented for an individual, hereafter identified as \( o \), who emailed frequently over the year, and for whom there are 12 individuals (of the 150 above) who contact \( o \) and that \( o \) contacts back. These are referred to by the identifiers \( 1, \ldots, 12 \).

We shall seek to determine whether \( i \to o \) triggers \( o \to j \), denoted \( i \to o \to o \to j \), using only the timing of events. When \( i = j \), a significant test is evidence for reciprocation (or \( o \) responding to emails), otherwise it suggests information flow. The point processes generated by \( i \to o \) and \( o \to j \) replace \( A \) and \( B \) respectively in Section 2.

6.1 Bayesian null model

To use the results of Section 2 we first require \( r \) in (1). In this section we propose to estimate this function via a generative model, described next.
Let $D = \{t_1, \ldots, t_N\}$ denote the times of all of the emails sent by $o$ and $D_j$ the times specific to a recipient $j$. For simplicity $D_j$ is modelled as a completely random subset of $D$, meaning that it can be generated, under the null, by randomly discarding each element of $D$ with a fixed probability.

Figure 2 presents o’s sent email times, with the y-axis showing the day and x-axis the time of day of each event. This brings out a daily pattern in emailing behaviour; for example between the hours of 1600 and midnight $o$ is markedly less busy, a period of 8 hours where $o$ is presumably away from work (the timezone in which he was operating is not known to us). On the other hand, looking vertically, there some suggestion of behavioural change at a larger time-scale. For example there appears to be a busy period around the middle of year (specifically, July). This visual analysis invites us to separate the likelihood for $D$ as

$$p(D) = p_1(N_1, \ldots, N_{365}) p_2(d_1, \ldots, d_N),$$

where $N_k$ is the number of events in day $k$ and $d_i = t_i \mod 24, i = 1, \ldots, N$ are the events collapsed onto a day.

**Model for the number of events per day.** Assume $n_i \sim \text{Poisson}(\lambda(i))$, independently, where $\lambda(t)$ is a step-function generated by

1. Placing $x$ steps $k_1, \ldots, k_x$ at random on $(1, 365)$, with $x \sim \text{Poisson}(\nu_1)$. Let $k_0 = 1$ and $k_{x+1} = 365$.

2. Choosing $x + 1$ heights $\gamma_1, \ldots, \gamma_{x+1}$ independently with $\gamma_i \sim \text{Gamma}(\alpha, \beta)$, and setting $\lambda(t) = \gamma_j$ for $t \in [k_{j-1}, k_j), j = 1, \ldots, x + 1$.

The hyper-parameters were set to $\nu_1 = 1, \alpha = 1$ and $\beta = .1$. 

![Emailing Behaviour of an Individual in the Enron Dataset](image-url)
Figure 3: Emailing behaviour of an individual in the Enron dataset: fitted intensity

**Model for the time-of-day.** Assume \( d_i \sim \mathcal{F} \) independently, where \( \mathcal{F} \) has a piecewise constant density \( f \) on \([0, 24]\) generated by

1. Placing \( y \) steps \( l_1, \ldots, l_y \) at random on \((0, 24)\), with \( y \sim \text{Poisson}(\nu_2) \). Let \( l_0 = 0 \) and \( l_{y+1} = 24 \).
2. Generating a vector of masses \( \phi_1, \ldots, \phi_{y+1} \) jointly from Dirichlet\((\frac{k_1-k_0}{24}, \ldots, \frac{k_{y+1}-k_y}{24})\).
3. Letting \( f(t) = \frac{\phi_j}{l_j - l_{j-1}} \) for \( t \in [l_{j-1}, l_j), j = 1, \ldots, y + 1 \).

\( \nu_2 \) was set to 2.

An advantage of this approach is a posteriori \( \lambda \) and \( \mathcal{F} \) are independent, and so posterior samples for each can be obtained in completely separate simulations. In each case these were obtained using reversible-jump Markov Chain Monte Carlo (Green, 1995), using one million samples and retaining every 100th term. (The sample is thinned because the computation below is more expensive than a sampling step.)

