

Point Process Intensity Shape Identification Based on Available Precedents Stochastic Descriptions

Viacheslav Antsiperov and Aleksei Morozov
Kotelnikov Institute of Radioengineering of RAS, Moscow, Russia
Email: antciperov@cplire.ru

Abstract—The article discusses the inhomogeneous point process intensity reconstruction for the case when a process is given by a realization of its discrete point set, samples, whose intensity of appearance is not known. The specificity of the problem under consideration is that a priori information is also assumed to be associated only with the registered data of previously analyzed similar processes - referred to in the paper as precedents. So, in the frames of the approach the reconstruction problem is posed as the statistical identification of the registered data with already observed precedents, rather than the traditional in statistics problem of hypothesis testing. The solution to the problem is proposed for the special class of point processes - Poisson Point Processes. Identification of the recorded PPP with one of the precedents stored in prepared database is implied up to a shift and scale transforms. The identification procedure synthesis and precedents database refill are considered in the frames of the maximum likelihood approach, whose implementation is carried out according to the principles of machine learning. The description of the point process registered and precedents are chosen as a mixture (superposition) of Gaussian components. Recurrent calculation of the log-likelihood function is structured in the form of EM-like algorithm adapted to the problem.

Index Terms—inhomogeneous point process, Poisson process intensity identification, machine learning, EM algorithm, Gaussian mixtures, effective computational schemes

I. INTRODUCTION

The main problem considered in the paper is associated with the statistical inference about the intensity of an inhomogeneous point process, whose realization is given as a set of discrete points (events) [1]. This problem is considered in relation to inhomogeneous Poisson process – a convenient and flexible tool for implementing imitation models and simulating real systems. It should be noted that although the Poisson process is a rather special case of general point processes – the simplest, maximally random process – the number of its applications is large today and continues to grow rapidly. As an example, we indicate such areas of Poisson processes applications as: astronomy [2], biology [3], ecology [4], seismology [5], telecommunications [6], etc.

As the number of applications is large, there is a great number of methods and approaches to Poisson process analysis and processing (including the above-mentioned problem). One of the first works dedicated to the analysis and estimation of Poisson process characteristics, its intensity are the papers of Cox [7], [8]. They are characterized by the systematic application of traditional statistical methods to the class of stochastic point processes. Several problems of evaluating the Poisson process intensities and comparing the intensities of several Poisson processes have been considered and solved. But it should be noted, that most problems were solved only for the case of homogeneous Poisson process. The inhomogeneous case analytically is greatly complicated, and therefore the problems of estimating the temporal dynamics of intensity where not practically considered by Cox (excluding the problem of trend estimation).

Along with the development of computer technology, statistical methods began to acquire an algorithmic character and increasingly correspond to the principles and approaches of machine learning (the branch of modern computer science) [9], [10]. Several rather complicated problems were solved by combining computational power of modern computers and computing schemes of some algorithms (usually of recurrent type), developed during the last 20-30 years [11]. For example, several problems of estimating the Poisson process intensity, that widely use methods of machine learning are set forth in the book [12]. The author of the book considers parametric intensities and discusses approaches to estimating parameters depending on the type of the problem – either it is related to the estimation of the intensity of a single process or to superposition of the processes. It is noted that in the latter case the EM (expectation-maximization) algorithm for Gaussian mixtures is most effective. Several different approaches concerning the estimation of random intensities are presented in [1]. The author of the proceedings considers Poisson processes with random intensities as twice stochastic, or “Cox processes”, and proposes his own method for estimating the intensity based on the Monte Carlo method for Markov processes.

In this paper, we also propose a new method for identifying the time dynamics of the intensity of

inhomogeneous Poisson process based on a certain variant of machine learning [13]. In accordance with the principles of statistical (machine) learning, the problem considered is posed as the statistical inference via already observed data, precedents [10], rather than the traditional statistics problem of parameters estimation. So, it is assumed that a finite set of precedents (examples) of Poisson process intensity (training set) has been observed, for each precedent some description has been formed and it is required to determine to which of the precedents observed should be attributed the newly registered inhomogeneous Poisson point process, given by a set of discrete points.

