Assessment of e-Social Activity in Psychiatric Patients

Pablo Bonilla-Escribano, David Ramírez, *Senior Member, IEEE*, Alba Sedano-Capdevila, Juan José Campaña-Montes, Enrique Baca-García, Philipe Courtet, and Antonio Artés-Rodríguez, *Senior Member, IEEE*

***Abstract*—This paper introduces a novel method to assess the social activity maintained by psychiatric patients using information and communication technologies. In particular, we model the daily usage patterns of phone calls and social and communication apps using point processes. We propose a novel nonhomogeneous Poisson process model with periodic (circadian) intensity function using a truncated Fourier series expansion, which is inferred using a trust-region algorithm. We also extend the model using a mixture of periodic intensity functions to cope with the different daily patterns of a person. The analysis of the usage of phone calls and social and communication apps of a cohort of 259 patients reveals common patterns shared among patients with relatively high homogeneity and differences among patient pathologies.**

***Index Terms*—E-social activity, expectation-maximization (EM) algorithm, maximum likelihood (ML), mixture model, point processes.**

Manuscript received December 31, 2018; revised March 27, 2019 and May 5, 2019; accepted May 20, 2019. Date of publication May 23, 2019; date of current version November 6, 2019. This work was supported in part by MINECO under Projects aMBITION (TEC2017-92552-EXP) and CLARA (RTI2018-099655-BI00), in part by MINECO jointly with EC under Projects ADVENTURE (TEC2015-69868-C2-1-R) and CAIMAN (TEC2017-86921-C2-2-R), in part by the Comunidad de Madrid under Projects PRACTICO-CM (Y2018/TCS-4705) and IND2017/TIC-7618, Instituto de Salud Carlos III (PI16/01852), Delegación del Gobierno para el Plan Nacional de Drogas (20151073), and AFSP (LSRG-1-005-16). The work of P. Bonilla-Escribano was supported by a PIPF UC3M Scholarship. *(Corresponding author: Pablo Bonilla-Escribano.)*

P. Bonilla-Escribano, D. Ramírez, and J. J. Campaña-Montes are with the Department of Signal Theory and Communications, Universidad Carlos III de Madrid, Leganés, Madrid 28911, Spain, and also with IIS Gregorio Marañón, Madrid 28007, Spain (e-mail: pbonilla@ing.uc3m.es; david.ramirez@uc3m.es; jjcampa@tsc.uc3m.es).

A. Sedano-Capdevila is with the Department of Psychiatry, IIS Foundatión Jiménez Díaz, Madrid 28040, Spain (e-mail: alba.s.capdevila@gmail.com).

E. Baca-García is with the Department of Psychiatry, IIS Foundatión Jiménez Díaz, Madrid 28040, Spain, Department of Psychiatry, Universidad Autónoma de Madrid, Madrid 28049, Spain, Universidad Católica del Maule, Talca, Chile, and also with CIBERSAM, Spain (e-mail: ebacgar2@yahoo.es).

P. Courtet is with the Department of Psychiatric Emergency and Acute Care, Lapeyronie Hospital, University of Montpellier, Montpellier 34295, France (e-mail: p-courtet@chu-montpellier.fr).

A. Artés-Rodríguez is with the Department of Signal Theory and Communications, Universidad Carlos III de Madrid, Leganés, Madrid 28911 Spain, IIS Gregorio Marañón, Madrid 28007, Spain, and also with CIBERSAM, Spain (e-mail: antonio@tsc.uc3m.es).

Digital Object Identifier 10.1109/JBHI.2019.2918687

I. Introduction

MENTAL disorders, which affect one out of four people in the world, are among the leading causes of disability [1], [2]. They represent one of the most expensive disorders to treat, and, for instance, the estimated cost of depression treatments entails more than €118 billions per year only in Europe [3].

One preponderant reason for this high treatment cost is the lack of self-awareness of the disease [4]. Indeed, if the patients' health condition worsens, yet, they do not try to look for help, they may eventually suffer a relapse and may need to be hospitalized, which involves high economical and human burdens [5]. In an attempt to alleviate these burdens, the research effort initially focused on developing ways of remotely monitoring the patients' health condition by regularly requesting them to answer a set of health-related questions. These questions are asked via one of their devices (usually a smartphone), in such a way that the patients could provide pure, extemporaneous feedback whilst performing their usual activities. With the widespread use of smartphones, it is feasible to do so. According to the GSMA's annual Mobile Economy report, at the end of 2017, there were 5 billion mobile phones. Surprisingly enough, the number of SIM cards (7.8 billion) surpassed the world population. Nonetheless, questionnaire-based approaches suffer from serious flaws [6], such as: 1) patients' (retrospective) bias; 2) frequent data losses due to low response rates; 3) significant lower adherence associated to some specific groups, such as cannabis abuse patients; 4) the so-called "fatigue effects", that is, a decrease in active participation over time; or even 5) withdrawal, i.e. the patient discontinues the procedure.

To avoid these flaws, a new research line has recently developed tools to first gather the interactions of the patients with electronic devices, in order to establish patterns of behavior (what is referred to as digital phenotype [7]) without requiring any action from the patient. Then, data processing techniques are used to transform variations in such patterns into reliable information to diagnose diseases, evaluate their progression, or even pre-empt worsening. For instance, in [8], we studied the feasibility of detecting relapses using location data, [9] analyzed emotions using data coming from Twitter, and the work in [10] inferred online communities with interest in depression, exploiting data from social networks.

Needless to say, one of the most important interactions with devices is the one that allows interaction with other people, e.g.,

communication through a social network or a phone call, which we refer to as *e-social* activity. Undoubtedly, e-social activity is strongly related to actual social interactions [11], and it is especially interesting in millennials, since the characteristics of their social relationships are unique due to the amount of time they spend with smartphones and due to the emotional aspect [12]. There are studies that endeavor to assess the e-social activity of psychiatric patients, such as [13], where a cohort of psychiatric patients was given a mobile phone app that (among other information) obtains their e-social activity. However, the features that are extracted to describe it are rather daunting, since they are basically counts or proportions of the different call types. It must be mentioned that this kind of feature extraction is not limited to the aforementioned study, since there are open-source software packages for the same purpose that obtain similar features, such as [14]. Hence, to extract meaningful conclusions out of the aforementioned counts or call proportions, which are currently taken as the indicators of the e-social activity of the patients, different machine learning techniques have been applied. For instance, in [15] decision and regression trees are used detect depression. The research of both [16] and [17] aims to predict stress, but the former uses k-nearest neighbors, SVM and PCA, while the latter employs ensembles of tree classifiers, SVM and neural networks. Lastly, it is worth mentioning that the bivariate regression analysis in [18] is capable of finding mental health-related changes in patients with schizophrenia.

In this paper, we propose a novel approach to assess the e-social activity of psychiatric patients via the usage of phone calls and social and communication apps. To model this type of data, from a mathematical point of view, it is important to consider their particular features. Concretely, they produce a set of point observations, which are composed of a start time (a timestamp), an end time (or a duration), and possibly some other features like incoming/outgoing/missed call tags, etc. Hence, the right mathematical tool to model such observations are point processes [19]. Typically, point processes are defined through the intensity function, which measures the number of events per unit of time and may even vary over time. Actually, this temporal dependence is the feature that allows us to model the e-social activity of the patients as it can show, for instance, whether they are more active during the day or at night. From a clinical point of view, point processes have the advantage compared to the current machine learning techniques that they directly describe the e-social activity in a way that the psychiatrist (or even the patient) can understand it without requiring any kind of black-box algorithm. Moreover, the proposed technique yields a generative model, i.e., it is a data exploration tool, whereas the aforementioned machine learning techniques are discriminative tools, that is, they solve concrete problems like the stress prediction one.

