EXPECTATION-MAXIMIZATION AND THE PREDPOL ALGORITHM

LUC TA

Abstract. We begin by reviewing the expectation-maximization algorithm in the context of Hawkes processes. A particularly concerning application of this technique is the PredPol algorithm, one of the first predictive policing algorithms deployed in the United States. We contextualize and outline a derivation of the mathematical model used in PredPol. This body of this piece is primarily based on the work of Johnson, McKenzie, and Wong in Section 2.2 of [3]. The footnotes are heavily influenced by the scholarship of Benjamin in [1] and the Yale course *Philosophy of Data Science* taught by Lily Hu.

1. Mathematical Preliminaries

A *point process* ξ is a locally finite counting measure on the Borel σ-algebra of a locally compact second countable Hausdorff space.[1] We say ξ is a *Poisson process* if (i) $\xi(B)$ is Poisson distributed for any bounded subset B and (ii) $\xi(B_1), \ldots, \xi(B_n)$ are independent whenever $B_1, \ldots, B_n$ are disjoint. A **Hawkes process** is a self-exciting Poisson process. While these definitions may seem *ad hoc*, Hawkes processes appear naturally in statistical models of epidemiology, seismology, and mathematical finance.[2]

In mathematical statistics, it is common to estimate parameters using *maximum likelihood estimation*, which relies on the presence of unobserved *latent variables*. In the absence of latent variables, the next best method of estimation is the **expectation-maximization (EM) algorithm**. This algorithm iterates an expectation (E) step and a maximization (M) step until convergence. In the context of Hawkes processes, we call $\xi(t)$ the *triggering kernel*, a function that determines the elevated risk of events following each event. The authors of [4] showed that the E- and M-steps are as follows:

Theorem 1. *Given a Hawkes process with exponential triggering kernel* $\xi(t) = \theta\omega e^{-\omega t}$ *on the time interval* $[0, T]$ *and background rate* $\mu(t)$*, let* p_{ij} *be the probability that event* i *triggers event* j*. Then, starting with an initial estimate* $(\mu_0, \theta_0, \omega_0)$*, the EM algorithm for estimating the parameters*

[1] Mathematical rigor often requires that even the most intuitive concepts be shrouded in layers of jargon and technical definitions. In the act of making mathematical and political discourses inaccessible to all but a few researchers, what work is being done? What does this say about the mapping between the logics of mathematics and the logics of settler colonialism, Empire, and the carceral state?

[2] On this note, readers should pay close attention to the relationships that mathematical and scientific discourses have with the logics of capitalism.

(μ, θ, ω) *is given by the E-step*

$$p_{ij}^k = \frac{\theta^k \omega^k e^{-\omega^k (t_i - t_j)}}{\mu^k + \sum_{j=1}^{i-1} \theta^k \omega^k e^{-\omega^k (t_i - t_j)}},$$
$$p_{ii}^k = \frac{\mu^k}{\mu^k + \sum_{j=1}^{i-1} \theta^k \omega^k e^{-\omega^k (t_i - t_j)}}$$

and the M-step

$$\mu^{k+1} = \frac{\sum_{i=1}^{n} p_{ii}^k}{T},$$
$$\theta^{k+1} = \frac{\sum_{i<j} p_{ij}^k}{n},$$
$$\omega^{k+1} = \frac{\sum_{i<j} p_{ij}^k}{\sum_{i<j} (t_i - t_j) p_{ij}^k}.$$

2. The PredPol Algorithm

Predictive policing is the application of mathematical models by law enforcement to predict potential criminal activity.[3] One of the first of many predictive policing algorithms used in the United States is the **PredPol algorithm**, developed by PredPol, Inc. with the support of the Los Angeles Police Department.[4] In this section, we summarize the patented mathematics behind the PredPol algorithm originally derived in [5]. This technical exposition is based chiefly on recent scholarship-activism of Johnson, McKenzie, and Wong in [3]. For a more detailed and nuanced analysis of predictive policing in the United States and its political and social implications, we also direct the reader to the brilliant work of Benjamin in *Race After Technology* [1]. In this paper, we only outline the technical workings of the PredPol algorithm as a case study "from the inside."

[3]It should go without saying that predictive policing algorithms encode racial bias at all steps of conceptualization and implementation. The most obvious of these ways is basing models on racially biased crime data (which creates a positive feedback loop with racialized overpolicing), but this is far from the only way racism embeds into predictive policing—after all, the carceral system is *designed* in opposition to Blackness. For further reading, I highly recommend the scholarship of Benjamin in [1].