II. MODEL

To solve the problem discussed in the introduction in the spirit of the principles of machine learning, it is necessary to clearly define the corresponding initial assumptions. Let us define two main groups of assumptions. The first group concerns the characteristics of Poisson process used in the description of precedents as database entities. The second is related to the type of dependence of identified process and target precedent from the formed database in terms of specified precedent characteristics.

Let us assume that each precedent is the intensity $\lambda_k(t)$ of Poisson process, which is predefined for $(0, T)$ time domain. We also assume that this intensity could be approximated by a superposition of N_k Gaussian components:

$$\lambda_k(t) = \Lambda_k \sum_{j=1}^{N_k} p_{jk} \frac{1}{\sqrt{2\pi\sigma_{jk}^2}} \exp\left(-\frac{(t - \tau_{jk})^2}{2\sigma_{jk}^2}\right) \quad (1)$$

where Λ_k is the integral over $(0, T)$ of intensity $\lambda_k(t)$, p_{jk} is the fraction of total intensity for j -th component, τ_{jk} , σ_{jk} are parameters of this component. Let us note, that Λ_k is also the parameter for Poisson distribution of total number n of representing consequent Poisson process points.

The set of the above listed parameters fully describes intensity (1). However, for comparing the intensities not all of them are equally important. For example, the normalization parameter Λ_k , that determines the total "energy" of the process and does not affect the temporal intensity shape, is not important when comparing the intensities in shape. Therefore, let us exclude Λ_k from the set of parameters describing the mixture (1). As a result, the characteristics of each precedent are primarily the N_k – the number of constituent (1) components and then the N_k -sets of component parameters $\{p_{jk}\}$, $\{\tau_{jk}\}$, $\{\sigma_{jk}\}$. Thus, assuming Λ_k in (1) is equal to unity, we arrive at a description of the precedents – description that is probability distributions of a special type – Gaussian mixtures.

Regarding the introduced precedents feature (parameter) space, we note the following. It is known that for a fixed total number n of Poisson process points, their time moments $\{t_1, \dots, t_n\}$ (without taking into account the

order) are distributed as identical independent random variables with the probability distribution $\lambda_k(t)$ (1) [12]. In this connection, the feature space can be interpreted as a space of parameters of Gaussian mixtures. This remark is not only of theoretical interest, but it also has an important practical significance. It is because within the framework of machine learning there are many effective algorithms that allow to find the maximum likelihood estimates of Gaussian mixtures parameters of the set of independent, identically distributed random points. This group of algorithms includes popular EM-like algorithms [11] that recursively refined parameters $\{p_{jk}\}$, $\{\tau_{jk}\}$, $\{\sigma_{jk}\}$ during M step.

Having determined the structure of the database of precedent descriptions and how it is formed from the implementation of discrete points based on EM-like algorithm, it is necessary to clarify the procedure of identification of the newly recorded process and the precedent from the database. Theoretically, it would be possible to assume that there are precedents in the database for all possible cases, so the description of the intensity of recorded Poisson process will coincide with the normalized version (1) of one of the precedents. Thus, if we form from the n discrete samples $\{t_1, \dots, t_n\}$ of the identified process (logarithmic) likelihood function for each (k -th) precedent:

$$L_k(\{t_i\}) = \ln \prod_{i=1}^n p(t_i | k) = \ln \prod_{i=1}^n \sum_{j=1}^{N_k} p_{jk} \frac{1}{\sqrt{2\pi\sigma_{jk}^2}} \exp\left(-\frac{(t_i - \tau_{jk})^2}{2\sigma_{jk}^2}\right) \quad (2)$$

than it would be possible to use the maximum likelihood (ML) principle to find the target precedent (K -th):

$$K = \arg \max_k L_k(\{t_i\}) \quad (3)$$

III. METHOD

Unfortunately, from the practical point of view, the above straightforward approach to identification is unlikely to be realized. Indeed, with this approach, for each shape of intensity (1) the shift in the origin of time, for example, will lead to a precedent different from the original one (with parameters $\{\tau_{jk}\}$ shifted to the constant). Considering this and a number of other reasons, we consider the identification of the recorded process and the target precedent from DB as a coincidence of their shapes up to affine time transformation (change in time origin on θ and change of time scale on s): $\lambda(t) \sim \lambda_k(st + \theta)$, $-\infty < \theta < \infty$, $s > 0$. The log-likelihood function in this case should be rewritten as:

