PSYCHOMETRIKA—VOL. 78, NO. 4, 793–814
OCTOBER 2013
DOI: 10.1007/S11336-013-9329-1

MODELLING DYADIC INTERACTION WITH HAWKES PROCESSES

PETER F. HALPIN AND PAUL DE BOECK

UNIVERSITY OF AMSTERDAM

We apply the Hawkes process to the analysis of dyadic interaction. The Hawkes process is applicable to excitatory interactions, wherein the actions of each individual increase the probability of further actions in the near future. We consider the representation of the Hawkes process both as a conditional intensity function and as a cluster Poisson process. The former treats the probability of an action in continuous time via non-stationary distributions with arbitrarily long historical dependency, while the latter is conducive to maximum likelihood estimation using the EM algorithm. We first outline the interpretation of the Hawkes process in the dyadic context, and then illustrate its application with an example concerning email transactions in the work place.

Key words: dyadic interaction, event sampling, Hawkes processes, EM algorithm, maximum likelihood.

1. Introduction

This paper addresses the time series analysis of event data collected within dyads. Recent advances in the study of dyadic interaction have focused mainly on continuous outcomes rather than event data (e.g., Card, Selig, & Little, 2008; Chow, Ferrer, & Hsieh, 2010; Ferrer, Steele, & Hsieh, 2012; Gottman, Murray, Swanson, Tyson, & Swanson, 2003; Hamaker, Zhang, & van der Maas, 2009; Kenny & Ledermann, 2010). When event data are treated, they are usually analysed only with respect to the sequence of events and not their timing *per se* (e.g., Bakeman & Quera, 2011; Bakeman & Gottman, 1997; Kenny, Kashy, & Cook, 2006, Chapter 14; Magnusson, 2005). The approach taken in this paper contributes to this research by explicitly modelling the timing of events.

We focus on events that can be treated as having negligible duration relative to the period of observation, which we refer to as *occurrences*. Occurrences may be contrasted with events that are treated as extended in time (e.g., states, regimes). Examples of occurrences over the course of a single conversation include types of statements (e.g., criticism, agreement), non-verbal behaviours (e.g., gestures, facial expressions), the onset of physiological responses (e.g., increased heart rate, perspiration), and so on. Examples of occurrences over the course of a relationship include entire conversations (e.g., face-to-face interactions, telephone calls) and specific ways of interacting (e.g., conflict, intimacy). In the domain of social media, occurrences include emails, instant messages, status updates, and tweets. Dyadic interaction also may be construed more broadly, to include, say, how a student interacts with an intelligent tutoring system, with occurrences then including item responses and hint requests.

The intuitive idea is to model the Bernoulli probability of an event happening in a small window of time $[t, t + \Delta)$, conditional on the events that have happened before $t \in \mathbb{R}+$. This is referred to as the *instantaneous probability* of an event and is denoted $p(t)$. We give a formal definition of $p(t)$ below. In the dyadic context, the rationale behind modelling $p(t)$ is to describe how the probability of each person's behaviour changes in continuous time, how this depends on

Peter F. Halpin is now at New York University; Paul De Boeck is now at Ohio State University.

Requests for reprints should be sent to Peter F. Halpin, Department of Humanities and Social Science in the Professions, New York University, 246 Greene St, Office 316E, 10013-6677, New York, USA. E-mail: peter.halpin@nyu.edu

© 2013 The Psychometric Society

previous behaviours, and the extent to which the ebb and flow of activity of two individuals is coordinated in time.

The formal framework is provided by the theory of point processes (e.g., Daley & Vere-Jones, 2003). Point processes have been applied in many fields of study including seismology (e.g., Ogata, 1988), ecology (e.g., Law, Illian, Burslem, Gratzer, Gunatilleke, & Gunatilleke, 2009), neuroscience (e.g., Dayan & Abbott, 2005; Truccolo, Eden, Fellows, Donoghue, & Brown, 2005), and human dynamics (e.g., Crane & Sornette, 2008). We focus on linear systems analysis (e.g., Brillinger, 1975) and specifically on a bivariate version of the Hawkes process (Hawkes, 1971; Hawkes & Oakes, 1974). The Hawkes process is a general model for self-excitatory events, meaning that it is appropriate for data that are clustered (i.e., over-dispersed relative to the homogeneous Poisson process). In the dyadic context, clustering describes interactions in which the individuals are mutually responsive. We outline this interpretation in more detail below.

Despite the wide application of specific types of point processes (e.g., Poisson, renewal), the development of general parametric models continues to pose technical problems. In particular, the log-likelihood of the Hawkes process contains the logarithm of a weighted sum of probability density functions, which leads to numerical optimization problems such as ill-conditioning and local maxima. Recent approaches have addressed these issues using the so-called "branching structure" representation of the Hawkes process (e.g., Rasmussen, 2011; Veen & Schoenberg, 2008). In intuitive terms, the branching structure allows us to rewrite the Hawkes process as a mixture of Poisson processes, thereby making its treatment analogous to that of finite mixture models. Using this approach, we formulate an EM algorithm for maximum likelihood estimation; the main details of which were first laid out by Veen and Schoenberg (2008).

The next section deals with specification of the Hawkes process and its interpretation in the context of dyadic interaction. We then discuss maximum likelihood estimation and subsequently apply the model to an example from interpersonal communication in the workplace. The data consist of email transactions between an employee and employer. Email records have featured prominently in the analysis of human dynamics (e.g., Barabási, 2005; Eckmann, Moses, & Sergi, 2004; Kalman, David, & Rafaeli, 2006), and our example provides a novel viewpoint on this topic. It also allows for illustration of known results concerning asymptotic inference and goodness of fit. The conclusion addresses limitations and extensions, and some mathematical details are collected in Appendix.

2. Model Specification and Interpretation

Our approach to dyadic interaction is presented schematically in Figure 1. The circles represent members of a dyad (A and B) and the arrows correspond to three types of occurrences. It should be emphasized that the classification of occurrences into one of the three types is to be inferred from the temporal structure of the data—it is not assumed to be known beforehand. Arrows 1A and 1B denote a person's spontaneous actions. These are events that are not in response to anything occurring previously within the observation period. We model the instantaneous probability of spontaneous action as a base rate parameter that is constant over time. Arrows 2A and 2B represent a person's responses to their own actions, while arrows 3A and 3B represent a person's responses to the actions of their partner. We model the instantaneous probability of each of these responses as a non-linear function of time with parameters describing (a) overall intensity, (b) duration, and (c) the shape of the response curve. We refer to this approach in terms of a dyadic responsiveness (DR) model.

The remainder of this section presents two equivalent ways of formalizing the DR model. The first represents the conditional intensity function (CIF) of a bivariate point process in terms

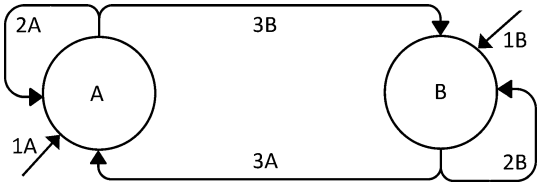


FIGURE 1.
Schematic representation of dyadic interaction. *Circles A and B* indicate members of a dyad. *Arrows* indicate various aspects of their interaction: *1* denotes spontaneous actions, which are treated as constant over time; *2* denotes responses to one's own previous actions; *3* denotes responses to the actions of one's partner. The response functions are treated as non-linear functions of time. See the text for further details.

of linear systems analysis, which provides the more general framework. After applying the restrictions required by the Hawkes process, the second representation is a cluster Poisson process with branching structure. In terms of the EM algorithm, the branching structure describes the complete data, whereas the CIF describes the incomplete data. We avoid unnecessary technical details and emphasize the substantive interpretation of the model. For more detailed treatments, the reader is referred to the classic book by Cox and Isham (1980) or the comprehensive work of Daley and Vere-Jones (2003).

2.1. *The Conditional Intensity Function*

A univariate temporal point process M is a random measure that assigns to each Borel set in $\mathbb{R}+$ a finite, positive integer. Intuitively, $M\{(a,b)\}$ counts the number of randomly occurring isolated points in the interval (a,b), and each point represents the time at which an event of interest occurs. Point processes extend immediately to the multivariate case. $M\{(a,b)\}$ is then vector-valued and each univariate margins gives the number of a different type of event occurring in the time period (a,b). In particular, a bivariate point process can represent each of two individuals' behaviours, one type of behaviour being recorded for each person. In this context, we interpret the margins of M as corresponding to the two individuals in a dyad and a realization of M as an interaction.

Under mild conditions, M is uniquely defined by its CIF (see Daley & Vere-Jones, 2003):

$$\lambda(t) = \lim_{\Delta \downarrow 0} \frac{E(M\{[t, t+\Delta)\} \mid H_t)}{\Delta} \tag{1}$$

where $E(M\{(a,b)\})$ is the expected number of points falling in (a,b) and $t \in \mathbb{R}+$. H_t is the history of the process up to time t. Letting t_k, $k \in \mathbb{N}$, denote the times at which an event has occurred before t, then $H_t = t_1, t_2, \ldots, t_k$. In substantive terms, the history of an interaction is defined by the times at which all previous behaviours have happened.

$\lambda(t)$ is an instantaneous conditional rate function. Its interpretation is aided by assuming that the probability of more than one event happening at a time is negligible:

$$\text{Prob}\big(M\big\{[t, t+\Delta)\big\} > 1\big) = o(\Delta) \tag{2}$$

for $\Delta \to 0$. When condition (2) holds, M is said to be *orderly* and we have the following intuitive relationship:

$$\text{Prob}\big(M\big\{[t, t+\Delta)\big\} = 1 \mid H_t\big) = \lambda(t)\Delta + o(\Delta). \tag{3}$$

As mentioned in the Introduction, Equation (3) is the instantaneous probability of an event and is denoted $p(t)$. The basic goal is to describe how $p(t)$ depends on the history of the event sequence.