To achieve more descriptive models, it is possible to consider mixtures of components for the (conditional) intensity function, since otherwise, the model would be forced to use *one single* (conditional) intensity to characterize all the e-social activity of the patient. In other words, it would be assumed that the patient always describes the same type of behavior; which may not be true, due to, e.g., potential relapses and disease remissions, or simply due to weekends and holidays. Although mixtures of Poisson point processes have been successfully applied in other fields, like vehicular accident data [20], insurance claims [21], economy [22], [23], RNA sequencing [24], or earthquake modeling [25], to the best of our knowledge, there is no previous work that has used them for the assessment of e-social activity of psychiatric patients. Actually, the use of these mixture models in psychiatry poses a series of challenges that have not been addressed in the aforementioned works. Concretely, the most important feature is the interpretability of the proposed model, i.e., a psychiatrist must be able to interpret it and obtain relevant results from the clinical point of view. To that end, only one component will be active on each day, allowing to define different types of days. Along these lines, the work in [20] divides the observations associated to each day and trains different models, which results in seven different intensities. However, our model classifies the days automatically. Further, it is important to correctly model the circadian rhythm in humans, thus, we propose to parametrize each of the components of the mixture model by means of periodic functions using a truncated Fourier series expansion. A block diagram of the proposed model is shown in Fig. 1.

The rest of the paper is organized as follows: Section II introduces a point process with periodic intensity function, as well as the estimation of the model parameters. In Section III, we present a novel point process that has an intensity function composed by a mixture of periodic functions. Moreover, the estimation of this model is also presented in this section. Finally, Section IV illustrates the good performance of the proposed methods using both synthetic and real data, and the main conclusions are summarized in Section V.

II. A Circadian Poisson Point Process Model

In this work, the e-social activity is basically measured by phone calls and the use of social networks. Thus, what we observe is *when* the activity took place (a timestamp) and *for how long* (the duration) and, additionally, we could also have other features, for instance, the type of call (incoming, outgoing, missed). This kind of data is typically modeled using point processes [19]. Concretely, since there is strong evidence supporting that human activities, such as socializing, follow a circadian rhythm, [26], [27], we propose to use a nonhomogeneous Poisson process with a *periodic* parametric function for the (conditional) intensity function.

In the following, we shall use a truncated Fourier series as a model for the intensity function, which must also be nonnegative. A sensible solution would be to take the square of the series, but we have found more convenient to use the Fejér-Riesz representation of the Fourier series [28], [29]. In particular, the Fourier series

$$\frac{a_0}{2} + \sum_{k=1}^{K} \left[a_k \cos\left(\frac{2\pi k}{T} t\right) + b_k \sin\left(\frac{2\pi k}{T} t\right) \right] \qquad (1)$$

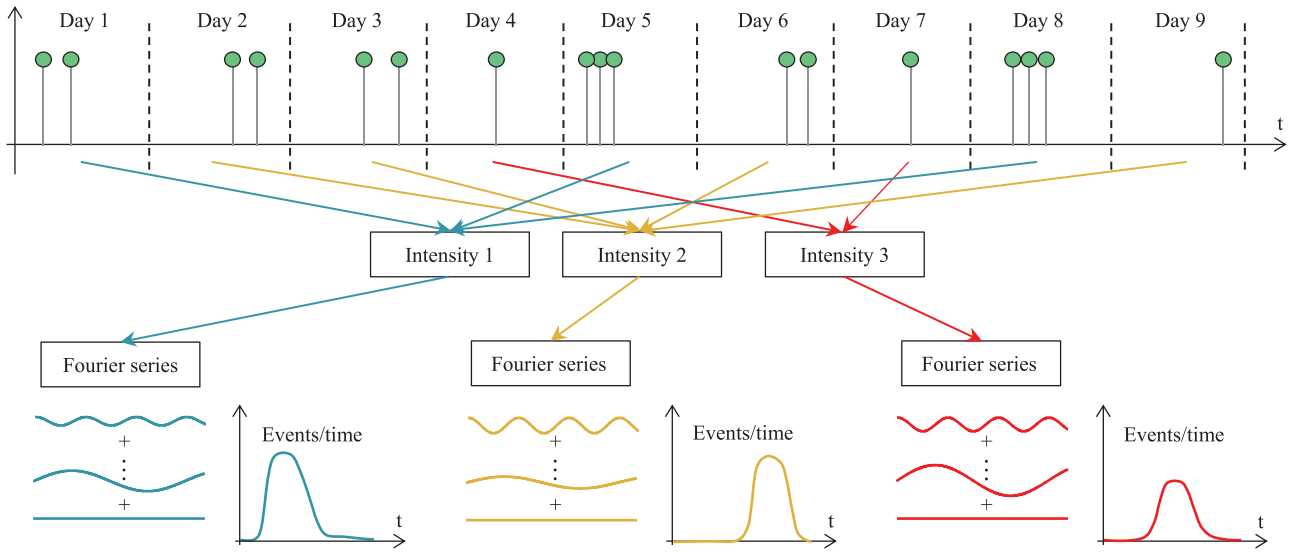

Fig. 1. Block diagram of a possible realization of the proposed model. Vertical lines with green circles on top represent the times at which an event has been observed.

is non-negative if the coefficients a_k and b_k are computed as

$$a_k + jb_k = 2\sum_{\nu=0}^{K-k} c_\nu c_{k+\nu}^*, \tag{2}$$

for any $c_k \in \mathbb{C}, k = 0, \ldots, K$, where T is the fundamental period and, in accordance with the circadian rhythm, will be set to $T = 24$ hours. Thus, (1) is the parametric family that we propose for the conditional intensity $\lambda^*(t)$.[1]

An important parameter of the proposed model is the order of the Fourier series, K. In order to achieve an interpretable model, we fix the order to $K = 3$, which is consistent with the existing literature. For instance, in [31], the authors found convenient to split a day into 4 intervals, i.e., from $0h$ to $6h$; $6h$ to $12h$; $12h$ to $18h$; and from $18h$ to $24h$, to better ascertain the circadian rhythm of mobile phone communications.

For later derivations, it is useful to express the sum in (2) as [32]

$$\sum_{\nu=0}^{K-k} c_\nu c_{k+\nu}^* = \mathbf{c}^H \mathbf{U}_k \mathbf{c}, \tag{3}$$

where $\mathbf{c} = [c_0, \ldots, c_K]^T$ and $\mathbf{U}_k \in \mathbb{R}^{(K+1)\times(K+1)}$ is a Toeplitz matrix whose entries on the kth diagonal are 1, and 0 elsewhere:

$$\mathbf{U}_k = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \cdots & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \end{bmatrix}. \tag{4}$$

[1]Throughout this paper, the asterisk follows the convention in [30], by which the conditional dependence on the past history up to time t, $\mathcal{H}_t$, is succinctly expressed as $\lambda^*(t) = \lambda(t|\mathcal{H}_t)$.