$$L_k(\{t_i\}) = \ln \int \rho_{apr}(\theta, s) \prod_{i=1}^n p(t_i | \theta, s, k) d\theta ds, \quad (4)$$

$$p(t | \theta, s, k) = \sum_{j=1}^{N_k} p_{jk} \frac{s}{\sqrt{2\pi\sigma_{jk}^2}} \exp\left(-\frac{(st + \theta - \tau_{jk})^2}{2\sigma_{jk}^2}\right),$$

where $\rho_{apr}(\theta, s)$ is the a priori probability distribution of not related to the precedent's parameters θ and s .

In order to give the log-likelihood function (4) an appropriate form for the application of the EM algorithm,

in addition to the registration data $\{t_1, \dots, t_n\}$, we introduce the hidden variables $\{z_1, \dots, z_n\}$, where $z_i \in \{1, \dots, N_k\}$ – the indicator of the i -th point belonging to the z_i component of the mixture. From this point of view $p(t | \theta, s, k)$ (4) is the marginal distribution of t from the joint distribution $p(t, j | \theta, s, k)$ of the observed and hidden data. With this in mind, we rewrite (4) as follows:

$$L_k(\{t_i\}) = \ln \int \rho_{apr}(\theta, s) \sum_{z_1, \dots, z_n} \prod_{i=1}^n p(t_i, z_i | \theta, s, k) d\theta ds, \quad (5)$$

$$p(t, z | \theta, s, k) = \frac{s}{\sqrt{2\pi\sigma_{zk}^2}} \exp\left(-\frac{(st + \theta - \tau_{zk})^2}{2\sigma_{zk}^2}\right) p_{zk}.$$

The computational aspects associated with calculating $L_k(\{t_i\})$ (5) can be significantly complicated by the necessity of determination of sums for $\{z_i\}$ and integrals over θ, s . To avoid this complication, let us introduce the conditional distribution $\{z_i\}, \theta, s$ (for given $\{t_i\}$):

$$p(\{z_i\}, \theta, s | \{t_i\}, k) = \frac{\rho_{apr}(\theta, s) \prod_{i=1}^n p(t_i, z_i | \theta, s, k)}{\int \rho_{apr}(\theta, s) \sum_{z_1, \dots, z_n} \prod_{i=1}^n p(t_i, z_i | \theta, s, k) d\theta ds}, \quad (6)$$

substitute it into $L_k(\{t_i\})$ (5), average the resulting expression over arbitrary distributions $v(\{z_i\}), w(\theta, s)$ and divide the result into two terms $L_k(\{t_i\}) = F_{vw}(\{t_i\}) + K_{vw}(\{t_i\})$ of the form:

$$F_{vw}(\{t_i\}) = \int \sum_{z_1, \dots, z_n} v(\{z_i\}) w(\theta, s) \ln \left(\frac{\rho_{apr}(\theta, s) \prod_{i=1}^n p(t_i, z_i | \theta, s, k)}{v(\{z_i\}) w(\theta, s)} \right) d\theta ds, \quad (7)$$

$$K_{vw}(\{t_i\}) = - \int \sum_{z_1, \dots, z_n} v(\{z_i\}) w(\theta, s) \ln \left(\frac{p(\{z_i\}, \theta, s | \{t_i\}, k)}{v(\{z_i\}) w(\theta, s)} \right) d\theta ds.$$

The first term $F_{vw}(\{t_i\})$ in (7) is usually called the “free energy” and, with accuracy of averaging over $v(\{z_i\}), w(\theta, s)$ (and the “information” term $\langle \ln[v(\{z_i\}) w(\theta, s)] \rangle$), it is obtained from $L_k(\{t_i\})$ (5) by permuting the sum and integral with the operation of taking the logarithm. This, obviously, greatly simplifies the calculations. The second term in (7) is the Kullback-Leibler divergence of the distribution density $p(\{z_i\}, \theta, s | \{t_i\}, k)$ with respect to $v(\{z_i\}) w(\theta, s)$, which is always nonnegative and vanishes in the only case – if the first distribution coincides with the second. That is why $F_{vw}(\{t_i\})$ is always a bottom estimate of $L_k(\{t_i\})$, and this estimate is as better as the free energy is greater (or divergence is smaller). Based on the above analysis, we can formulate the following variational method of the likelihood function determination:

$$L_k(\{t_i\}) \approx \max_{v(\{z_i\}); w(\theta, s)} F_{vw}(\{t_i\}), \quad (8)$$

and equality is possible only in the case of factorization $p(\{z_i\}, s, k | \{t_i\}, k)$ with respect to the hidden variables $\{z_i\}$ and parameters θ, s .