We draw on linear systems analysis by writing $\lambda(t)$ as a bivariate causal filter (e.g., Brillinger, 1975). This may be represented in various equivalent ways, and here we prefer a

summation over the events in H_t. Letting $i, j \in \{1, 2\}$ denote the margins of M and letting t_{ik} denote the time of the kth event of the ith margin, $k \in \{1, \ldots, N_i\}$, the marginal CIFs are:

$$\lambda_i(t) = \mu_i + \sum_{t_{ik}<t} \phi_{ii}(t - t_{ik}) + \sum_{t_{jl}<t} \phi_{ij}(t - t_{jl}), \quad i \neq j. \tag{4}$$

We unpack the general interpretation of $\lambda_i(t)$ in the following 3 points:

1. The parameter μ_i is the base rate at which an individual can be expected to perform a behaviour, in the absence of the influence of any previous behaviours. It corresponds to paths 1A and 1B in Figure 1. In a particular application, μ_i might be conceptualized in terms of a person's habitual propensity toward a behaviour, or the influence of the context in which the interaction is situated. By way of comparison, the base rate is referred to as the "immigrant intensity" in ecological applications; it describes the rate with which new organisms are expected to arrive from other territories. Note that μ_i can be a function of time, although we do not assume that here.

2. The function ϕ_{ij} describes a person's responsiveness. The first subscript denotes the person who is responding, and the second subscript denotes the person to whom s/he is responding. In more generic terms, the first subscript is the output process and the second subscript the input process. When $i = j$ we refer to ϕ_{ij} as a self-response function and this corresponds to paths 2A and 2B in Figure 1; when $i \neq j$ we refer it as an other-response function and it describes paths 3A and 3B.

The role of $\phi_{ij}(t - t_{jk})$ is to model how the instantaneous probability of the ith margin deviates from μ_i following the occurrence of an input event t_{jk}. Note that for the input events, the first subscript denotes the members of the dyad, but the second subscript indicates the history of behaviours performed by that person (for visual clarity we prefer this to the more accurate notation t_{j_k}). We assume that $\phi_{ij}(u) = 0$ for $u < 0$ and $\int_0^\infty \phi_{ij}(u)\, du < \infty$, which defines Equation (4) as a bivariate causal filter. This means that a person cannot respond to an event that has not yet happened, and that the "total amount" of responsiveness to each event is finite. In the present context, these can be viewed as regularity conditions. A more specific definition of the response functions is given below.

3. The summation in Equation (4) is perhaps the most crucial component of the linear systems approach—it is the mechanism by which past events influence future events. In substantive terms, all of the previous behaviours to which I am still responsive, both my own behaviours and those of my interaction partner, determine the overall probability of my non-spontaneous action at time t. A theory of responsiveness as cumulative can be contrasted with the use of the Markov property in sequential analysis (e.g., Bakeman & Gottman, 1997).

Given the linear systems framework, the main issue is to obtain the response functions. There are two basic approaches: (a) we can estimate the ϕ_{ij} in the frequency domain, or (b) we can write down explicit expressions for the ϕ_{ij} and estimate the parameters in the time domain. Both approaches have their merits, but the focus of this paper is the latter. In particular, we desire a model in which the parameters describe relevant aspects of the individuals' responsiveness. It would also be convenient if ϕ_{ij} had the same functional form for each value of i and j, thereby facilitating comparison over types of responsiveness.

With these goals in mind, we propose

$$\phi_{ij}(u) = \alpha_{ij} \times f(u; \xi_{ij}). \tag{5}$$

Here, f is a probability density function defined on $\mathbb{R}+$ with parameter ξ_{ij} and $\alpha_{ij} \in \mathbb{R}$ governs the degree to which $\phi_{ij}(u)$ departs from μ_i. The exact choice of f is relatively arbitrary—it can be made to suit specific applications without affecting the main results of this paper. A common choice in many applications is the exponential density, although in the dyadic context this places relatively strict assumptions on the immediacy of a person's responsiveness. Instead we use the

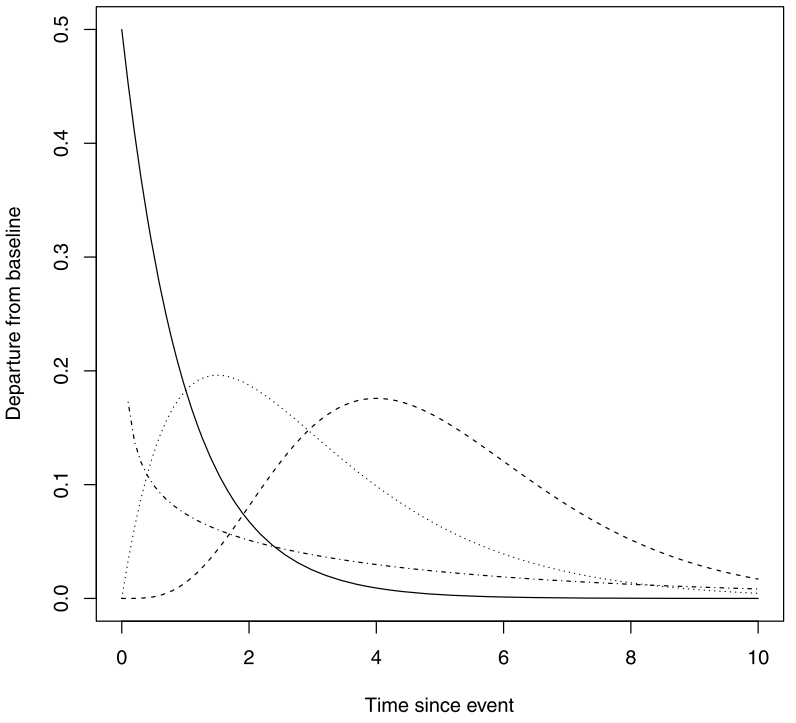


FIGURE 2.
Examples of the response curves from Equation (5) where f is the two-parameter gamma density with shape parameter κ and scale parameter β. The *solid black line* shows exponential decay ($\beta = \kappa = 1$) with an intensity parameter of $\alpha = 0.5$. Other curves have parameters selected from $\kappa = [0.5, 5]$, $\beta = [1, 5]$, $\alpha = [0.25, 0.75]$.

two-parameter gamma density, writing $\xi_{ij} = (\kappa_{ij}, \beta_{ij})$ to denote the shape and scale parameters, respectively.

Note that the parameters of the response functions do not depend on the history of the process, meaning that ϕ_{ij} describes the average response of person i over all input events t_{jk}, $k \in \{1, \ldots, N_j\}$. In general, it is possible to define response functions that depend on characteristics of the input events. This is discussed further in the conclusion.

We summarize the interpretation of the proposed response functions as follows. Some examples are depicted in Figure 2.

1. The weight α_{ij} represents the overall intensity of responsiveness of person i to a behaviour of person j. The precise meaning of "overall intensity" is given by the relation $\alpha_{ij} = \int_0^\infty \phi_{ij}(t)\, dt$. In other words, α_{ij} is the total responsiveness of person i to each event t_{jk} over an infinitely long interaction.

2. The shape parameter of the gamma density, κ_{ij}, governs the overall nature of the responsiveness of person i to person j. This parameter determines the excess skew and kurtosis of the gamma distribution, and in conjunction with the scale parameter it determines the location and variance.

3. The scale parameter of the gamma density, β_{ij}, describes the period of time over which person i is responsive to a behaviour of person j. This responsiveness window is interpreted in terms of the rescaling property of the gamma density: If $X \sim \Gamma(\kappa, \beta)$, then $cX \sim \Gamma(\kappa, c\beta)$.

2.2. Hawkes Process

Equations (4) and (5) result in a log-likelihood that contains the logarithm of a weighted sum of densities (see Equation (A1) in the Appendix). As is familiar from the literature on mixture modelling (e.g., McLachlan & Peel, 2000) and non-linear regression (e.g., Seber & Wild, 2003), this leads to problems with numerical optimization. In order to address this problem, we impose the following two restrictions:

$$\mu_i > 0 \quad \text{for } i = 1 \text{ or } i = 2, \tag{6}$$

$$1 > \alpha_{ij} \geq 0 \quad \text{for } i, j \in \{1, 2\}. \tag{7}$$

Restrictions (6) and (7) define Equations (4) and (5) as a bivariate Hawkes process (Hawkes, 1971; Hawkes & Oakes, 1974). We first consider the relevance of these restrictions in the context of dyadic interaction and then explain how they lead to a more tractable estimation problem via the branching structure representation of the process.

2.2.1. Interpretation of Equations (6) and (7) Equation (6) is quite mild. It basically requires that at least one behaviour during an interaction is treated as a spontaneous action. But the first behaviour of an interaction must be spontaneous, because there are no previous behaviours to which it may be considered a response. So Equation (6) amounts to saying that one member of the dyad must initiate the interaction. A general problem here concerns edge effects, or how to adjust for situations where the beginning (or end) of an interaction is not observed; this takes us beyond the scope of the present paper.

Equation (7) is more restrictive. It means that we can only describe processes in which the occurrence of one event has a non-negative influence on the probability of further events occurring. Such processes are referred to as excitatory. In other words, a Hawkes process is not appropriate when an interaction is characterized in terms of ongoing or accumulating inhibition of the responses of one or either party (e.g., stone-walling; Gottman, 1994). This is a significant limitation. We accept it as a necessary evil because it is required to obtain stable maximum likelihood estimates (MLEs).