To obtain an estimate of \( r \), we proceed as follows. We take the average \( \bar{R} \) of

\[
R(t) = R_n[d(t)] + R_d(t \mod 24) \left( R_n[d(t) + 1] - R_n[d(t)] \right),
\]

where \( d(t) \) is the day on which \( t \) occurs, and \( R_n \) and \( R_d \) are formed from individual posterior draws of \( \lambda \) and \( f \) as

\[
R_n(i) = \frac{\sum_{j=1}^{i-1} \lambda(j)}{\sum_{j=1}^{365} \lambda(j)}; \quad R_d(t) = \int_0^t f(y) \, dy,
\]

for \( i = 2, \ldots, 365, t = [0, 24] \), defining \( R_n(1) = 0 \). Then we set \( \hat{\rho}[a, b] = \bar{R}(a) - \bar{R}(b), \hat{r}(t) = d\bar{R}/dt \).

Figure 3 illustrates our model fit to o’s emailing behaviour. The crosses denote event times, now on the \( x \)-axis. For visibility purposes only data from the first 30 days are
presented. The line is $\hat{r}(t)$, which we will use in place of $r$ in (1). The model finds a unimodal daily pattern and, for instance, a period of high activity between 20th and the 25th of January.

Figure 4: Evidence of $i \rightarrow o \iff o \rightarrow j$, for $i,j = 1,...,12$. The black circles indicate results retained by FDR set at 10%, the half-circles $p$-values not retained by FDR but still smaller than 0.05, while the white circles indicate a p-value greater than 0.05.

Next we turn to detecting triggering effects between $o$'s sent and received emails. In Figure 4, the results of testing $i \rightarrow o \iff o \rightarrow j$ are shown for $i,j = 1,...,12$, limiting the range of $\tau$ to $\tau_{max} = 1$ week, as discussed in Section 3.1. The black circles are the p-values that are retained for analysis if we use a false discovery rate (FDR) of 10% (Benjamini and Hochberg, 1995). The half-circles are for $p$-values less than 5% and white are the (non-significant at the 5% level) $p$-values that remain.

Most of the entries on the bottom-left to top-right diagonal are black, meaning that there is compelling evidence for reciprocation. Because reciprocation is largely to be expected, the two white circles on that diagonal warrant additional inspection: they indicate a lack of evidence for $o$ responding to emails from either 7 or 10.

$7 \rightarrow o \not\iff o \rightarrow 7$ There is only one email 7 to o and one other from o to 7. They are sent about one month apart (and appear to be unrelated judging by their subject-lines).

The p-value is automatically 1 because $\tau_{max}$ was set to a week.

$10 \rightarrow o \not\iff o \rightarrow 10$ This example is more interesting. The $p$-value is only 0.28 despite there being 14 emails from 10 to o and 9 from o to 10, the most coincidental email times falling in July, about $3\frac{1}{2}$ hours from each other. The reason why no effect is detected is in part because o is estimated to be relatively busy in July, with $\rho$(July) $\approx 0.13$ as opposed to the average $1/12 \approx 0.08$, meaning that we are less sensitive to coincidental timings during that month than at other times. In fact, upon inspecting the subject-lines of $10 \rightarrow o$ and $o \rightarrow 10$, it does appear as if 10 and o do not reciprocate. For example, the subject-lines of the two most coincidental emails are “FW: Enron Complaint” and “Dunn hearing link?”, not obviously related.
Consider now the p-values that were retained by the FDR procedure. Table 1 shows the subject-lines of the emails that the test based its decision on. More precisely, for each of the retained $i, j$ pairs, we find the closest two $i \rightarrow o$ and $o \rightarrow j$ events, subject to the former preceding the latter. Thus we have the ‘most triggering’ email event $e$, and its subject-line is displayed first. Next, we display the subject-lines of all the emails from $o$ to $j$ that fall within $[e, e + \hat{\tau}]$, for the $\hat{\tau}$ used by the test.

Table 1 shows that the method succeeds in picking out ‘real’ excitation periods. Consider for example the most significant detection $11 \rightarrow o \xrightarrow{\sim} o \rightarrow 11$. 4 emails fall within $\hat{\tau}$ of the ‘most triggering’ email. These all have the subject-line “RE: DWR - Gas Daily” corresponding quite compellingly to the subject-line of the original email by 11, “DWR - Gas Daily”. Furthermore, the next email in the $o \rightarrow 11$ sequence has a different subject-line, “RE: DWR and Edison Meetings”, suggesting that $\hat{\tau}$ fits well.