Then, the real coefficients may be recovered as

$$a_k = 2\mathbf{c}_r^T \mathbf{U}_k \mathbf{c}_r + 2\mathbf{c}_i^T \mathbf{U}_k \mathbf{c}_i, \tag{5}$$

and

$$b_k = 2\mathbf{c}_r^T \left(\mathbf{U}_k - \mathbf{U}_k^T\right) \mathbf{c}_i, \tag{6}$$

where $\mathbf{c}_r = \Re(\mathbf{c})$ is the real part of $\mathbf{c}$ and $\mathbf{c}_i = \Im(\mathbf{c})$ its imaginary part. Moreover, plugging (5) and (6) into (1), the intensity becomes

$$\lambda^*(t) = \mathbf{c}_r^T \mathbf{C}(t) \mathbf{c}_r + \mathbf{c}_i^T \mathbf{C}(t) \mathbf{c}_i + 2\mathbf{c}_r^T \mathbf{S}(t) \mathbf{c}_i, \tag{7}$$

where

$$\mathbf{C}(t) = \mathbf{I}_{K+1} + \sum_{k=1}^{K} \cos\left(\frac{2\pi k}{T} t\right) \left(\mathbf{U}_k + \mathbf{U}_k^T\right), \tag{8}$$

and

$$\mathbf{S}(t) = \sum_{k=1}^{K} \sin\left(\frac{2\pi k}{T} t\right) \left(\mathbf{U}_k - \mathbf{U}_k^T\right), \tag{9}$$

with $\mathbf{I}_{K+1}$ being the identity matrix of dimension $K+1$. Finally, defining

$$\mathbf{T}(t) = \begin{bmatrix} \mathbf{C}(t) & \mathbf{S}(t) \\ \mathbf{S}^T(t) & \mathbf{C}(t) \end{bmatrix}, \tag{10}$$

and

$$\mathbf{d} = [\mathbf{c}_r^T, \mathbf{c}_i^T]^T, \tag{11}$$

the intensity is given by

$$\lambda^*(t) = \mathbf{d}^T \mathbf{T}(t) \mathbf{d}. \tag{12}$$

Next, we derive the maximum likelihood estimator of the intensity function given a set of M realizations of the point process $\{t_j\}_{j=1}^M$ measured over the time interval $[0, T_{\text{obs}})$. Concretely,

the ML estimator of $\lambda^*(t)$ is the solution of the minimization problem [33]

$$\min_{\lambda^*(t)} \quad \int_0^{T_{\text{obs}}} \lambda^*(t)\, dt - \sum_{j=1}^{M} \log \lambda^*(t_j), \tag{13}$$

and taking into account the parametric form in (12), it becomes

$$\min_{\mathbf{d}} \quad \underbrace{\mathbf{d}^T \overline{\mathbf{T}}_{T_{\text{obs}}} \mathbf{d} - \sum_{j=1}^{M} \log \left(\mathbf{d}^T \mathbf{T}(t_j)\mathbf{d}\right)}_{J(\mathbf{d})}, \tag{14}$$

where

$$\overline{\mathbf{T}}_{T_{\text{obs}}} = \begin{bmatrix} \overline{\mathbf{C}}_{T_{\text{obs}}} & \overline{\mathbf{S}}_{T_{\text{obs}}} \\ \overline{\mathbf{S}}_{T_{\text{obs}}}^T & \overline{\mathbf{C}}_{T_{\text{obs}}} \end{bmatrix}, \tag{15}$$

with the blocks of $\overline{\mathbf{T}}_{T_{\text{obs}}}$ being

$$\begin{aligned} \overline{\mathbf{C}}_{T_{\text{obs}}} &= \int_0^{T_{\text{obs}}} \mathbf{C}(t)\, dt \\ &= T_{\text{obs}} \mathbf{I}_{K+1} + \sum_{k=1}^{K} \frac{T}{2\pi k} \sin\left(\frac{2\pi k}{T} T_{\text{obs}}\right) \left(\mathbf{U}_k + \mathbf{U}_k^T\right), \end{aligned} \tag{16}$$

and

$$\begin{aligned} \overline{\mathbf{S}}_{T_{\text{obs}}} &= \int_0^{T_{\text{obs}}} \mathbf{S}(t)\, dt \\ &= \sum_{k=1}^{K} \frac{T}{2\pi k} \left[1 - \cos\left(\frac{2\pi k}{T} T_{\text{obs}}\right)\right] \left(\mathbf{U}_k - \mathbf{U}_k^T\right). \end{aligned} \tag{17}$$

It is easy to verify that the optimization problem in (14) is not convex [34], thus requiring the use of non-convex techniques. In particular, we propose to use the trust-region method in [35], which, in turn, is based on the interior-reflective Newton's method. This technique requires both the gradient and the Hessian of the cost function, which are computed next.

Using [36] and the cost function, denoted as $J(\mathbf{d})$ in (14), we obtain the required gradient, which is given by

$$\nabla_{\mathbf{d}} J(\mathbf{d}) = 2\overline{\mathbf{T}}_{T_{\text{obs}}} \mathbf{d} - 2 \sum_{j=1}^{M} \frac{1}{\mathbf{d}^T \mathbf{T}(t_j)\mathbf{d}} \mathbf{T}(t_j)\mathbf{d}. \tag{18}$$

Similarly, the Hessian of the cost function is given by

$$\begin{aligned} \mathbf{H}J(\mathbf{d}) = 2\overline{\mathbf{T}}_{T_{\text{obs}}} + \sum_{j=1}^{M} \Bigg[&\left(\frac{2}{\mathbf{d}^T \mathbf{T}(t_j)\mathbf{d}}\right)^2 \mathbf{T}(t_j)\mathbf{d}\mathbf{d}^T \mathbf{T}(t_j) \\ &- \frac{2}{\mathbf{d}^T \mathbf{T}(t_j)\mathbf{d}} \mathbf{T}(t_j) \Bigg], \end{aligned} \tag{19}$$

and its derivation can be found in Appendix A.

One final comment is in order. Despite being a non-convex problem, the *local* minima that we have observed in an extensive number of experiments obtained with a large number of initializations are, in fact, *global* as all achieve the same value of the cost regardless of the initial point. The theoretical analysis of this behavior is out of the scope of this work and will be reported elsewhere.

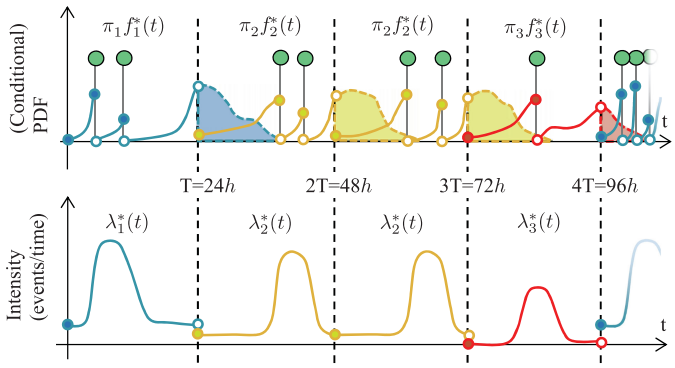

Fig. 2. Intensity mixture model. Vertical lines with green circles on top represent the times $t_{m,n}$ at which an event has been observed.

III. A Circadian Poisson Processes Mixture Model

The model presented in the previous section allows for modeling the e-social activity of psychiatric patients taking into account the circadian rhythm, i.e., the periodic structure. However, it assumes the same behavior for every day, which may not be accurate, for instance, by reason of the disease progression or weekends and holidays. Hence, in this section, we introduce a novel point process whose intensity is based on a mixture of intensities and, moreover, only one intensity may be active on each day. That is, we consider a model wherein there is a known number of (conditional) intensities N, each of which remains constant during a time slot of duration T, and could only change at times $T \times m$, where $m = 1, 2, \ldots, m'$. Here, $m' = \lfloor \frac{T_{\text{obs}}}{T} \rfloor$ is the number of time slots, since T_{obs} is the observation time, and $\lfloor \cdot \rfloor$ is the floor operator.