Despite this limitation, the Hawkes process remains applicable to a wide range of human interactions that may be regarded as excitatory in nature. In particular, Hawkes processes are suited to describe periods of high activity interspersed with periods of inactivity, which is referred to as clustering. It has been argued that such a phenomenon is quite general in human dynamics (e.g., Barabási, 2005; Kalman et al., 2006). Plausible examples from the dyadic context include turn taking in conversation (e.g., Sacks, Schegloff, & Jefferson, 1974), mimetic desire (Girard, 1976), joint attention (e.g., Moore & Dunham, 1995), and affective reciprocity (e.g., Gottman, 1994).

It is possible to infer whether or not data are clustered by testing for over-dispersion in the stationary Poisson process, which we illustrate in the example. But, if Equation (7) becomes problematic in a given application, there are various ways of getting around it. For example, it may be possible to code the absence of one event as the presence of another, thereby replacing an inhibitory response with an excitatory one. It is also possible to omit one or more response functions should they be incompatible with either the theory or data of a particular interaction. Moreover, we can readily include refractory periods following an event by redefining the base rate as a function of time (e.g., Schoenberg & Bolt, 2000). For example, define $\mu_i = c_i(1 - f(t - t^*_{ik}))$ where c_i is a non-negative constant, t^*_{ik} is the most recent event due to person i, and f is a density function. Thus, a temporary, non-accumulating inhibition can be incorporated into the model, which may be relevant if event duration or response latency becomes an issue.

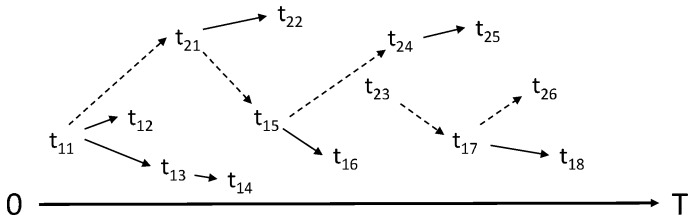


FIGURE 3.
An example branching structure for a bivariate process observed over the time interval $[0, T]$. Events are denoted by their times, t_{ik}. The vertical orientation of events is arbitrary. Each reaction ("offspring") is connected to its antecedent event by an arrow. *Solid arrows* indicate self-reactions, *dashed arrows* indicate other-reactions. Events without antecedents are actions ("immigrants").

2.2.2. The Branching Structure Representation of the Hawkes Process. In ecology, the branching structure describes the growth of a population in terms of subsequent generations of offspring due to each immigrant (e.g., Rasmussen, 2011). In the context of disease control, it is interpreted as the number of people contaminated by each subsequent carrier (e.g., Daley & Vere-Jones, 2003). In the present context, the branching structure parses an interaction into chains of responses following from each spontaneous action.

The branching structure can be explained with reference to Figure 3. The figure consists of a number of behaviours, t_{ik}, observed over the period, $[0, T]$. The arrows connect responses to their antecedent events. Self-responses are depicted by solid arrows, and other-responses are depicted by dashed arrows. Importantly, it is assumed that each response has exactly one antecedent. Events that do not have antecedents are spontaneous actions.

The following four statistical assumptions establish the equivalence between the branching structure and the DR model given in Equations (4) and (5) (see Hawkes & Oakes, 1974; Daley & Vere-Jones, 2003; Rasmussen, 2011).

1. The number of spontaneous actions of person i over the interval $[0, T]$ is a Poisson process with rate μ_i given in Equation (4).
2. The number of self-responses to each event t_{ik} over the interval $[t_{ik}, T]$ is a Poisson processes with rate $\phi_{ii}(t - t_{ik})$ given in Equation (5).
3. The number of other-reactions to each event t_{ik} over the interval $[t_{ik}, T]$ is a Poisson processes with rate $\phi_{ji}(t - t_{ik})$ given in Equation (5), $i \neq j$.
4. All of the Poisson processes in Assumptions 1–3 are independent.

When the branching structure is known, the estimation of a Hawkes process is routine. It is much more usual that the branching structure is not known, in which case it is intuitive to compare the situation with that found in finite mixture modelling. We may view each observation as belonging to one of three possible processes, namely those for spontaneous actions, self-responses, or other-responses. The distribution of each process is known to be of the Poisson family, and we desire the parameters of these distributions. Based on this analogy, it is intuitive to introduce a missing variable that identifies the process to which each event belongs and proceed by means of the EM algorithm.

The main dissimilarity with mixture modelling is the absence of explicit mixing proportions. As we show in the Appendix, these are readily obtained from the rate function of the Poisson process. There is also the additional complication that the Poisson processes governing self- and other-responses are defined with reference to specific antecedent events. Since the timing of each event is known, this is mainly a matter of indexing the possible antecedents.

The remainder of this section develops a notation for the treatment of the branching structure as missing data. The notation is quite dense but is required for the results presented in the following section. We begin by writing

$$\lambda_{ijk}(t) = \begin{cases} \phi_{ij}(t - t_{jk}) & \text{for } t \geq t_{jk} \\ 0 & \text{for } t < t_{jk} \end{cases} \tag{8}$$

for $i, j \in \{1, 2\}$, $k \in \{1, \ldots, N_j\}$. The subscripts on $\lambda_{ijk}(t)$ are interpreted as follows: the first subscript is the output process, the second is the input process, and the third is the specific stimulus event in the input process. We also let $\lambda_{i00} = \mu_i$. Note that the rate function of the Poisson process is in fact a CIF as defined above in Equation (1); because the Poisson process does not depend on its history, its CIF is equal to its unconditional intensity function (i.e., its rate).

Next, we write z_{ijk} to denote the process of responses of person i to behavioural event t_{jk}. The subscripts on z_{ijk} have the same interpretation as for λ_{ijk}. If t_{il} is a response to event t_{jk}, then $t_{il} \in z_{ijk}$. If t_{il} is not a response but a spontaneous action, then $t_{il} \in z_{i00}$. We let $z_i = \{z_{i00}, z_{ii1}, \ldots, z_{iiN_i}, z_{ij1}, \ldots z_{ijN_j}\}$, $i \neq j$, denote the collection of all processes defined over the ith margin. Because there is a one-to-one correspondence between the processes z_{ijk} and their rate functions $\lambda_{ijk}(t)$, we make use of the shorthand notation $\lambda_z(t)$ for $z \in z_i$.

Lastly, we introduce the variables $Z_i = (Z_{i1}, \ldots, Z_{iN_i})$ and write $Z_{ik} = z$ if $t_{ik} \in z$ for $z \in z_i$. Thus, Z_{ik} indicates the process to which t_{ik} belongs. In some applications it may be convenient to represent the values of Z_{ik} as integers, although that is not required here. Treating $Z = (Z_1, Z_2)$ as missing leads to an EM algorithm, which is the focus of the following section.

Although we have introduced Z for the purpose of estimation, its substantive interpretation is also quite interesting. By treating Z as random, each Z_{ik} is associated with a categorical distribution describing the probability that the behaviour t_{ik} is a response to every other behaviour. These probabilities can be used to assign behaviours to their most likely type and to assign to each response its most likely antecedent. More radically, the probabilities may be viewed as weights, each behaviour then being "comprised" of multiple independent sources (i.e., part spontaneous action, part self-response, part other-response). It is also possible to allow the response functions to depend on the type of antecedent, so that, for instance, I may become more responsive to you once you have responded to me. We can also impose structural constraints on the branching structure and compare these against an unconstrained model, thereby testing for specific sequences of interactions. In short, there are various interesting applications of the branching structure in the context of dyadic interaction.

2.3. Summary

We have specified the DR model as a bivariate Hawkes process. The model has two representations, one as a CIF that is a causal filter and one as a cluster Poisson process with branching structure. The former seems to us to have a more general interpretation, but the branching structure is useful for parameter estimation and provides some interesting possibilities for further analysis. Our general approach to dyadic interaction is similar in spirit to Gottman's equations of marriage (e.g., Gottman et al., 2003; Hamaker et al., 2009; Madhyastha, Hamaker, & Gottman, 2011). The parallel between his notion of influence and our notion of other-responsiveness is especially apparent. The essential difference is that we focus on event data.

3. Estimation

In this section, we formulate an EM algorithm for the maximum likelihood estimation of the DR model. In common with the case of finite mixture modelling, the complete data log-likelihood

replaces a weighted sum of densities with a single density, which greatly simplifies numerical optimization. In the current application, the EM algorithm has the further advantage of reducing the dimensionality of the numerical optimization problem through closed form and concentrated solutions. Additionally, basing estimation on the EM algorithm means that the distribution of the branching structure is available.

Let $l(\theta \mid X)$ denote the incomplete data log-likelihood, where $\theta = (\theta_1, \theta_2)$ contains the parameters of the model and $X = (X_1, X_2)$ with $X_i = (t_{i1}, \ldots, t_{iN_i})$. The EM approach involves conditioning $l(\theta \mid X)$ on the missing variable Z introduced above. This leads to an expression for the complete data log-likelihood, denoted $l(\theta \mid X, Z)$. The objective function of the M step, $Q = Q(\theta \mid \theta^{(n)})$, is the expectation of $l(\theta \mid X, Z)$ with respect to the posterior distribution of Z given the observed time points X and a "current value" of the parameter vector, denoted $\theta^{(n)}$. For further discussion of the EM algorithm, see McLachlan and Krishnan (2008).