Solving the variational problem for the functional $F_{vw}(\{t_i\})$ (7) in the usual way (using the Lagrange multiplier method), we obtain the following system for optimal solutions:

$$w(\theta, s) = \frac{1}{\Sigma_w} \exp \left\{ \sum_{z_1, \dots, z_n} v(\{z_i\}) \ln \left(\rho_{apr}(\theta, s) \prod_{i=1}^n p(t_i, z_i | \theta, s, k) \right) \right\}, \quad (9)$$

$$v(\{z_i\}) = \frac{1}{\Sigma_v} \exp \left\{ \int w(\theta, s) \ln \left(\rho_{apr}(\theta, s) \prod_{i=1}^n p(t_i, z_i | \theta, s, k) \right) d\theta ds \right\},$$

where Σ_w and Σ_v are the normalization constants (partition function). The system (9) can be simplified to:

$$w(\theta, s) = \frac{1}{\Sigma_w} \exp \left\{ \sum_{i=1}^n \sum_{z=1}^N V(t_i, z) \ln p(t_i, z | \theta, s, k) \right\} \rho_{apr}(\theta, s), \quad (10)$$

$$V(t, z) = \frac{1}{\Sigma_v} \exp \left\{ \int w(\theta, s) \ln p(t, z | \theta, s, k) d\theta ds \right\}, \quad v(\{z_i\}) = \prod_{i=1}^n V(t_i, z_i),$$

and takes the final explicit form when $p(t, z | \theta, s, k)$ is substituted into it from (5):

$$w(\theta, s) = \frac{1}{\Sigma_w} \exp \left\{ n \ln s - \sum_{i=1}^n \left\langle \frac{1}{2\sigma_{zk}^2} (st_i + \theta - \tau_{zk})^2 \right\rangle_{V(t_i, z)} \right\} \rho_{apr}(\theta, s), \quad (11)$$

$$V(t, z) = \frac{1}{\Sigma_v} \frac{p_{zk}}{\sqrt{2\pi\sigma_{zk}^2}} \exp \left\{ -\frac{1}{2\sigma_{zk}^2} \langle (st + \theta - \tau_{zk})^2 \rangle_{w(\theta, s)} \right\}, \quad v(\{z_i\}) = \prod_{i=1}^n V(t_i, z_i),$$

In a number of cases, with a suitable choice of $\rho_{apr}(\theta, s)$, system (11) can be reduced to (nonlinear) algebraic and for its solution well-known methods can be used. However, one can do otherwise: find a good (asymptotically) approximate solution using the fact that $w(\theta, s)$ in (11) has a narrow ($\sim 1/n$) maximum at θ_m, s_m , which can be found from the following equations:

$$\theta_m = \theta_k - s_m T_k, \quad \theta_k = \frac{\sum_{i=1}^n \sum_{z=1}^N \tau_{zk} V(t_i, z) / \sigma_{zk}^2}{\sum_{i=1}^n \sum_{z=1}^N V(t_i, z) / \sigma_{zk}^2}, \quad (12)$$

$$T_k = \frac{\sum_{i=1}^n \sum_{z=1}^N t_i V(t_i, z) / \sigma_{zk}^2}{\sum_{i=1}^n \sum_{z=1}^N V(t_i, z) / \sigma_{zk}^2}, \quad \frac{1}{\sigma_k^2} = \frac{1}{n} \sum_{i=1}^n \sum_{z=1}^N \frac{V(t_i, z)}{\sigma_{zk}^2},$$

$$s_m D_k^2 + [\theta_k T_k - \theta T_k] = \frac{\sigma_k^2}{s_m},$$

$$\theta T_k = \frac{\sum_{i=1}^n \sum_{z=1}^N t_i \tau_{zk} V(t_i, z) / \sigma_{zk}^2}{\sum_{i=1}^n \sum_{z=1}^N V(t_i, z) / \sigma_{zk}^2}, \quad D_k^2 = \frac{\sum_{i=1}^n \sum_{z=1}^N (t_i - T_k)^2 V(t_i, z) / \sigma_{zk}^2}{\sum_{i=1}^n \sum_{z=1}^N V(t_i, z) / \sigma_{zk}^2}.$$