The aforementioned model can be defined by means of a hidden variable $z(t)$ whose value can only take on the values $1, 2, \ldots, N$, and indicates which component has generated the event at time t, i.e., which component is active at time t. Further, $\boldsymbol{\pi} = [\pi_1, \pi_2, \ldots, \pi_N]^T$ is the vector of prior probabilities on $z(t)$.

As it was presented in Section I, Fig. 1 depicts a block diagram of the proposed model for a possible scenario with a mixture of 3 components, which are based on a Poisson point process whose (conditional) intensities are calculated using a Fourier series. Needless to say, the parameters will be different depending on which (conditional) intensity is active. Hence, the subscript l will be used to indicate that those parameters take the values corresponding to the lth (conditional) intensity. In this way, we define $f_l^*(t)$ as the (conditional) density that an event happens at time t when the lth mixture component is active, and $F_l^*(t)$ as the corresponding (conditional) cumulative distribution function (CDF). Let us denote by $t_{m,n}$ the time instant of the nth observation within the mth day. Thus, $n \in \mathcal{N}(m)$, where $\mathcal{N}(m)$ is the set of indexes of the observations contained in the mth day. Finally, a single subscript represents any time instant within the mth day, i.e., t_m. With all this in mind, Fig. 2 shows graphically the first observation days that were depicted in Fig. 1, but paying especial attention to the transitions and the mathematical quantities. This will turn useful when defining the likelihood function.

The likelihood of all observations contained in the mth time slot is given by the joint density of all observed points multiplied by the probability that no point appears from the last sample until the end of the day. This probability is given by the (conditional) survival function for the lth component, $S_l^*(mT) = 1 - F_l^*(mT)$. Then, it follows that the likelihood of the complete observed sequence of days, L_{mixture}, is given by

$$L_{\text{mixture}} = \prod_{m=1}^{m'} \left\{ \prod_{l=1}^{N} \left[\pi_l \left(\prod_{n \in \mathcal{N}(m)} f_l^*(t_{m,n}) \right) \times \left(1 - F_l^*(mT)\right) \right]^{\mathbb{I}\left(z(t_m)=l\right)} \right\}. \quad (20)$$

For notational convenience, we use the "empty product" [37], which is defined to be 1 if $\mathcal{N}(m)$ is empty. Moreover, the indicator function, denoted by $\mathbb{I}(\cdot)$, is applied to all the points that are located in the same time slot. Therefore, (20) is indirectly imposing that

$$z(t_{m,n}) = z(t_{m,p}), \quad (21)$$

where $p \in \mathcal{N}(m)$, and may be different from n, ensuring that one, and only one of the N intensities is active in each of the m' time slots.

The proposed mixture model has been defined in terms of (conditional) densities and CDFs, which are cumbersome to work with. Thus, we shall rewrite (20) in terms of intensities. To that end, following [33], the densities may be rewritten as

$$f_l^*(t_{m,n}) = \lambda_l^*(t_{m,n}) e^{-\int_{t_{m,n-1}}^{t_{m,n}} \lambda_l^*(s) ds}, \quad (22)$$

where s is used in lieu of $t_{m,n}$ to emphasize that the value of the integral only depends on its limits, which yields

$$L_{\text{mixture}} = \prod_{m=1}^{m'} \left\{ \prod_{l=1}^{N} \left[\pi_l \left(\prod_{n \in \mathcal{N}(m)} \lambda_l^*(t_{m,n}) \right) \times \left(e^{-\int_{(m-1)T}^{mT} \lambda_l^*(s) ds} \right) \right]^{\mathbb{I}\left(z(t_m)=l\right)} \right\}. \quad (23)$$

Then, the ML estimation of the unknown parameters, namely $\boldsymbol{\pi}$ and the N intensities $\lambda_l^*(t)$, boils down to

$$\underset{\boldsymbol{\pi}, \{\lambda_l^*(t)\}_{l=1}^N}{\operatorname{argmax}} L_{\text{mixture}} = \underset{\boldsymbol{\pi}, \{\lambda_l^*(t)\}_{l=1}^N}{\operatorname{argmax}} \log L_{\text{mixture}}. \quad (24)$$

A. ML Estimation Based on the EM Algorithm

The log-likelihood expression in (23) cannot be maximized in closed form since there are some quantities, referred to as *hidden variables*, which are not directly observed from the data. To circumvent this, the so-called Expectation-Maximization (EM) algorithm will be employed [38]. The basic principle behind the EM algorithm is to iterate between the expectation or E step and the maximization or M step. In the E step, the expectation of the log-likelihood is computed with respect to the hidden variables, while the unknown parameters are considered fixed. Afterwards, in the M step, the estimates of the unknown parameters are obtained by maximizing the expected log-likelihood, which does not depend on the hidden variables. Finally, if these two steps are repeated, it can be shown that the algorithm converges to a local maximum of the log-likelihood [39]. A block diagram that serves as an overview of how the EM algorithm is employed to solve (24) is shown in Fig. 3.

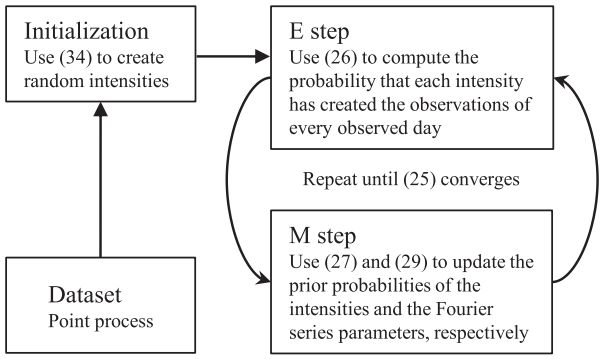

Fig. 3. Block diagram of the EM algorithm.

To proceed, let us compute the expected value with respect to the hidden variables conditioned on the observed data points and model parameters (E step). This expectation is given for the proposed model by

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(g-1)}) = \sum_{m=1}^{m'} \sum_{l=1}^{N} r_{m,l} \left[\log\left(\pi_l\right) - \int_{(m-1)T}^{mT} \lambda_l^*(s)\, ds + \sum_{n \in \mathcal{N}(m)} \log\left(\lambda_l^*(t_{m,n})\right) \right], \quad (25)$$

where $\boldsymbol{\theta}$ stands for the model parameters, g is the current iteration number, and thus $\boldsymbol{\theta}^{(g-1)}$ are the previously computed parameters. Further, let $\{t_m\}$ denote the set of all data points contained in the mth time slot, and $r_{m,l} = p\left(z(t_m) = l | \{t_m\}, \boldsymbol{\theta}^{(g-1)}\right)$ be the responsibilities. The responsibilities can be understood as the probability that the samples contained in the mth day have been produced by the lth intensity and they can therefore be computed as

$$r_{m,l} = \frac{\pi_l \left(\prod_{n \in \mathcal{N}(m)} \lambda_l^*(t_{m,n}) \right) e^{-\int_{(m-1)T}^{mT} \lambda_l^*(s) ds}}{\sum_{l'=1}^{N} \pi_{l'} \left(\prod_{n \in \mathcal{N}(m)} \lambda_{l'}^*(t_{m,n}) \right) e^{-\int_{(m-1)T}^{mT} \lambda_{l'}^*(s) ds}}. \quad (26)$$

This expression may suffer from numerical problems, which may be partially avoided using the so-called the log-sum-exp trick [40].