In Appendix, we derive the following four results presented in Equations (9)–(12). First, we have the objective function

$$Q\big(\theta \mid \theta^{(n)}\big) = \sum_{i=1}^{2}\left(\sum_{z\in z_i}\left(\sum_{k=1}^{N_i}\log\big(\lambda_z(t_{ik})\big) \times \operatorname{Prob}\big(Z_{ik} = z \mid X_i, \theta_i^{(n)}\big) - \Lambda_z(T)\right)\right) \tag{9}$$

where $\Lambda_z(T) = \int_0^T \lambda_z(s)\,ds$ and $[0, T]$ is the time over which the interaction is observed. The posterior probabilities of Z are obtained from Bayes' rule:

$$\operatorname{Prob}\big(Z_{ik} = z \mid X_i, \theta_i^{(n)}\big) = \frac{\lambda_z^{(n)}(t_{ik})}{\sum_{m\in z_i}\lambda_m^{(n)}(t_{ik})}. \tag{10}$$

Equations (9) and (10) provide the necessary components of the EM algorithm described below; Equation (10) is readily computed on the E step, and Equation (9) can be submitted to a generic optimizer on the M step. It is desirable to simplify the M step because the parameter space of the model is quite large; a total of 14 parameters, 3 for each of 4 response functions, plus 2 base rates. This is accomplished by the following two results.

Firstly, there is a closed form solution for the base rate parameters:

$$\mu_i^{(n+1)} = \bar{N}_{i00}^{(n)}/T. \tag{11}$$

Here, $\bar{N}_{i00}$ is the expected number of actions of person i. Equation (11) is the EM analogue of the well known result for a homogeneous Poisson process.

Secondly, the updated intensity parameters $\alpha_{ij}^{(n+1)}$ can be written in terms of the $\beta_{ij}^{(n+1)}$ and $\kappa_{ij}^{(n+1)}$:

$$\alpha_{ij}^{(n+1)} = \bar{N}_{ij}^{(n)} \Big/ \sum_{k=1}^{N_j} F\big(T - t_{jk}; \beta_{ij}^{(n+1)}, \kappa_{ij}^{(n+1)}\big), \tag{12}$$

where F denotes the distribution function of the kernel density f, and $\bar{N}_{ij}$ is the expected number of responses of person i to person j. Equation (12) leads to a concentrated maximum likelihood solution by substituting its right-hand side into Q on the M step. Equation (12) also shows that $\bar{N}_{ij}/N_j$ is not an unreasonable upper bound on α_{ij}, leading to a rough interpretation of the intensity parameters in terms of proportions of events. We make use of this interpretation in the example.

Using Equations (11) and (12), the numerical optimization of Q is reduced from 14 to 8 dimensions. These results are also independent of the particular kernel used for the response functions. The shape parameter of the gamma kernel cannot be solved for using Equation (9).

TABLE 1.
Bias and error of parameter recovery.

	μ_i	α_{ii}	α_{ij}	β_{ii}	β_{ij}	κ_{ii}	κ_{ij}
True values	0.020	0.500	0.300	1.000	2.000	1.000	3.000
CIF	0.021	0.495	0.292	1.096	2.643	1.103	3.292
	(0.003)	(0.037)	(0.043)	(0.554)	(1.690)	(0.442)	(1.672)
EM	0.021	0.492	0.295	0.972	2.107	1.029	3.175
	(0.003)	(0.030)	(0.039)	(0.205)	(0.832)	(0.121)	(0.920)

Note: CIF denotes estimation based on the incomplete data log-likelihood. EM denotes estimation based on Algorithm 1. For each method the mean of the parameter estimates over the simulated data sets is reported, and the standard deviation appears in parentheses.

The rate parameter has a concentrated maximum likelihood solution, but this is subject to edge effects and we do not pursue it here.

Equations (9) through (12) lead to the following EM algorithm. Let θ_+ contain the parameters of the gamma kernel.

Algorithm 1. An EM algorithm for the DR model.

Step 0 Set $n = 1$; select strictly positive starting values for the parameters $\theta^{(n)}$.

Step 1 (E step) Use $\theta^{(n)}$ to compute $\text{Prob}\{Z_{ik} = z \mid X_i, \theta_i^{(n)}\}$ in Equation (10) for $i \in \{1, 2\}$, $k \in \{1, \ldots, N_i\}$, and $z \in z_i$.

Step 2 (M step) For $i \in \{1, 2\}$:

1. Compute $\mu_i^{(n+1)}$ using Equation (11).
2. Using $\theta_+^{(n)}$ as starting values, find $\theta_+^{(n+1)} = \arg\max_{\theta_+} Q(\theta_+ \mid \theta^{(n)})$ subject to the constraints in Equation (12).
3. Use $\theta_+^{(n+1)}$ to compute $\alpha_{ij}^{(n+1)}$ from Equation (12), $j \in \{1, 2\}$.
4. Store the resulting parameter values in $\theta^{(n+1)}$.

Step 3 If $l(\theta^{(n+1)} \mid X) - l(\theta^{(n)} \mid X) < \epsilon$ for some convergence criterion ϵ, return $\theta^{(n+1)}$; otherwise set $n = n + 1$ and repeat Steps 1–3.

Algorithm 1 is similar to the algorithm presented by Veen and Schoenberg (2008). The main differences are the details of the specific models considered, and that we have tried to be more explicit in motivating the EM approach. Their simulation study indicated that the estimation error of the EM algorithm is superior to that of the incomplete-data likelihood. Table 1 presents a small simulation study that leads to the same conclusion for the present application. We generated 100 data sets of approximately $N_i = 400$ occurrences from a fixed parameterization of the DR model. For each dataset, the model was estimated by maximizing the incomplete data log-likelihood (see the Appendix) and by using Algorithm 1. Due to running time considerations, we only estimated a single margin for each dataset. Starting values were obtained by randomly perturbing the true parameter values. R (R Development Core Team, 2011) code for both the simulation and estimation of the DR model is available from the first author's website.

The table shows that the performance of the two approaches is similar for the base rate and intensity parameters. For the duration and shape parameters, the error of Algorithm 1 is much smaller. This is also illustrated in Figure 4, which shows the distributions of the κ_{ij} estimates. A larger simulation study would be useful to provide a more complete evaluation of Algorithm 1 and how it compares to other methods (e.g., Rasmussen, 2011).

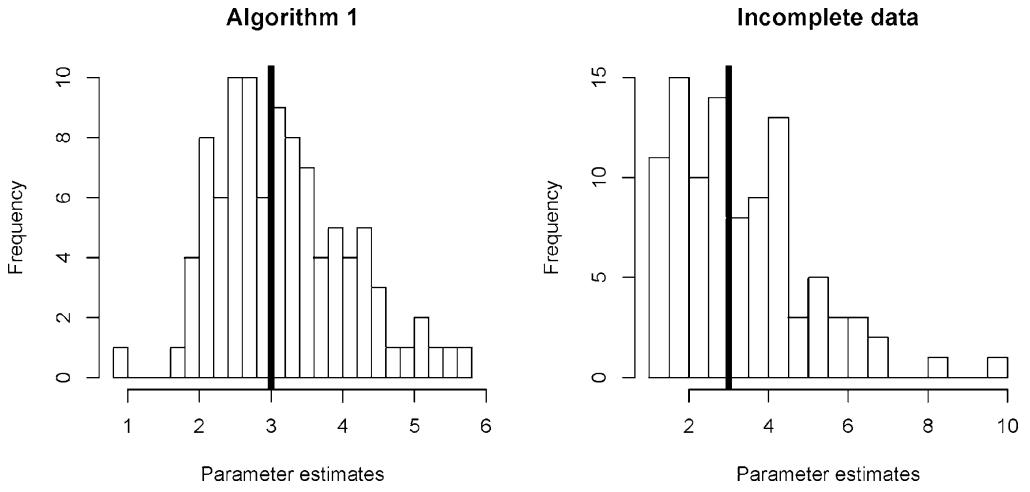


FIGURE 4.
Distribution of parameter estimates of κ_{ij} for Algorithm 1 and the incomplete data log-likelihood (see Equation (A1)). 100 data sets were generated from a DR model with true value $\kappa_{ij} = 3$ indicated in *bold*.

Before moving on, we make a note about local maxima. As with the situation in finite mixture modelling, the presence of local maxima in the Q function requires the use of multiple starting values when applying the EM algorithm without knowledge of the data generating parameters (McLachlan & Peel, 2000; also see Ogata, 1978). The simulation study purposely avoided this problem, but we have encountered a number of real and simulated data sets that resulted in local maxima. In practice, this can be dealt with by using multiple starting values, which is the approach we take in the example. It would also be desirable to have results for the curvature of Q.

4. Example: Email Interactions in the Workplace

We used the DR model to analyse email transactions between an employee and employer. The beginning of the record coincides with the beginning of the employment period and lasts for 515 days ($T = 12360$ hours). The record consists of the sending times of $N_1 = 226$ emails sent from the employee to the employer, and $N_2 = 195$ sent from the employer to the employee. The record is complete with respect to the two individuals' work-related email accounts over the observation period. The goals of this example are to (a) describe the email behaviour of the two individuals in terms of the model parameters, (b) describe the dependence between the two individuals' email behaviour, and (c) illustrate general results concerning goodness of fit and asymptotic inference. We also mention the relevance of the findings for existing literature on human dynamics.

4.1. Descriptive Statistics and Goodness of Fit

Table 2 reports descriptive statistics. The waiting times are the intervals between subsequent emails sent from the same email account (i.e., within-individual). The response times are the intervals between an email sent from one individual's account and the next reply sent from the respondent (i.e., between-individual). The respondents are indicated in the columns of the table.

For the employee, the waiting times ranged from less than a minute to almost 20 days. Nearly half of the emails sent by the employee were within 24 hours of the previous email, and about

TABLE 2.
Descriptive statistics for raw waiting times (hours).