Although the method largely found evidence of reciprocation, there are some places where real information flow was identified. The email from 9 to $o$ with subject “California Update–Legislative Push Underway” being followed by a string of emails from $o$ to 3 with subject “Re: California Update–Legislative Push Underway” is a particularly compelling example. A concern could be that 3 was simply been ‘cc’ed’ while $o$ was responding to 9. This is not the case: two of the four emails displayed, the second and the fourth, are sent from $o$ directly to 3 with no other party involved. For reference, there were 28 emails in $[e, e + \hat{\tau}]$. Only those with the matching subject-lines are displayed. The vertical dots indicate the position of the omitted emails.

There are admittedly what appear to be false positives in the results, notably the detection $5 \rightarrow o \xrightarrow{\sim} o \rightarrow 4$. Although there are over 40 hours separating the ‘most triggering email’ from the two emails that are considered ‘triggered’, these are the only two emails from $o$ to 4 that occur after the first email from 5 to $o$, about half a year earlier. The problem is that our model for $r$ does not really capture ‘bursty’ behaviour — our intuition tells us that the second email, just five hours later, is very likely to be related to the first (and the subject-lines suggest this) rather than adding much evidence to $5 \rightarrow o$ triggering $o \rightarrow 4$. The step-function that is fit, however, looks for changes on a much larger time scale. Thus we postulate that a more tailored model for $r$, especially one that captured these local bursts of activity would improve detections.

7 Conclusion

This article proposes a family of tests for association between two point processes, and analyses their finite-sample and asymptotic properties. Depending on implementation details the procedure can be used to determine different forms of association, including triggering, correlation, inhibition or anti-correlation.

It is remarkable that such a simple and exact technique should exist and not have been discovered until now. Many exciting extensions are now possible. In the analysis of dynamic network data, it may be of interest to find paths or other motifs, rather than simple pairwise dependence. Here the difficulty would be computational as well as statistical, because of the potential complexity of the search. Another important addition would be to incorporate information associated to each point in order to better identify ‘pairings’ between the events of both processes.

This work also leads to an interesting probabilistic question. Consider that the approach relies on the self-superposition and time-transformation of point processes — different tests are obtained depending on the order in which these operations are performed. To extend results to situations where the null intensity is estimated from the data, the key
Table 1: Subject-lines of sent emails that are estimated to be triggered. Further details in main text.

| Pattern | P-value | Time lag | Subject |
|---------|---------|----------|---------|
| 11 → j → j → 11 | $9.6 \times 10^{-12}$ | 0:00:00 | DWR - Gas Daily |
| 11 → j → j → 2 | $4.7 \times 10^{-9}$ | 0:00:00 | RE: CPUC Questions on DA |
| 12 → j → j → 12 | $1.5 \times 10^{-6}$ | 0:00:00 | RE: CA Unbundling |
| 2 → j → j → 3 | $4.6 \times 10^{-5}$ | 0:00:00 | California Update–Legislative Push Underway |
| 9 → j → j → 3 | $4.6 \times 10^{-5}$ | 0:00:00 | California Update–Legislative Push Underway |
| 1 → j → j → 1 | $4.7 \times 10^{-5}$ | 0:00:00 | Re: Comments to Gov’s Proposals |
| 3 → j → j → 3 | $9.3 \times 10^{-5}$ | 21:37:00 | Update from EES Call this Morning |
| 9 → j → j → 9 | $3.7 \times 10^{-4}$ | 0:00:00 | California Update–Legislative Push Underway |
| 2 → j → j → 6 | $8.6 \times 10^{-4}$ | 0:00:00 | HERE IS MY DRAFT |
| 6 → j → j → 6 | $1.7 \times 10^{-3}$ | 0:00:00 | Re: FW: SoCalGas Capacity Forum |
| 5 → j → j → 5 | $2.1 \times 10^{-3}$ | 0:00:00 | Re: Response to ORA/TURN petition |
| 8 → j → j → 5 | $2.2 \times 10^{-3}$ | 0:00:00 | RE: Call to Discuss Possible Options to Mitigate Ef... |
| 4 → j → j → 10 | $2.4 \times 10^{-3}$ | 0:00:00 | RE: Transwestern Hearing |
| 8 → j → j → 8 | $3.8 \times 10^{-3}$ | 1:31:13 | Attorneys |
| 5 → j → j → 4 | $7.1 \times 10^{-3}$ | 0:00:00 | FW: EPSA report |
| 2 → j → j → 11 | $7.5 \times 10^{-3}$ | 0:00:00 | Willie Brown INFO |
| 7 → j → j → 1 | $8.4 \times 10^{-3}$ | 0:00:00 | Governor DavisPress conference Highlights – wil... |
| 6 → j → j → 5 | $9.1 \times 10^{-3}$ | 2:22:00 | FW: SoCalGas Capacity Forum |
obstacle seems to be characterizing point processes that have been time-transformed using an estimated conditional intensity. In particular it would be interesting to investigate conditions under which the resulting points would be more evenly dispersed than Poisson, in some sense to be formalised, and whether this could allow conservative testing.