Now, we are left with the maximization of $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(g-1)})$ with respect to the prior probabilities and the intensities. The optimization of the prior probabilities is independent of that of the remaining parameters, and may be carried out using the method of Lagrange multipliers, which yields

$$\pi_l = \frac{1}{m'} \sum_{m=1}^{m'} r_{m,l}. \quad (27)$$

Thus, the estimated prior probabilities are given by the weighted number of days assigned to the lth component divided by the number of observed days.

The next step is to find the estimates of the N (conditional) intensity functions. Looking carefully at (25), it is easy to note that the optimization with respect to each (conditional) intensity is independent from the others, and ignoring constant terms is equivalent to

$$\underset{\lambda_l^*(t)}{\operatorname{argmin}} \sum_{m=1}^{m'} r_{m,l} \left[\int_{(m-1)T}^{mT} \lambda_l^*(s)\, ds - \sum_{n \in \mathcal{N}(m)} \log\left(\lambda_l^*(t_{m,n})\right) \right]. \tag{28}$$

As can be seen, (28) is almost identical to the ML estimation problem for one intensity in (14). Thus, we may use in a straightforward manner the periodic model in Section II and its ML estimator, yielding a mixture of periodic intensities. Concretely, the resulting optimization problem is

$$\underset{\mathbf{d}_l}{\operatorname{argmin}}\ \mathbf{d}_l^T \left(\sum_{m=1}^{m'} r_{m,l} \overline{\mathbf{T}}_m \right) \mathbf{d}_l - \left[\sum_{m=1}^{m'} \sum_{n \in \mathcal{N}(m)} r_{m,l} \log\left(\mathbf{d}_l{}^T \mathbf{T}(t_{m,n}) \mathbf{d}_l \right) \right], \tag{29}$$

where

$$\overline{\mathbf{T}}_m = \begin{bmatrix} \overline{\mathbf{C}}_m & \overline{\mathbf{S}}_m \\ \overline{\mathbf{S}}_m^T & \overline{\mathbf{C}}_m \end{bmatrix}, \tag{30}$$

with the blocks of $\overline{\mathbf{T}}_m$ being

$$\overline{\mathbf{C}}_m = \int_{(m-1)T}^{mT} \mathbf{C}(t)\, dt, \tag{31}$$

and

$$\overline{\mathbf{S}}_m = \int_{(m-1)T}^{mT} \mathbf{S}(t)\, dt. \tag{32}$$

B. Initialization

Once both the E and M steps are set up, the only element that is left to establish is the initialization. In this regard, it is typically considered that all intensities are equally probable at the first iteration [41], i.e.,

$$\pi_l = \frac{1}{N}, \quad l = 1, \ldots, N. \tag{33}$$

On the other hand, concerning the parameters of the functions to be optimized in EM routines, in this case the complex vector $\mathbf{c}_{\text{ini}}$, the usual approach is to consider several initializations and select the one that achieves a larger value of the auxiliary function. However, we must choose these initializations wisely since, otherwise, the algorithm will converge to a solution in which one single intensity prevails. The reason for this can be found in (26), which shows that the responsibilities depend on the exponential of the negative of the integral of the (conditional) intensities over a time slot. Therefore, the initializations are chosen such that

$$\int_{(m-1)T}^{mT} \lambda^*(t)\, dt = \|\mathbf{c}\|^2\, T, \tag{34}$$

is constant. Specifically, the initialization scheme consists in sampling from a zero-mean Gaussian distribution with identity covariance matrix, which is afterwards normalized to have the norm of the solution obtained for a model with just one intensity component.

TABLE I
SYNTHETIC DATA RESULTS

Amplitude	Number of data-points	Number of components		NSE($\times 10^{-4}$)			
		BIC	AIC	a_0	$a_{k\neq 0}$	b_k	π_l
15	2721	4	4	6.5	29	13	587
10	1935	4	4	9.8	35	24	3778
5	942	4	4	78	182	156	2889
2	417	4	4	99	756	136	1196
0.5	137	2	3	1511	6183	2415	4010

IV. RESULTS

In this section, the performance of the proposed method is evaluated. First, to validate the inference algorithm, we consider an experiment with synthetic data. Second, to illustrate its applicability in psychiatry, we apply the proposed technique to real data acquired by the smartphone of psychiatric patients.

A. Validation of the Estimation Procedure Using Synthetic Data

In the following, we show the good performance of the proposed inference algorithm. Concretely, we consider a mixture of $N = 4$ intensity functions, each of which is a Fourier series with order $K = 3$. The number of simulated days is 30 and we vary the amplitudes of the Fourier series, which yields a different number of observations, as shown in Table I. This table also shows the normalized squared error (NSE) of the estimated parameters obtained by the proposed EM algorithm in one execution, showing good performance values. We took the NSE as the square of the norm of the difference between the estimated and simulated parameters, divided by the square of the norm of the simulated parameters. Additionally, for the EM algorithm, we found that 10 initializations suffice and increasing them does not provide any significant performance improvement.

The aforementioned analysis assumed that the number of components, N, was known in advance. When this is not the case, it must be estimated. To that end it is possible to use methods based on information theoretic criteria, such as the Akaike Information Criterion (AIC) [42] or the Bayesian Information Criterion (BIC) [43]. Thus, we compute the model for orders $N = 1, 2, \ldots, 8$ and select the one with smallest value of the criterion. The results for the order estimation are also shown in Table I, where we can see that both criteria select the right order

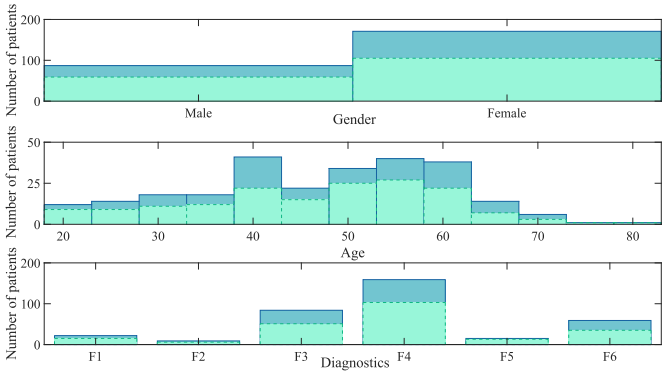

Fig. 4. Gender, age, and main diagnostic of the sample population. Blue: phone usage sample; green: app usage sample.

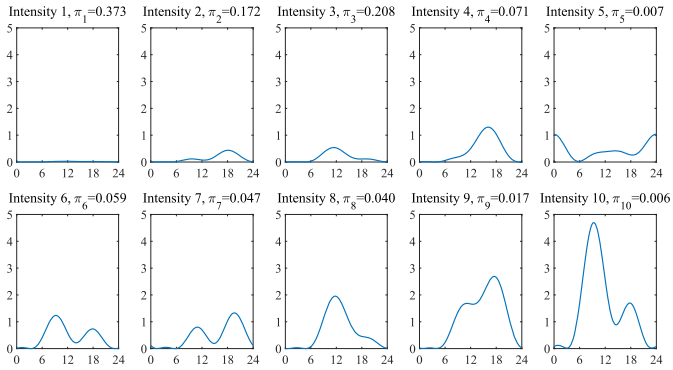

Fig. 5. Intensity functions obtained by the proposed method from the calls made by the sample of 259 patients. The abscissa represents the hour of the day, while the ordinate represents the intensity in events/h.

in most cases. Actually, the fact that the method may filter out some weak intensities can be a desirable feature, since it may remove some spurious intensities, what can greatly simplify the clinical interpretation of the results.