	Employee	Employer
Waiting times		
Range (Median)	[0.02, 473.90] (27.50)	[0.02, 522.95] (28.93)
Frequency (%) < 24 hours	111 (49.12)	88 (45.13)
Frequency (%) < 1 hours	16 (7.07)	17 (8.71)
Response times		
Range (Median)	[0.03, 517.72] (19.97)	[0.02, 331.28] (7.40)
Frequency (%) < 24 hours	127 (56.19)	136 (69.75)
Frequency (%) < 1 hours	26 (11.50)	39 (20.00)

Note: Waiting times are the times between subsequent emails from the same sender. Response times are the times between each email from the sender and the subsequent reply; respondents are indicated by the column titles. A total of $N_1 = 226$ emails were sent by the employee and $N_2 = 195$ by the employer.

7 % were sent within one hour. A similar situation held for the waiting times of the employer. Overall, the response times were shorter than the waiting times. The median response time for the employer was only 7.4 hours, and 20 % of the replies came within one hour. For the employee, the apparent differences between the response times and waiting times were less pronounced.

Based on Table 2, we conclude that the waiting times and response times were highly right-skewed for both the employee and the employer. There has been some discussion as to whether and in what capacity email behaviour follows a power law distribution (e.g., Barabási, 2005; Ebel, Mielsch, & Bornholdt, 2002; Eckmann et al., 2004; Kalman et al., 2006; Newman, 2005). On this point, we note that log-log plots (not shown) of the data described in Table 2 did not demonstrate good agreement with a power law distribution, but were comparable to Figure 5(a) of Newman (2005). This could be due to the relatively small number of emails in the present record.

Although the waiting times and response times did not follow a power law distribution, they nonetheless demonstrated clustering. Here, we refer to the top two rows of Figure 5, which show the goodness of fit of the homogeneous Poisson process to our data. In general, the residuals of a point process can obtained by means of the time change theorem. The theorem states that, under a correctly specified model, the observed time points can be rescaled via the integral of $\lambda(t)$ so that their waiting times are exponentially distributed with rate equal to one (see Daley & Vere-Jones, 2003). Thus, for a point process that is specified through its CIF, a large number methods are available for testing goodness of fit. As shown in the figure, the email behaviour of our dyad was not well explained by the homogeneous Poisson process, showing both too many short waiting times and several waiting times that were longer that expected (i.e., clustering). We conclude that application of the Hawkes process is warranted.

To estimate the DR model, we implemented Algorithm 1 in `R` (R Development Core Team, 2011). For numerical optimization, we used the Nelder–Mead method of the `optim` package. To address the possibility of local maxima, we divided estimation into two stages. In the first stage, we ran a total of 100 starting values for a maximum of 20 iterations with convergence set at $\epsilon = 0.01$. Starting values for the base rate parameters were randomly selected in the range $[0.001, N_i/T]$, and values for the intensity parameters were randomly selected from $[0.1, 0.9]$. These intervals reflect the range of possible values. For the parameters of the gamma density, however, there are no natural upper bounds. To allow for a large range of configurations, we randomly selected scale parameters from $[1, 5]$ and shape parameters from $[1, 3]$.

In the second stage, we ran the best 20 solutions from the first stage until convergence at $\epsilon = 1 \times 10^{-5}$. All 20 solutions converged to $l = -1948.42$ within 50 second-stage iterations.

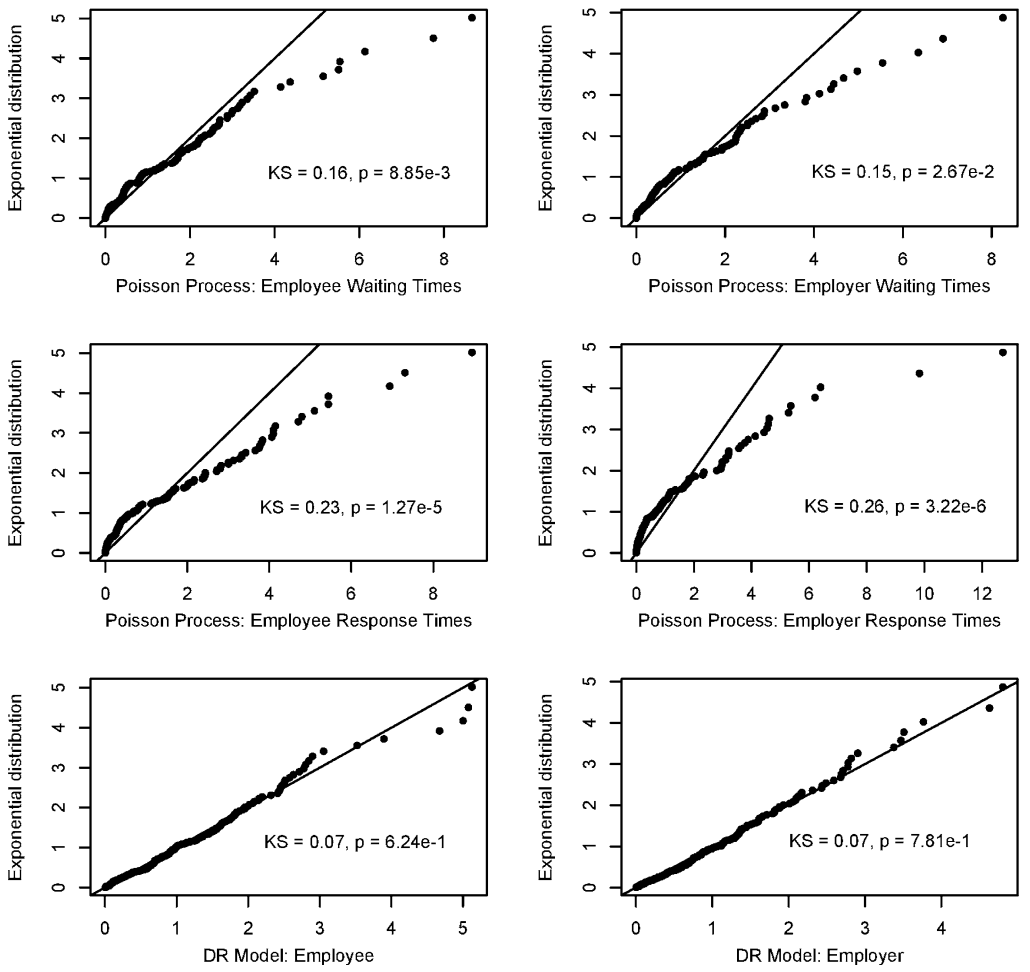


FIGURE 5.
Quantile–quantile plots for email data. *Insets* give values of Kolmogorov–Smirnov statistic (KS) and its two-tailed probability under the hypothesis that the distributions are equivalent. Data for the employee are in the *left column*, and the employer in the *right*. The *top row* shows the residuals of the homogeneous Poisson process fitted to the within-individual waiting times. The *middle row* shows the between-individual response times. The *lower row* shows the residuals for the DR model.

The condition number of the Q function at the solution was reasonable (3.47×10^3), indicating local identification of the model. By comparison, the condition number of the incomplete data log-likelihood at the solution was 4.22×10^7. This is in keeping with the results of Veen and Schoenberg (2008) and again indicates the usefulness of Algorithm 1.

As shown in the bottom row of Figure 5, the DR model adequately accounted for the clustering in the data. We now turn to consider how the model parameters can be used to describe the email transactions.

4.2. *Interpreting the Model Parameters*

In the present context, we interpret the three-way classification of events (see Figure 1) as follows: spontaneous actions correspond to starting of new threads, self-responses are follow-up emails, and other-responses are reply emails. Table 3 reports the MLEs of the model. The table also contains the estimated asymptotic standard errors, which were computed via the Hessian

TABLE 3.
Parameter estimates and standard errors.

	Employee	Employer
Baseline (*new threads*)	0.011 (0.001)	0.006 (0.001)
$\bar{N}_{i00}$	131.875 (11.485)	74.342 (8.623)
Self-Responses (*follow-ups*)		
Intensity	0.057 (0.016)	0.070 (0.019)
Duration	2.520 (1.275)	0.856 (0.390)
Shape	0.522 (0.170)	0.739 (0.243)
$\bar{N}_{ii}$	12.913 (1.441)	13.556 (1.540)
Other-Responses (*replies*)		
Intensity	0.417 (0.046)	0.475 (0.046)
Duration	17.066 (3.388)	9.672 (1.642)
Shape	0.574 (0.076)	0.602 (0.069)
$\bar{N}_{ij}$	81.212 (10.483)	107.102 (10.488)

Note: Entries are maximum likelihood estimates with asymptotic standard errors in parentheses. $\bar{N}$ denotes expected number of events.

of Q evaluated at the MLEs (see Jamshidian & Jennrich, 2000). Additionally, the table reports the MLEs for the expected number of new threads, $\bar{N}_{i00}$, the expected number of follow-up emails, $\bar{N}_{ii}$, and the expected number of replies, $\bar{N}_{ij}$. These estimates were obtained by solving Equations (11) and (12) for the expected counts, and their standard errors were obtained using the delta method.

Based on Table 3, we make the three following comparisons of the email behaviour of the employee and employer. We do not explicitly discuss the results in terms of statistical significance, but the reader may compute the usual Wald tests for differences from the table. Note that testing whether estimates differ from zero would require selection methods for non-nested models (e.g., information criteria) since all of the parameters are bounded by zero. In practice, this involves substituting the MLEs from Algorithm 1 into Equation (A1) in the Appendix and computing the relevant penalty.

1. The employee had a higher base rate ($\hat{\mu}_1 = 0.011$) than the employer ($\hat{\mu}_2 = 0.006$), meaning that the employee started a larger number of new threads. But the individuals did not differ in the intensity of their follow-ups or replies. As per Equation (12), we can interpret the intensities roughly in terms of proportions of emails. In particular, of the $N_1 = 226$ total emails sent by the employee, the employer responded to $\hat{N}_{21} \approx 107$ or about 47 %; of the $N_2 = 195$ total email sent by the employer; the employee responded to $\hat{N}_{12} \approx 81$ or about 41 %.