The optimal parameters θ_m, s_m (12) are easily computed if conditional distributions of the indicators $V(t_i, z)$ (11) are given for all $\{t_i\}$. However, within the framework of the approximation in question $w(\theta, s) = \delta(\theta - \theta_m, s - s_m)$ these distributions can also be easily found:

$$V(t_i, z) = \frac{1}{\Sigma_{Vi}} \frac{p_{zk}}{\sqrt{2\pi\sigma_{zk}^2}} \exp \left\{ -\frac{1}{2\sigma_{zk}^2} (s_m t_i + \theta_m - \tau_{zk})^2 \right\}, \quad (13)$$

$$\Sigma_{Vi} = \sum_{z=1}^N \frac{p_{zk}}{\sqrt{2\pi\sigma_{zk}^2}} \exp \left\{ -\frac{1}{2\sigma_{zk}^2} (s_m t_i + \theta_m - \tau_{zk})^2 \right\}.$$

Combining (12) and (13) into a common recurrence scheme with a counter of iterations r , we obtain the structure of the algorithm for finding $v\{z_i\}$ и $w(\theta, s)$ very similar to the structure of the EM algorithm:

$$\begin{aligned}
 E: v^{(r+1)}(\{z_i\}) &= \prod_{i=1}^n V^{(r+1)}(t_i, z_i) : \\
 V^{(r+1)}(t_i, z) &= \frac{1}{\sum_{V_i} \sqrt{2\pi\sigma_{zk}^2}} \exp\left\{-\frac{1}{2\sigma_{zk}^2}(s_m^{(r)} t_i + \theta_m^{(r)} - \tau_{zk})^2\right\}, \quad (14) \\
 M: w^{(r+1)}(\theta, s) &= \delta(\theta - \theta_m^{(r+1)}, s - s_m^{(r+1)}) : \\
 s_m^{(r+1)} &= \frac{S_k}{2} + \sqrt{\left(\frac{S_k}{2}\right)^2 + \left(\frac{\sigma_k}{D_k}\right)^2}, \quad \theta_m^{(r+1)} = \theta_k - s_m^{(r+1)} T_k,
 \end{aligned}$$

where the abbreviation $S_k = (\Theta T_k - \theta_k T_k)/D_k^2$, which has the meaning of the scale factor in the case of intensities with weak overlapping of components ($\sigma_k/D_k < 1$), is used.

Substitution of the optimal $v^{(r+1)}(\{z_i\})$ и $w^{(r+1)}(\theta, s)$ (14) into the expression for the free energy $F_{vw}(\{t_i\})$ (7) after a series of transformations gives:

$$\begin{aligned}
 F_{vw}^{(r+1)}(\{t_i\}) &= \ln \rho_{apr}(\theta_m^{(r+1)}, s_m^{(r+1)}) + \sum_{i=1}^n \sum_{z=1}^N V^{(r+1)}(t_i, z) \ln \left(\frac{p(t_i, z | \theta_m^{(r+1)}, s_m^{(r+1)}, k)}{V^{(r+1)}(t_i, z)} \right) = \\
 &= \ln \left(\rho_{apr}(\theta_m^{(r+1)}, s_m^{(r+1)}) \prod_{i=1}^n \sum_{z=1}^N \right) = \ln \left(\rho_{apr}(\theta_m^{(r+1)}, s_m^{(r+1)}) \prod_{i=1}^n p(t_i | \theta_m^{(r+1)}, s_m^{(r+1)}, k) \right) \quad (15)
 \end{aligned}$$

It follows from (15) that to calculate $F_{vw}(\{t_i\})$, and ultimately the likelihood functions $L_k(\{t_i\})$ (8), only the parameters θ_m, s_m calculated at step M are needed. Of course, to calculate them, we need a series of auxiliary quantities $\theta_k, T_k, \dots, D_k$ (12), which in turn are determined by the parameters of the model $\{p_{jk}\}, \{\tau_{jk}\}, \{\sigma_{jk}\}$ and by the set of quantities $\{V(t_i, z)\}$ calculated at step E. However, calculations in the explicit form of $v^{(r+1)}(\{z_i\})$ and $w^{(r+1)}(\theta, s)$ are not required in (14), they play only a conceptual role. Thus, the algorithm obtained is a constructive solution to the problem posed.