B. Analysis of e-Social Activity

The aim of this preliminary study is to discover e-social activity patterns in psychiatric patients and whether patients with different pathologies have different patterns. Concretely, in this study, a cohort of psychiatric patients from Hospital Fundación Jiménez Díaz in Madrid volunteered to participate in a long-term study by downloading a smartphone app designed to collect their interactions with mobile devices. The eB2 app [44] gathered the following data: 1) actigraphy; 2) GPS location; 3) Google location; 4) app usage; 5) log of both phone calls and messages; 6) nearby Wi-Fi and Bluetooth devices; and 6) inertial measurement unit signals. Although this study only considers the usage of apps and phone calls.

The exclusion criteria for the study were patients that match at least one of the following conditions: 1) failure to provide signed consent (IRB approval: Pic 66/2017 FJD and EO 76/2013_FJD_HIE_HRJC); 2) the patient is illiterate; 3) the patient does not own a smartphone that is compatible with the app; 4) the patient suffers from a mental disease classified as F0, or F7 through F9 according to the ICD codes [45]; and 5) there are less than 30 days of observations available.

Since we analyze two aspects of the e-social activity of the patients, namely the usage of phone calls and social and communication apps, there are two subsets of patients that are considered: one for the 259 psychiatric patients (mean age of approximately 47 years and 8 months, 88 males and 171 females) for which there were more than 1 month of phone call data; and out of them, the 164 patients (mean age of roughly 47 years and 3 months, 59 males and 105 females) for which there were also more than 1 month of data about the usage of the social and communication apps. Fig. 4 summarizes the demographics of both groups (note that some patients may have comorbidity).

First, we apply the proposed Poisson mixture model to obtain the intensity functions of all the calls in which the patients were involved (regardless of the type of calls). Fig. 5 shows such

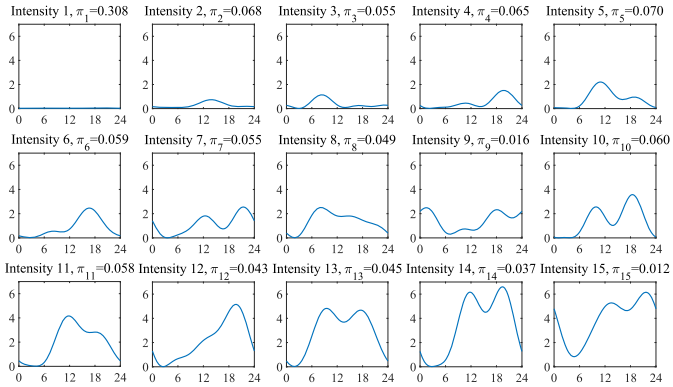

Fig. 6. Intensity functions obtained by the proposed method from the usage of social and communication apps of the sample of 164 patients. The abscissa represents the hour of the day, while the ordinate represents the intensity in events/h.

intensity functions, ordered in increasing number of expected number of calls. As can be seen, the method selects 10 intensities to describe the behavior of the 259 patients. This figure shows that the intensity 1 can be understood as a sign of poor social interaction via phone calls, whereas intensities 2, 6, 7, and 10 show two periods of call use: morning and afternoon. Hence, patients that have these intensities are more active during those periods. On the other hand, intensities 3, 4, 8, and 9 do not show this two-period activity variation. Instead, the activity increases throughout the day until the summit and then decreases. Moreover, note that intensities 3 and 8 are similar in shape but not in amplitude. Additionally, intensity 5 is interesting, since it reveals a nocturnal pattern, which is quite different from the rest. Similar conclusions can be drawn from the usage of social and communication apps, but the intensities tend to be higher, as depicted in Fig. 6.

Finally, Fig. 7 shows the proportion of each call intensity, and we observe that the nocturnal intensity (number 5) is one of the least likely in all diagnoses, albeit it is more frequent in the F5 group, where patients with insomnia are included. It is interesting to notice that patients suffering from depression (F3) show a higher proportion of the intensities

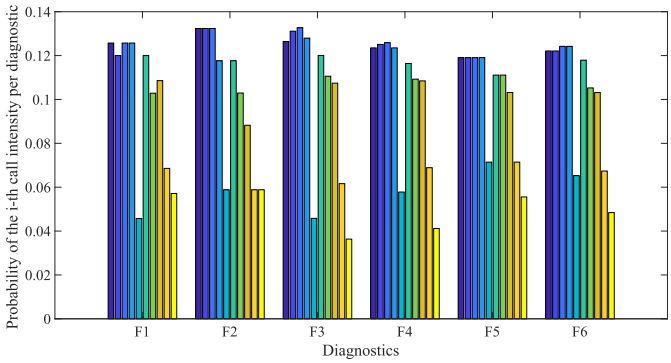

Fig. 7. Probability of the call intensities for each type of diagnosis.

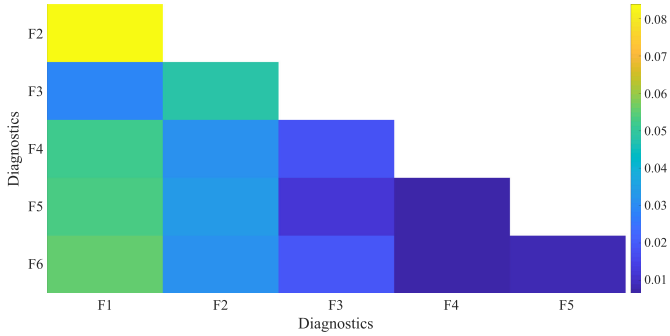

Fig. 8. Differences in the probability of the intensities describing the usage of social and communication apps, computed via the symmetric KL divergence.

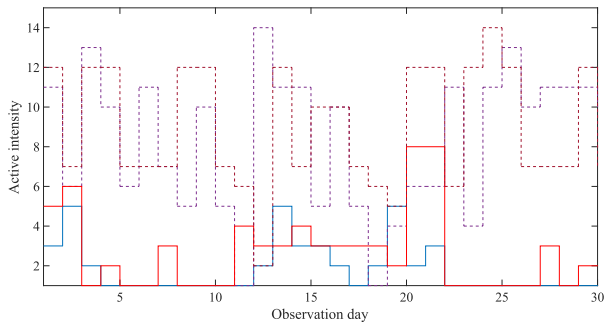

Fig. 9. Activation of the intensities describing the usage of social and communication apps. Solid lines: 2 example patients diagnosed with F1; dashed lines: 2 example patients diagnosed with F2.

that involve a lower number of calls. A different picture is observed in the patients suffering from obsessive disorders, included in the F4 group, since they have higher proportions of the intensities involving more phone calls. Finally, we assess how the proportion of the used (or activated) intensities on each day differ amongst the different diagnostic groups. In order to do so, we compute the symmetric KL divergence [46] of the probability of using each of the social and communication intensities. In Fig. 8 we can notice that the groups with more different behaviors from each other are the patients suffering from substance abuse (F1), and the patients with psychotic disorders (F2). This can be explained by the higher proportion of low-activity intensities in F1 patients, whilst F2 patients have a considerably big proportion of the 6th intensity (socializing during the afternoon), as depicted in Fig. 9.