The DR model leads us to infer that less than 50 % of the emails sent between the two individuals were in fact replies. This may be contrasted with the common methodological assumption that a subsequent email from one party is necessarily a reply to the previous email of the other party (e.g., Barabási, 2005; Eckmann et al., 2004). Emails that do not receive replies may be of a declarative nature, their content may be followed up by other modes of communications such as meetings and phone calls, or they may simply be ignored. In general, to the extent that the behaviours comprising an interaction are not a strict sequence of other-responses, the analysis of raw response times lacks a clear rationale.

2. Although the individuals did not differ in the overall proportion of reply and follow-up emails, they did differ in response times. In particular, the duration of the employee's other-response function ($\hat{\beta}_{12} = 17.066$) was substantially longer than that of the employer ($\hat{\beta}_{21} = 9.672$). This translates to median response time of 4.984 hours for the employee, versus 3.068 hours for the employer. Comparing these values to Table 2, we may conclude that the sending

time of the emails identified as replies by the DR model were overall much shorter than the response times from the table. In intuitive terms, this means that a person is likely to respond quickly or not all.

3. Lastly, we consider the shape of the response curves. At first, we were quite surprised to find that all response curves had $\hat{\kappa}_{ij} < 1$, meaning that a large proportion of follow-ups and replies were sent immediately, not after a period of some delay. Moreover, we were surprised to find that the estimates were approximately equal for all response functions. This led us to fit the DR model again, this time using the estimates in Table 3 as starting values, but constraining the shape parameters to be equal with starting value set to their mean. The model converged at $\epsilon = 1 \times 10^{-5}$ in six iterations, and the usual likelihood ratio test for nested models showed the constraints to be viable ($\chi^2 = 0.516$; $df = 3$; $p = 0.915$). We conclude that the follow-ups and the replies showed equivalent sub-exponential decay for both individuals, with an estimated shape parameter of $\hat{\kappa} = 0.593$ in the constrained model.

This finding requires some explanation. In the case of the follow-up emails, the immediate responsiveness combined with the relatively short duration of responsiveness admits the following interpretation: Senders follow up their own emails when an error is noticed (e.g., omitted attachments). After an initial period, these types of mistakes are likely to have been missed. Hence, most follow-up emails have a short duration window with immediate onset.

In the case of the reply emails, the immediate responsiveness may appear rather implausible. In particular, it seems that reading an incoming email and composing a reply should require a non-negligible period of time. Here, we recall that a relatively large proportion of the raw response times were less than one hour (see Table 2). Using the MLEs $\hat{\theta}$ and the posterior probabilities $\text{Prob}(Z_{ik} = z \mid X, \hat{\theta})$ from Equation (10), we computed the probability that the emails with response times of less than one hour were replies to the previous email sent by the other individual. For the employee, the average probability over 26 emails was 0.743 and for the employer the average probability was 0.865 over 39 emails. We conclude that emails with relatively short response times (a) were to be found in the data and (b) were likely to be replies.

In summary, Table 3 shows that the employee initiated more email transactions than the employer, but was also slower to reply. The specific interpretation in terms of the relationship between the two individuals, or in terms of email interaction in the work place, depends on substantive theory. For instance, the relative promptness the employer's replies may be indicative of an egalitarian relationship (Anderson, 2008). In other respects, the individuals showed similar email behaviour, and it was notable that both the follow-up emails and replies were most likely to be sent immediately rather than after a period of delay. The fact that both types of emails showed the same rate of decay for both individuals, but that these distributions were not well approximated by a power law, leads to further questions about the possibility of a general pattern underlying email behaviour.

4.3. Dependence Between the Individuals

We now consider the historical dependence exhibited by the dyad's email behaviour. The first 2000 hours ($\approx$ 83 days) of the bivariate time series are depicted in Figure 6. In the figure, A denotes spontaneous action (initiation of new threads), O denotes other-responses (reply emails), and S denotes self-responses (follow-up emails). The central part of the figure shows the univariate CIFs and the upper and lower margins show the event times. We first describe the event margins and then the CIFs.

Each event was classified into one of the three types based on $P_{ik} = \max_z \text{Prob}(Z_{ik} = z \mid X, \hat{\theta})$. In the present example, the classification was not very accurate, with averages $\bar{P}_{1\bullet} = 0.75$ and $\bar{P}_{2\bullet} = 0.80$. Nonetheless, the classified event streams make apparent the possibilities of further analysis using the branching structure.

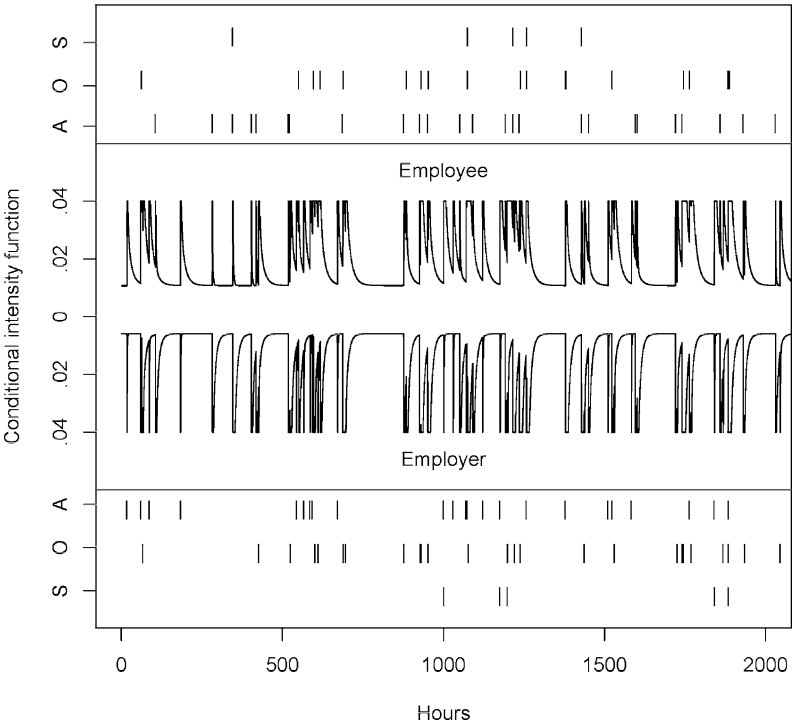


FIGURE 6.
Observed time series and fitted conditional intensity functions (CIFs) for employee (*above*) and employer (*below*). The *vertical axis* is reflected about zero to emphasize the gross level symmetry of the CIFs (i.e., all values are non-negative). The time series are depicted as hash marks in the *upper* and *lower margins* and were classified as A = spontaneous action, O = other-response or S = self-response using the branching structure (see text for details).

Turning to the CIFs, it is remarkable that they are quite symmetrical in their overall appearance, with clusters of spikes and gaps between clusters appearing to be highly synchronized between the two individuals. While clustering has been observed in population level human dynamics (e.g., Barabási, 2005; Crane & Sornette, 2008; Kalman et al., 2006), the synchronicity apparent in Figure 6 is intrinsically dyadic in nature. Therefore, its explanation motivates a micro-level approach.

The CIFs in Figure 6 provide a relatively clear depiction of the overall coordination between the email behaviour of the employer and employee, but it is also desirable to have an analytic quantity that describes this dependency. There are various quantities that are appropriate to the linear case (e.g., Brillinger, 1975). Here, we consider mutual information, which for a general bivariate point process, is defined as

$$I = E_{M_1} E_{M_2} \int_0^T \ln\left(\frac{\lambda_1(t)\lambda_2(t)}{\gamma_1(t)\gamma_2(t)}\right) dt \tag{13}$$

where $\gamma_i(t)$ is the CIF of the i-th margin obtained by allowing it to depend only on its own history, not on the history of the jth margin. For the DR model,

$$\gamma_i(t) = \lambda_i(t) - \sum_{t_{ik}<t} \phi_{ii}(t - t_{ik}), \quad i \neq j. \tag{14}$$

In the context of dyadic interaction, I quantifies how much the behaviour of one individual depends on the previous behaviours of the other individual. In the case that a dyads' behaviours are statistically independent of one another $I = 0$, and otherwise $I > 0$. It is usual to treat I as a

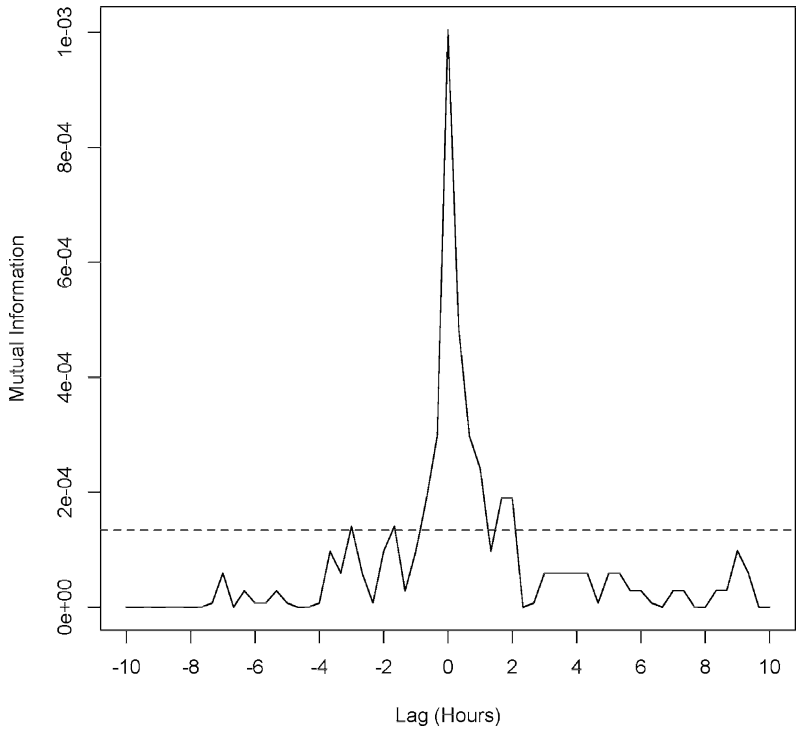


FIGURE 7.
Estimated mutual information between the email behaviour of the employer and employee as a function of time lag applied to the employer. The *dashed line* represents the upper 99 % confidence limit under the hypothesis that the two processes are independent. Mutual information was estimated using bins of 20 minutes duration. Note that the zero point corresponds to window $[-10, 10]$ minutes; however, the mutual information is not defined for the DR model at lag $= 0$.

function of continuous time lag u, but note that the assumption of orderliness (see Equation (2)) implies that I is not defined when $u = 0$. If $I(u) > 0$ for $u \neq 0$, then the two processes are said to be dependent upon one another at lag u.