IV. CONCLUSIONS

In this paper a complete solution of the problem of temporal dynamics of Poisson process intensity identification is presented. For the implementation of the solution, a computational scheme is proposed. This scheme is very close in structure to the popular in machine learning EM algorithm. Since many technical aspects of the computer implementation of EM algorithms are now well-developed [11], one can expect that the implementation of the proposed solution will be also quick and effective.

ACKNOWLEDGMENT

The authors wish to thank the Russian Foundation for Basic Research (RFBR). This work was supported by a RFBR grant N 18-07-01295.

REFERENCES

- [1] R. P. Adams, I. Murray, and D. J. C. MacKay. "Tractable nonparametric Bayesian inference in Poisson processes with Gaussian process intensities," in *Proc. 26th Annu. Inter. Conf. on Machine Learning*, Montreal, Quebec, Canada, 2009, pp. 9-16.
- [2] M. Gennaro, K. Tchernyshyov, T. M. Brown, and K. D. Gordon. "A new method for deriving the stellar birth function of resolved Stellar Populations," *The Astrophysical Journal*, vol. 808, no. 1, p. 20, 2015.
- [3] D. L. Borchers and R. Langrock, "Double-observer line transect surveys with Markov-modulated Poisson process models for overdispersed animal availability," *Biometrics*, vol. 71, no. 4, pp. 1060-1069, 2015.
- [4] D. I. Warton and L. C. Shepherd, "Poisson point process models solve the 'pseudo-absence problem' for presence-only data in ecology," *Ann. Appl. Stat.*, vol. 4, no. 3, pp. 1383-1402, 2010.
- [5] J. Zhuang, C. Chang, Y. Ogata, and Y. Chen, "A study on the background and clustering seismicity in the Taiwan region by using point process models," *J. of Geophysical Research*, vol. 110, no. B5, p. B05S18, 2005.
- [6] M. Haenggi, J. Andrews, et al., "Stochastic geometry and random graphs for the analysis and design of wireless networks," *IEEE JSAC*, vol. 27, no. 7, pp. 1029-1046, 2009.
- [7] D. R. Cox and P. A. W. Lewis, *The Statistical Analysis of Series of Events*, London: John Wiley & Sons, Inc., 1966.
- [8] D. R. Cox and V. Isham. *Point Processes*, London: Chapman & Hall, 1980.
- [9] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed., New York: Springer, 2009.
- [10] T. Mitchell, *Machine Learning*, New York: McGraw Hill, 1997.
- [11] G. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, New York: Wiley, 1977.
- [12] R. Streit, *Poisson Point Processes—Imaging, Tracking, and Sensing*, New York: Springer, 2010.
- [13] V. Antsiperov, "Precedent-based low count rate image intensity estimation using maximum likelihood distribution descriptions," in *Proc. of the Inter. Conf. on Pattern Recognition and Artificial Intelligence*, Montreal, May 13-17, 2018, pp. 707-711.



Viacheslav Antsiperov was born in 1959. Graduated from the Moscow Institute of Physics and Technology in 1982. He received Candidate's degree (Physics and Mathematics) in 1986. At present, He is leading researcher at the Kotelnikov Institute of Radioengineering and Electronics of the Russian Academy of Sciences. His scientific interests are information systems, processing and analysis of signals including image and speech recognition, biomedical informatics.



Aleksei Morozov was born in 1968. Graduated from Bauman Moscow State Technical University in 1991. He received his Candidate's degree in Physics and Mathematics in 1998 and Senior researcher at the Kotelnikov Institute of Radio Engineering and Electronics, Russian Academy of Sciences. His scientific interests are logic programming, intelligent video surveillance, and processing of biomedical signals (EEG, EMG, and MEG).