A On the maximum likelihood estimates

Procedure is established as follows. Firstly, any \( \tau \) for which \( K(\tau)/n < \rho\{T(\tau)\} \) can be rejected. To see this, note that for any given \( \tau \), \( K(\tau)/n \) is maximised over \( \lambda_1 \geq 0, \lambda_2 \geq 0 \) at \( \lambda_1^* = K(\tau)/\rho[|T(\tau)|] \) and \( \lambda_2^* = [n - K(\tau)]/\rho[|T(\tau)|] \), giving

\[
\left( \frac{K(\tau)/n}{\rho[|T(\tau)|]} \right)^{K(\tau)/n} \left( \frac{(n - K(\tau))/n}{1 - \rho[|T(\tau)|]} \right)^{n-K(\tau)/n}.
\]

These estimates must be rejected if \( \lambda_1^* < \lambda_2^* \), that is, if \( K(\tau)/n < \rho\{T(\tau)\} \). In this case, the maximum over the restriction \( \lambda_1 \geq \lambda_2 \) must satisfy \( \lambda_1^* = \lambda_2^* \) because, while \( \lambda_1^* > \lambda_2^* \), we can continue to increase the likelihood by bringing at least one of \( \lambda_1^* \), \( \lambda_2^* \) closer to \( K(\hat{\tau})/n \) or \( (n - K(\hat{\tau}))/n \), respectively. But if we can only consider solutions where \( \lambda_1^* = \lambda_2^* \), then we are effectively fitting the single parameter model from the null hypothesis. That is automatically worse than choosing \( \tau \) just large enough that \( T(\tau) \) contains all of the events, where we would estimate \( \lambda_1^* = n, \lambda_2^* = 0 \).

The second point is that amongst the \( \tau \)s that satisfy \( K(\tau)/n \geq \rho\{T(\tau)\} \), the likelihood is always improved by shifting \( \tau \) towards the closest \( b_i - a(b_i) \) on the left of it and refitting \( \lambda_1, \lambda_2 \), essentially producing a ‘sharper fit’. Thus, we can restrict our search on \( \tau \) to a subset of \( b_1 - a(b_1), \ldots, b_n - a(b_n) \) or equivalently restrict our search on \( \rho\{T(\tau)\} \) to \( u_1, \ldots, u_n \) defined in the main text, leading to the solution given.

B Superposition based on conditional intensity

Let \( L = 1, A = \{0, 1/2\} \) and consider a point process \( B \) with just one point uniformly distributed on \( [0, 1) \). Of course Lemma cannot hold since there can only be one \( u \)-variable whereas for \( \lambda_1 \geq \lambda_2 > 0 \) the Lemma would admit any number with positive probability. Nevertheless, to illustrate the problem of extending the Lemma to a general point process, consider the distribution of \( u \) if \( r \) in (I) is replaced by the point process conditional intensity \( 1/(1 - t) \), with \( \lambda_1 = \lambda_2 = 1 \). Let \( v \) be a uniform random variable over \( [0, 1/2) \). We find that \( u \) is distributed as a mixture \( -\log(1 - v) \) (an exponential variable conditioned on being smaller than \( \log 2 \)) with probability \( 1/2 \) and \( -\log(1 - v)^{1-2v} \) (a non-standard distribution).