One way to obtain an estimate of I is to replace the expectation in Equation (13) with its observed value. This leads to a model-implied quantity, but one whose sampling error is not readily obtained. Therefore, we followed the method of Brillinger (2004) by estimating I directly from the email data. This required binning the data, and we used T/δ uniform bins of width $\delta = 20$ minutes. The resulting estimate is shown in Figure 7, with a lag of $u = [-10, 10]$ hours applied to the employer. The dashed line represents the upper 99 % confidence limit under the hypothesis that the processes are independent. There is a clear maximum at $u = 0$, which corresponds to the interval $[-10, 10]$ minutes. This maximum sharply drops to zero within $u = [-2, 2]$ hours.

Figure 7 provides a more precise way of describing the synchronization depicted in Figure 6: for both individuals, maximal dependency on the behaviour of the other occurred immediately and decayed quickly. Given that many modes of communication ostensibly have less intrinsic time delay than email (e.g., face-to-face, instant messaging), we conjecture that the synchronization observed in the present example may occur much more generally. Because this phenomena is intrinsically dyadic in nature, it demands an explanation in terms of micro-level human interaction.

5. Conclusion

This paper has addressed application of the Hawkes process to event data collected within dyads. We outlined its interpretation in the dyadic context, presented an EM algorithm for parameter estimation, and illustrated its application and related results with an example concerning email transactions in the work place.

Our general view is that event sampling provides a rich way of gaining information about dyadic interaction. There is a large variety of ready-made occurrences available to the behavioural researcher, notably in the domain of communication and information technology. On the other hand, there is a relative lack of statistical methodology available for the analysis of this type of data. The notable exception here is sequential analysis, and this is limited to sequences of events. The approach we have presented is an initial step towards modelling a basic and neglected phenomenon, the timing of human interactions.

It is important to emphasize the limitations of the DR model. First, the model is principally descriptive. For it to be properly explanatory, it is necessary to incorporate time-varying covariates into the model. In psychological applications, it is especially relevant to allow for continuous covariates with measurement error; and, as mentioned, the model itself also provides useful covariates in the form of its branching structure. These issues fall under the rubric of marked point processes (see Daley & Vere-Jones, 2003), and we are currently addressing computational challenges in this area.

Second, there is the challenge of comparison across dyads. The model can be fitted for each dyad and the parameter estimates related to outcomes of interest, but this does not allow for the simultaneous analysis of multiple dyads. For this purpose, it is desirable to introduce random effects into the model, leading to a multilevel approach (e.g., Scott, 2011). For aggregate behaviour of larger numbers of individuals (e.g., networks), more parsimonious parameterizations of similar models have proven useful (e.g., Eckmann et al., 2004; Crane & Sornette, 2008).

Third, the DR model is intrinsically limited to interactions that are excitatory in nature. While the Hawkes process has been regarded as the closest equivalent of an autoregressive moving-average (ARMA) framework for event data (Daley & Vere-Jones, 2003) and has found many applications in other fields, there remains the question of its breadth of applicability in the behavioural sciences. This question can best be answered by applications of the model. There are also questions related to the small sample properties of the MLEs and whether it is preferential to employ Bayesian estimators (e.g., Rasmussen, 2011) in such circumstances. This paper has only considered asymptotic inference. Related areas of extension include development of models for non-excitatory interactions (e.g., Isham & Westcott, 1979), derivation of model-implied quantities for small-sample inference about the nature of an interaction, and problems in forecasting and on-line updating of estimates. In all of these areas, the demands of behavioural research can drive important contributions to the point processing literature.

Acknowledgements

This research was funded in part by a postdoctoral grant from the National Science and Engineering Research Council of Canada.

Appendix. Derivation of Estimating Equations

In this Appendix, we derive the equations required for Algorithm 1. We first present the incomplete and complete data log-likelihoods for the DR model, which leads to Equation (9).

We then obtain Equation (10) using Bayes' rule, and Equations (11) and (12) are obtained from the expected complete data log-likelihood.

The general form of the log-likelihood for a univariate point process with CIF $\lambda(t)$ is (see Daley & Vere-Jones, 2003)

$$l(\theta \mid X) = \sum_k \ln\bigl(\lambda(t_k)\bigr) - \int_0^T \lambda(s)\, ds, \tag{A1}$$

where $[0, T]$ is the observation period, $X = (t_1, \ldots, t_N)$ denotes the observed event times, and θ contains the parameters of the model. The integral of $\lambda(t)$ over the interval $[0, u]$ is sometimes referred to as the compensator and is denoted as $\Lambda(u)$.

For an orderly, multivariate point process the log-likelihood can be treated through the univariate margins. In particular, if the log-likelihood for each univariate margin is l_i, the overall likelihood is given as $l = \sum_i l_i$. We present results in terms of the univariate margins to simplify notation.

Substituting Equations (4) through (7) into the bivariate version of (A1) leads directly to an expression for the incomplete data log-likelihood of the DR model. As discussed above, this does not produce stable parameter estimates because it contains the logarithm of a weighted sum of densities. Therefore, we employ the branching structure representation of the Hawkes process to obtain the complete data log-likelihood:

$$l_i(\theta_i \mid X_i, Z_i) = \sum_{z \in z_i} \Biggl(\sum_{t_{ik} \in z} \ln\bigl(\lambda_z(t_{ik})\bigr) - \Lambda_z(T) \Biggr) \tag{A2}$$

for $i = \{1, 2, \}$ and $k \in \{1, \ldots, N_i\}$. The collection of processes z_i and the CIFs $\lambda_z(t)$ were defined in our discussion of the branching structure. The summation over $z \in z_i$ is implied by the independence assumptions of the branching structure. The equivalence between Equation (A2) and the incomplete data representation was initially established by Hawkes and Oakes (1974). The overall result is to replace the logarithm of a sum of densities with the sum of their logarithms, which is also the role of the complete data log-likelihood in finite mixture modelling (McLachlan & Peel, 2000, pp. 47–49).

The unknown part of Equation (A2) is represented by the summation over $t_{ik} \in z$; because we do not know the branching structure, we do not know which process each t_{ik} belongs to. It is important to recall here that the notation $\lambda_z(t)$ is shorthand for $\lambda_{ijk}(t)$ and does not indicate dependence of the rate functions on the missing variable Z_i. For explicitness, we also note that the number of processes in z_i is $N_1 + N_2 + 1$, which is necessarily larger than the number of events N_i. This implies that some of the processes in z_i contain no events. For these cases, we evaluate the Poisson process with zero arrivals, so only the compensator contributes to the log-likelihood.

In order to obtain the expectation of Equation (A2) over the posterior distribution of Z, we write it in the equivalent form:

$$l_i(\theta_i \mid X_i, Z_i) = \sum_{z \in z_i} \Biggl(\sum_{k=1}^{N_i} \ln\bigl(\lambda_z(t_{ik})\bigr) \times \delta(Z_{ik} = z) - \Lambda_z(T) \Biggr), \tag{A3}$$

where

$$\delta(Z_{ik} = z) = \begin{cases} 1 & \text{if } Z_{ik} = z, \\ 0 & \text{if } Z_{ik} \neq z. \end{cases} \tag{A4}$$

This is again the same approach employed in mixture modelling, and the expected value is equally as straightforward to obtain:

$$Q_i(\theta) = E_{Z_i|X_i,\theta_i}\, l_i(\theta_i \mid X_i, Z_i)$$
$$= \sum_{z\in z_i}\left(\sum_{k=1}^{N_i} \ln\big(\lambda_z(t_{ik})\big) \times \text{Prob}(Z_{ik} = z \mid X_i, \theta_i) - \Lambda_z(T)\right). \tag{A5}$$

This leads directly to the expression for $Q(\theta \mid \theta^{(n)})$ in Equation (9).

We now obtain the posterior probabilities in Equation (10) by following the usual procedure of providing the likelihoods and the priors and then applying Bayes' rule. For the likelihoods, we have

$$f(t_{ik} \mid Z_{ik} = z, X_i, \theta_i) = f(t_{ik} \mid Z_{ik} = z, \theta_i)$$
$$= \lambda_z(t_{ik})/\Lambda_z(T), \tag{A6}$$

where the first equality follows from the independence assumptions of the branching structure and the second from the definition of the conditional density of the inhomogeneous Poisson process (Daley & Vere-Jones, 2003, p. 23).