V. Conclusion and Future Work

In this paper, we have presented a novel Poisson process mixture model, which can be applied to characterize the e-social activity of psychiatric patients. Moreover, to capture the circadian rhythm present in the considered application, we have proposed that each component of the mixture is given by a truncated Fourier series. From an algorithmic point of view, we have shown the good performance of the estimation technique, which is based on the EM algorithm.

From the clinical point of view, we have found a set of interpretable patterns analyzing data coming from the usage of phone calls and social and communication apps. For instance, the method has discovered patterns of nocturnal activity and poor social interactions. Moreover, since social activity is an integral part of the patient's behavior, we believe that the methods presented in this work could serve as a basis to help with the diagnosis of mental illnesses and to design the treatment.

In the future, to facilitate the clinical adoption of this novel approach in psychiatry, it would be of great interest to extend the model by incorporating complementary information, such as the type of call or the duration. Finally, notice that the approach of this paper has been *agnostic*, in the sense that a *global* model has been used to produce personalized profiles. Then, those profiles were *naturally* found to be significantly different amongst different patient's groups, but the group that each patient belongs to was not known to the model *beforehand*. This approach has the advantage that it will tend to reduce the bias of the results. Nonetheless, it would be interesting to see if more descriptive results could be obtained if the demographics and/or the diagnostics of the patients are explicitly taken into account by the model as well.

Appendix A
Derivation of the Hessian of the Cost Function

The Hessian of the cost function can be computed as the derivative of the gradient in (18), that is,

$$\begin{aligned}\mathbf{H}J(\mathbf{d}) &= \frac{\partial}{\partial \mathbf{d}^T}\left(2\overline{\mathbf{T}}_{T_{\text{obs}}}\mathbf{d} - 2\sum_{i=1}^{N}\frac{1}{\mathbf{d}^T\mathbf{T}(t_i)\mathbf{d}}\mathbf{T}(t_i)\mathbf{d}\right)\\ &= 2\frac{\partial}{\partial \mathbf{d}^T}\overline{\mathbf{T}}_{T_{\text{obs}}}\mathbf{d} - 2\sum_{i=1}^{N}\frac{\partial}{\partial \mathbf{d}^T}\frac{1}{\mathbf{d}^T\mathbf{T}(t_i)\mathbf{d}}\mathbf{T}(t_i)\mathbf{d}\\ &= 2\overline{\mathbf{T}}_{T_{\text{obs}}} - 2\sum_{i=1}^{N}\frac{\partial}{\partial \mathbf{d}^T}\frac{1}{\mathbf{d}^T\mathbf{T}(t_i)\mathbf{d}}\mathbf{T}(t_i)\mathbf{d}. \qquad (35)\end{aligned}$$

Now, we need to compute the derivative of the right-hand side of (35). First, using the product rule [36], this derivative becomes

$$\begin{aligned}\frac{\partial}{\partial \mathbf{d}^T}\left(\frac{1}{\mathbf{d}^T\mathbf{T}(t_i)\mathbf{d}}\mathbf{T}(t_i)\mathbf{d}\right) &= \mathbf{T}(t_i)\mathbf{d}\frac{\partial\left(\mathbf{d}^T\mathbf{T}(t_i)\mathbf{d}\right)^{-1}}{\partial \mathbf{d}^T}\\ &+ \left(\mathbf{I}_{2K+2}\otimes\left(\mathbf{d}^T\mathbf{T}(t_i)\mathbf{d}\right)^{-1}\right)\frac{\partial \mathbf{T}(t_i)\mathbf{d}}{\partial \mathbf{d}^T}, \qquad (36)\end{aligned}$$

and the proof concludes by applying the chain rule [47] and simplifying the above equation as follows

$$\begin{aligned}&\mathbf{H}J(\mathbf{d})\\&= \mathbf{T}(t_i)\mathbf{d}\frac{\partial\left(\mathbf{d}^T\mathbf{T}(t_i)\mathbf{d}\right)^{-1}}{\partial\left(\mathbf{d}^T\mathbf{T}(t_i)\mathbf{d}\right)}\frac{\partial\left(\mathbf{d}^T\mathbf{T}(t_i)\mathbf{d}\right)}{\partial\mathbf{d}^T}+\frac{1}{\mathbf{d}^T\mathbf{T}(t_i)\mathbf{d}}\mathbf{T}(t_i)\\&= -\frac{2}{\left(\mathbf{d}^T\mathbf{T}(t_i)\mathbf{d}\right)^2}\mathbf{T}(t_i)\mathbf{d}\mathbf{d}^T\mathbf{T}(t_i)+\frac{1}{\mathbf{d}^T\mathbf{T}(t_i)\mathbf{d}}\mathbf{T}(t_i). \quad (37)\end{aligned}$$

Appendix B
List of Symbols

Refer to Table II for a list of the main symbols used in this work.

TABLE II
List Of Main Symbols (By Order of Appearance)

Symbol	Explanation
a_k, b_k	Real coefficients of the Fourier series
c_k	Fejér-Riesz coefficients for the Fourier series
$\mathbf{c}$	Vector with Fejér-Riesz coefficients
K	Order of the Fourier series
$\lambda^*(t)$	(Conditional) intensity function
N	Number of mixture components
T	Fundamental period of the Fourier series
T_{obs}	Observation time
$z(t)$	Hidden variable selecting the mixture component
$\boldsymbol{\pi}$	Vector of prior probabilities on $z(t)$, with components π_l
$f_l^*(t)$	(Conditional) density when the lth intensity is active
$F_l^*(t)$	Cumulative distribution function (CDF) when the lth intensity is active
$t_{m,n}$	Time instant of the nth observation of the mth day
t_m	Any time instant within the mth day
$S_l^*(t)$	Survival function when the lth intensity is active
$\mathbb{I}(\cdot)$	Indicator function
$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(g-1)})$	Auxiliary function of the EM algorithm
$\{t_m\}$	Set of all data points contained in the mth day
$r_{m,l}$	Probability that the lth intensity has generated the data contained in the mth day

References

[1] World Health Organization, "The world health report 2001 - Mental health: New understanding, new hope," 2001.

[2] H. A. Whiteford, A. J. Ferrari, L. Degenhardt, V. Feigin, and T. Vos, "The global burden of mental, neurological and substance use disorders: An analysis from the global burden of disease study 2010," *PloS One*, vol. 10, no. 2, 2015, Art. no. e0116820.

[3] P. Sobocki, B. Jönsson, J. Angst, and C. Rehnberg, "Cost of depression in Europe," *J. Mental Health Policy Econ.*, vol. 9, no. 2, pp. 87–98, 2006.

[4] X. F. Amador *et al.*, "Awareness of illness in schizophrenia and Schizoaffective and mood disorders," *Archives Gen. Psychiatry*, vol. 51, no. 10, pp. 826–836, 1994.

[5] B. L. Svarstad, T. I. Shireman, and J. Sweeney, "Using drug claims data to assess the relationship of medication adherence with hospitalization and costs," *Psychiatric Services*, vol. 52, no. 6, pp. 805–811, 2001.

[6] E. Moitra, B. A. Gaudiano, C. H. Davis, and D. Ben-Zeev, "Feasibility and acceptability of post-hospitalization ecological momentary assessment in patients with psychotic-spectrum disorders," *Comprehensive Psychiatry*, vol. 74, pp. 204–213, 2017.

[7] S. H. Jain, B. W. Powers, J. B. Hawkins, and J. S. Brownstein, "The digital phenotype," *Nature Biotechnol.*, vol. 33, no. 5, pp. 462–463, 2015.