References

Akman, V. and Raftery, A. (1986). Asymptotic inference for a change-point Poisson process. The Annals of Statistics, 14(4):1583–1590.

Anderson, T. and Darling, D. (1952). Asymptotic theory of certain “goodness of fit” criteria based on stochastic processes. The Annals of Mathematical Statistics, 23(2):193–212.
Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 289–300.

Berman, M. (1986). Testing for spatial association between a point process and another stochastic process. *Applied Statistics*, pages 54–62.

Blundell, C., Heller, K., and Beck, J. (2012). Modelling reciprocating relationships with Hawkes processes. In *NIPS*, pages 2609–2617.

Brillinger, D. (1988a). Maximum likelihood analysis of spike trains of interacting nerve cells. *Biological Cybernetics*, 59(3):189–200.

Brillinger, D. (1988b). Some statistical methods for random process data from seismology and neurophysiology. *The Annals of Statistics*, pages 1–54.

Brillinger, D. (1992). Nerve cell spike train data analysis: a progression of technique. *Journal of the American Statistical Association*, 87(418):260–271.

Brillinger, D., Bryant Jr, H., and Segundo, J. (1976). Identification of synaptic interactions. *Biological Cybernetics*, 22(4):213–228.

Brillinger, D. and Segundo, J. (1979). Empirical examination of the threshold model of neuron firing. *Biological Cybernetics*, 35(4):213–220.

Chicheportiche, R. and Bouchaud, J.-P. (2012). Weighted Kolmogorov-Smirnov test: Accounting for the tails. *Physical Review E*, 86(4):041115.

Daley, D. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes, Volume 1 (2nd ed.)*. Springer, New York.

Daley, D. and Vere-Jones, D. (2007). *An introduction to the theory of point processes: volume II: general theory and structure*, volume 2. Springer.

Davis, R. and Brockwell, P. J. (1991). Time series: theory and methods.

Donsker, M. (1952). Justification and extension of Doob’s heuristic approach to the Kolmogorov-Smirnov theorems. *The Annals of Mathematical Statistics*, 23(2):277–281.

Doss, H. (1989). On estimating the dependence between two point processes. *The Annals of Statistics*, 17(2):749–763.

Green, P. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4):711–732.

Hawkes, A. (1971a). Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 438–443.

Hawkes, A. (1971b). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90.

Loader, C. R. (1991). Inference for a hazard rate change point. *Biometrika*, 78(4):749–757.

Lotwick, H. and Silverman, B. (1982). Methods for analysing spatial processes of several types of points. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 406–413.
Matthews, D., Farewell, V., and Pyke, R. (1985). Asymptotic score-statistic processes and tests for constant hazard against a change-point alternative. *The Annals of Statistics*, 13(2):583–591.

Müller, H.-G. and Wang, J.-L. (1994). Change-point models for hazard functions. *Lecture Notes-Monograph Series*, pages 224–241.

Noé, M. (1972). The calculation of distributions of two-sided Kolmogorov-Smirnov type statistics. *The Annals of Mathematical Statistics*, pages 58–64.

Noé, M. and Vandewiele, G. (1968). The calculation of distributions of Kolmogorov-Smirnov type statistics including a table of significance points for a particular case. *The Annals of Mathematical Statistics*, 39(1):233–241.

Perkel, D., Gerstein, G., and Moore, G. (1967). Neuronal spike trains and stochastic point processes: II. simultaneous spike trains. *Biophysical journal*, 7(4):419–440.

Perry, P. and Wolfe, P. (2013). Point process modelling for directed interaction networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

Ripley, B. (1976). The second-order analysis of stationary point processes. *Journal of applied probability*, pages 255–266.

Ripley, B. (1977). Modelling spatial patterns. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 172–212.

Worsley, K. (1988). Exact percentage points of the likelihood-ratio test for a change-point hazard-rate model. *Biometrics*, pages 259–263.