Next, we require an expression for the prior probability that an event belongs to process z. These can be derived directly from the result, shown in Equation (A6), that the intensity function $\lambda_z(t)$ is proportional to the density of the process z. But it is perhaps more intuitive to begin by interpreting the prior probability as the proportion of the total number of events N_i that are due to process z, which is analogous with the treatment in finite mixture modelling. This leads to

$$\text{Prob}(Z_{ik} = z \mid \theta_i) = N_z/N_i. \tag{A7}$$

Within the maximum likelihood framework, the counts N_z can be obtained from the complete data log-likelihood given in Equation (A3). This is done by solving the following two maximum likelihood equations:

$$\frac{\partial}{\partial \mu_i} l_i(\theta_i \mid X_i, Z_i) = \frac{\partial}{\partial \mu_i}\left(\sum_{k=1}^{N_i} \ln(\mu_i) \times \delta(Z_{ik} = z_{i00}) - \mu_i\, T\right)$$
$$= \frac{N_{i00}}{\mu_i} - T = 0, \tag{A8}$$

where $N_{i00} = \sum_{k=1} \delta(Z_{ik} = z_{i00})$ is the number of events in the baseline process. This implies

$$N_{i00} = \mu_i T = \Lambda_{i00}(T). \tag{A9}$$

Secondly, we consider the complete data log-likelihood for a single process $z_{ijk} \neq z_{i00}$:

$$\frac{\partial}{\partial \alpha_{ij}}\, l_i(\theta_i \mid X_i, Z_i = z_{ijk}) = \frac{\partial}{\partial \alpha_{ij}}\left(\sum_{l=1}^{N_i} \ln\big(\alpha_{ij}\ f_{ijk}(t_{il})\big) \times \delta(Z_{il} = z_{ijk}) - \alpha_{ij}\ F_{ijk}(T)\right)$$
$$= \frac{N_{ijk}}{\alpha_{ij}} - F_{ijk}(T) = 0 \tag{A10}$$

where $N_{ijk} = \sum_k \delta(Z_{ik} = z_{ijk})$ is the total number of events in process z_{ijk}, f_{ijk} is the kernel density of process z_{ijk}, and F_{ijk} the cumulative distribution function. This gives

$$N_{ijk} = \alpha_{ij} F_{ijk}(T) = \Lambda_{ijk}(T). \tag{A11}$$

From Equations (A7), (A9), and (A11) it follows that

$$\text{Prob}(Z_{ik} = z \mid \theta_i) = \Lambda_z(T) / \sum_{m\in z_i} \Lambda_m(T). \tag{A12}$$

Then using the law of total probability to obtain $f(t_{ik} \mid \theta_i)$ from Equations (A6) and (A7), Equation (10) follows from application of Bayes' rule. We note that Veen and Schoenberg (2008) did not explicitly relate the rates of the Poisson process to the posterior probabilities of the branching structure. The foregoing argument fills this void.

The derivation of Equations (11) and (12) from Equation (A5) proceeds in a manner similar to that shown in Equations (A8) and (A10), respectively. It is simply required to replace the N_z with $\bar{N}_z = \sum_k \text{Prob}(Z_{ik} = z \mid X_i, \theta_i)$ and, in Equation (A10), to sum over all processes $z \in \{z_{ij1}, \ldots, z_{ijN_j}\}$.

References

Anderson, P.A. (2008). Positions of power: status and dominance in organizational communication. In L. Guerrero & M.L. Hecht (Eds.), *The nonverbal communication reader: classical and contemporary perspectives* (3rd ed., pp. 450–468). Prospect Heights: Waveland.

Bakeman, R., & Gottman, J.M. (1997). *Observing interaction: an introduction to sequential analysis* (2nd ed.). New York: Cambridge University Press.

Bakeman, R., & Quera, V. (2011). *Sequential analysis and observations methods of the behavioural sciences*. Cambridge: Cambridge University Press.

Barabási, A.L. (2005). The origin of bursts and heavy tails in human dynamics. *Nature*, *435*, 207–211.

Brillinger, D.R. (1975). The identification of point process systems. *Annals of Probability*, *3*, 909–924.

Brillinger, D.R. (2004). Some data analyses using mutual information. *Brazilian Journal of Probability and Statistics*, *18*, 163–182.

Card, N., Selig, J., & Little, T. (2008). *Modeling dyadic and interdependent data in the developmental and behavioral sciences*. Hillsdale: Psychology Press.

Chow, S.M., Ferrer, E., & Hsieh, F. (2010). *Statistical methods for modeling human dynamics: an interdisciplinary dialogue*. New York: Taylor & Francis.

Cox, D.R., & Isham, V. (1980). *Point processes*. New York: Chapman & Hall/CRC.

Crane, R., & Sornette, D. (2008). Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 15649–15653.

Daley, D.J., & Vere-Jones, D. (2003). *An introduction to the theory of point processes: elementary theory and methods* (Vol. 1, 2nd ed.). New York: Springer.

Dayan, P., & Abbott, L.F. (2005). *Theoretical neuroscience: computational and mathematical modeling of neural systems*. Cambridge: MIT Press.

Ebel, H., Mielsch, L.-I., & Bornholdt, S. (2002). Scale-free topology of e-mail networks. *Physical Review E*, *66*, 035103.

Eckmann, J.P., Moses, E., & Sergi, D. (2004). Entropy of dialogues creates coherent structures in e-mail traffic. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 14333–14337.

Ferrer, E., Steele, J.S., & Hsieh, F. (2012). Analyzing the dynamics of affective dyadic interactions using patterns of intra- and inter-individual variability. *Multivariate Behavioral Research*, *47*, 136–171.

Girard, R. (1976). *Deceit, desire, and the novel: self and other in literary structure*. Baltimore: John Hopkins University Press.

Gottman, J.M. (1994). *Why marriages succeed or fail*. New York: Simon and Schuster.

Gottman, J.M., Murray, J.D., Swanson, C., Tyson, R., & Swanson, K.R. (2003). *The mathematics of marriage*. Cambridge: MIT Press.

Hamaker, E.L., Zhang, Z., & van der Maas, H.L.J. (2009). Using threshold autoregressive models to study dyadic interactions. *Psychometrika*, *74*, 727–745.

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

Hawkes, A.G., & Oakes, D. (1974). A cluster representation of a self-exciting process. *Journal of Applied Probability*, *11*, 493–503.

Isham, V., & Westcott, M. (1979). A self-correcting point process. *Stochastic Processes and Their Applications*, *8*, 335–347.

Jamshidian, M., & Jennrich, R.I. (2000). Standard errors for EM estimation. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *62*, 257–270.

Kalman, Y.M., David, G., & Rafaeli, D.R.R.S. (2006). Pauses and response latencies: a chronemic analysis of asynchronous CMC. *Journal of Computer-Mediated Communication*, *12*, 1–23.

Kenny, D.A., Kashy, D.A., & Cook, W.L. (2006). *Dyadic data analysis*. New York: Guilford Press.

Kenny, D.A., & Ledermann, T. (2010). Detecting, measuring, and testing dyadic patterns in the actor-partner interdependence mode. *Journal of Family Psychology*, *24*, 359–366.

Law, R., Illian, J., Burslem, D.F.R.P., Gratzer, G., Gunatilleke, C.V.S., & Gunatilleke, I.A.U.N. (2009). Ecological information from spatial patterns of plants: insights from point process theory. *Journal of Ecology*, *97*, 616–628.

Madhyastha, T.M., Hamaker, E.L., & Gottman, J.M. (2011). Investigating spousal influence using moment-to-moment affect data from marital conflict. *Journal of Family Psychology*, *25*, 292–300.

Magnusson, M.S. (2005). Discovering hidden structure with model and algorithms. In L. Anolli, S. Duncan Jr., M.S. Magnusson, & G. Riva (Eds.), *The hidden structure of interaction: from neurons to culture patterns* (pp. 4–22). Amsterdam: IOS Press.

McLachlan, G.J., & Krishnan, T. (2008). *The EM algorithm and extensions* (2nd ed.). New York: Wiley.
McLachlan, G.J., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.
Moore, C., & Dunham, P. (1995). *Joint attention: its origins and role in development*. Hillsdale: Erlbaum.
Newman, M.E.J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, *46*, 323–351.
Ogata, Y. (1978). The asymptotic behaviour of maximum likelihood estimates for stationary point processes. *Annals of the Institute of Statistical Mathematics*, *30*, 243–261.
Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, *83*, 9–27.
R Development Core Team (2011). R: a language and environment for statistical computing [Computer software manual]. Vienna, Austria. Available from http://www.R-project.org/ (ISBN 3-900051-07-0).
Rasmussen, J.G. (2011). Bayesian inference for Hawkes' processes. *Methodology and Computing in Applied Probability*. doi:10.1007/s11009-011-9272-5.
Sacks, H., Schegloff, E.A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*, 696–735.
Schoenberg, F., & Bolt, B. (2000). Short-term exciting, long-term correcting models for earthquakes. *Bulletin of the Seismological Society of America*, *90*, 849–858.
Scott, M. (2011). Affinity models for career sequences. *Journal of the Royal Statistical Society. Series C*, *60*, 417–436.
Seber, G.A.F., & Wild, C.J. (2003). *Non-linear regression* (2nd ed.). Hoboken: Wiley.
Truccolo, W., Eden, U.T., Fellows, M.R., Donoghue, J.P., & Brown, E.N. (2005). A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *Journal of Neurophysiology*, *93*, 1074–1089.
Veen, A., & Schoenberg, F.P. (2008). Estimation of space-time branching process models in seismology using an EM-type algorithm. *Journal of the American Statistical Association*, *103*, 614–624.

Manuscript Received: 4 JUL 2012
Final Version Received: 28 SEP 2012
Published Online Date: 22 FEB 2013