[8] S. Berrouiguet *et al.*, "Combining continuous smartphone native sensors data capture and unsupervised data mining techniques for behavioral changes detection: The feasibility study of the Evidence Based Behavior (eB2) platform," *JMIR mHealth uHealth*, vol. 6, no. 12, 2018, Art. no. e197.

[9] M. E. Larsen, T. W. Boonstra, P. J. Batterham, B. O'Dea, C. Paris, and H. Christensen, "We feel: Mapping emotion on Twitter," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 4, pp. 1246–1252, Jul. 2015.

[10] B. Saha, T. Nguyen, D. Phung, and S. Venkatesh, "A framework for classifying online mental health-related communities with an interest in depression," *IEEE J. Biomed. Health Inform.*, vol. 20, no. 4, pp. 1008–1015, Jul. 2016.

[11] N. B. Ellison, C. Steinfield, and C. Lampe, "The benefits of Facebook "friends:" Social capital and college students' use of online social network sites," *J. Comput.-Mediated Commun.*, vol. 12, no. 4, pp. 1143–1168, 2007.

[12] W. Youyou, M. Kosinski, and D. Stillwell, "Computer-based personality judgments are more accurate than those made by humans," *Proc. Nat. Acad. Sci.*, vol. 112, no. 4, pp. 1036–1040, 2015.

[13] D. Ben-Zeev *et al.*, "Crosscheck: Integrating self-report, behavioral sensing, and smartphone use to identify digital indicators of psychotic relapse," *Psychiatric Rehabil. J.*, vol. 40, no. 3, pp. 266–275, 2017.

[14] Y.-A. de Montjoye *et al.*, "bandicoot: A Python toolbox for mobile phone metadata," *J. Mach. Learn. Res.*, vol. 17, pp. 1–5, 2016.

[15] M. N. Burns *et al.*, "Harnessing context sensing to develop a mobile intervention for depression," *J. Med. Internet Res.*, vol. 13, no. 3, 2011, Art. no. e55.

[16] A. Sano and R. W. Picard, "Stress recognition using wearable sensors and mobile phones," in *Proc. Humaine Assoc. Conf. Affective Comput. Intell. Interact.*, 2013, pp. 671–676.

[17] A. Bogomolov, B. Lepri, M. Ferron, F. Pianesi, and A. S. Pentland, "Daily stress recognition from mobile phone data, weather conditions and individual traits," in *Proc. 22nd ACM Int. Conf. Multimedia*, 2014, pp. 477–486.

[18] R. Wang *et al.*, "Crosscheck: Toward passive sensing and detection of mental health changes in people with schizophrenia," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2016, pp. 886–897.

[19] J. F. C. Kingman, *Poisson Processes*. Oxford, U.K.: Clarendon Press, 1992, vol. 3.

[20] A. T. Ihler and P. Smyth, "Learning time-intensity profiles of human activity using non-parametric Bayesian models," in *Proc. 20th Annu. Conf. Neural Inf. Process. Adv. Neural Inf. Process. Syst.*, 2007, pp. 625–632.

[21] G. O. Brown and W. S. Buckley, "Experience rating with Poisson mixtures," *Ann. Actuarial Sci.*, vol. 9, no. 2, pp. 304–321, 2015.

[22] S. Faria and F. Gonçalves, "Financial data modeling by Poisson mixture regression," *J. Appl. Statist.*, vol. 40, no. 10, pp. 2150–2162, 2013.

[23] M. Burda, M. Harding, and J. Hausman, "A Poisson mixture model of discrete choice," *J. Econometrics*, vol. 166, no. 2, pp. 184–203, 2012.

[24] H. Wu, Z. Qin, and Y. Zhu, "PM-Seq: Using finite Poisson mixture models for RNA-Seq data analysis and transcript expression level quantification," *Statist. Biosci.*, vol. 5, no. 1, pp. 71–87, 2013.

[25] A. C. Micheas and J. Chen, "sppmix: Poisson point process modeling using normal mixture models," *Comput. Statist.*, vol. 33, no. 4, pp. 1767–1798, 2018.

[26] W. Huang, K. M. Ramsey, B. Marcheva, and J. Bass, "Circadian rhythms, sleep, and metabolism," *J. Clin. Investigation*, vol. 121, no. 6, pp. 2133–2141, 2011.

[27] A. Sorokin, A. Maksimov, and J. Jermain, "Human circadian rhythm synchronization by social timers: The role of motivation: IV. Individual features of the free-running 24-hour sleep-wake cycle under simulated conditions of vital activity," *Human Physiol.*, vol. 26, no. 1, pp. 41–47, 2000.

[28] L. Fejér, "Über trigonometrische polynome," *J. Für Die Reine Und Angewandte Mathematik*, vol. 146, pp. 53–82, 1916.

[29] F. H. Fröhner, "The Riesz–Fejér theorem: Missing link between probability theory and quantum mechanics," *Wissenschaftliche Berichte FZKA*, 1997.

[30] D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods.* New York, NY, USA: Springer-Verlag, 2003.

[31] T. Aledavood *et al.*, "Daily rhythms in mobile telephone communication," *PloS One*, vol. 10, no. 9, 2015, Art. no. e0138098.

[32] L. Zhao, J. Song, P. Babu, and D. P. Palomar, "A unified framework for low autocorrelation sequence design via majorization–minimization," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 438–453, Jan. 2017.
[33] D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure*. New York, NY, USA: Springer-Verlag, 2008.
[34] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[35] T. F. Coleman and Y. Li, "An interior trust region approach for nonlinear minimization subject to bounds," *SIAM J. Optim.*, vol. 6, no. 2, pp. 418–445, 1996.
[36] J. R. Magnus and H. Neudecker, "Matrix differential calculus with applications to simple, Hadamard, and Kronecker products," *J. Math. Psychol.*, vol. 29, no. 4, pp. 474–492, 1985.
[37] O. Boxma, O. Kella, and M. Mandjes, "On a generic class of Lévy-driven vacation models," *Probability Eng. Informational Sci.*, vol. 24, no. 1, pp. 1–12, 2010.
[38] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc. Ser. B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.
[39] C. F. J. Wu, "On the convergence properties of the EM algorithm," *Ann. Statist.*, vol. 11, no. 1, pp. 95–103, 1983.
[40] D. Ramos-López *et al.*, "Scalable importance sampling estimation of Gaussian mixture posteriors in Bayesian networks," *Int. J. Approx. Reasoning*, vol. 100, pp. 115–134, 2018.
[41] B. J. Frey and N. Jojic, "Transformation-invariant clustering using the EM algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 1, pp. 1–17, Jan. 2003.
[42] Y. Sakamoto, M. Ishiguro, and G. Kitagawa, *Akaike Information Criterion Statistics* (ser. Mathematics and its Applications). The Netherlands: Springer, 1986.
[43] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, 1978.
[44] "Evidence-Based Behavior, Improving mental health knowledge." 2018. [Online]. Available: https://eb2.tech/our-solution/
[45] World Health Organization, *International Statistical Classification of Diseases and Related Health Problems*, 2nd ed. Geneva, Switzerland: World Health Organization, 2004, vol. 2.
[46] P. J. Moreno, P. P. Ho, and N. Vasconcelos, "A Kullback–Leibler divergence based kernel for SVM classification in multimedia applications," in *Proc. Advances Neural Inf. Process. Syst.*, 2004, pp. 1385–1392.
[47] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus With Applications in Statistics and Econometrics*. Hoboken, NJ, USA: Wiley, 2019.