[4]Given the United States' transnational military and economic power as an ongoing project of settler colonialism, it should come as no surprise that the United States also researches, develops, and exports predictive policing algorithms and other carceral technologies abroad. In particular, the United States has its fingerprints on Red Wolf, Blue Wolf, and Wolf Pack, a system of facial recognition systems and databases containing personal data on Palestinians, all collected without their consent. The Israel Defense Forces use these technologies to arrest and detain Palestinians automatically. In 2023, Amnesty International found that Israeli forces "gamify" the application of these technologies by providing prizes to commanders who register the most Palestinians in these databases [2]. I will take this moment to remind the reader of the mathematical framework of *maximization* employed in the EM algorithm. How do the quantitatively motivated frameworks of maximization and optimization come into play here and in the applications of mathematics and carceral technologies more generally? What do they imply about the relationships between military technologies and capitalism?

Interestingly, the quantitative bases of many predictive policing algorithms originate in mathematical models in the physical sciences.[5] For example, the algorithm described in this section takes inspiration from reaction-diffusion models, which are typically used to study systems of chemical reactions.[6] Mathematically, the PredPol algorithm is based on **epidemic-type aftershock (ETAS) models**, which are typically used to predict the locations of earthquakes.[7]

To apply the EM algorithm, the PredPol developers model crime as a discretized Hawkes process interrupted by police.[8] Specifically, policing areas are taken as square boxes indexed by a finite set of natural numbers. Then, the developers define the *conditional intensity* (i.e., probabilistic rate) of events in box n at time t as

$$\lambda_n(t) := \mu_n + \sum_{t_n^i < t} \theta\omega e^{-\omega(t-t_n^i)},$$

where the t_n^i's are the times of events in box n in the history of the Hawkes process.[9] The background rate μ is a nonparametric histogram estimate of a stationary Poisson process.[10] As in Theorem 1, the triggering kernel is $\xi(t) = \theta\omega e^{-\omega t}$. The authors of [5] describe ξ in this context as modeling "near-repeat" or "contagion" effects in crime data.[11]

Letting T denote the time window of observation, Theorem 1 now yields the PredPol algorithm given a starting estimate $(\mu_0, \theta_0, \omega_0)$.[12] The algorithm

---

[5]What are the politics of modeling human behavior, particularly in the racialized contexts of policing and carcerality, as if they were inhuman, physical phenomena? Whose perspectives and what information gets lost in this depersonalizing process of mathematical formulation?

[6]Here, "motivated offenders," targets, and victims take the role of enzyme activators, while law enforcement plays the role of enzyme inhibitors [3]. What is lost in parameterizing the social construction of "crime" as something caused by "offenders" and inhibited by the police? Whom does such a simplistic causal model serve, and whom does it hurt?

[7]Once again, what are the carceral logics that motivate such a mapping between natural phenomena like chemical reactions and earthquakes and the social construction of "crime"?

[8]Cf. footnotes 5, 6, and 7.

[9]In other words, $\lambda_n(t)$ is the expected rate at which events are expected to occur around time t given the history of box n at times prior to t. If this expected rate is itself something the developers at PredPol, Inc. define—and immortalize in their police-backed quantitative research—then what does that say about the algorithms, discourses, and logics that rely on these invented quantities? More generally, who gets to determine how we define "crime," "criminality," and "innocence"? Which organizations, systems, and discourses sustain these constructions? (In the specific case of PredPol, one of these systems seems to be the purportedly apolitical field of mathematics.) What does this say about the carceral logics that sustain and are sustained by these definitions? How can understanding these positive feedback loops of definition and implementation help us disrupt the power structures they posit as common sense?

[10]Cf. footnote 1.

[11]Cf. footnotes 5, 6, and 7. What work does the developers' language of "contagion" do?

[12]What is the purpose of defining the PredPol algorithm in such opaque, mathematical terms, without any attention to the social or the qualitative or even any simplified summary for non-mathematicians to read? Whom does this work serve? Cf. footnotes 1, 5, and 9.

is as follows: until convergence, alternate between the E-step

$$p_{ij}^k = \frac{\theta^k \omega^k e^{-\omega^k (t_i - t_j)}}{\lambda_k(t_j)},$$

$$p_{ii}^k = \frac{\mu^k}{\lambda_k(t_j)}$$

and the M-step[13]

$$\mu = \frac{\sum_k \sum_i p_{ii}^k}{T},$$

$$\theta = \frac{\sum_k \sum_{i<j} p_{ij}^k}{\sum_k \sum_j 1},$$

$$\omega = \frac{\sum_k \sum_{i<j} p_{ij}^k}{\sum_k \sum_{i<j} (t_i - t_j) p_{ij}^k}.$$

References

[1] Ruha Benjamin, *Race after technology*, Polity Press, Oxford, England, 2019 (en).

[2] Amnesty International, *Israeli authorities are using facial recognition technology to entrench apartheid*, 2023.

[3] Joseph Johnson, Theo McKenzie, and Tian An Wong, *Predictive policing: a mathematical primer*, Notices Amer. Math. Soc. **71** (2024), no. 7, 929–937. MR4776110

[4] Erik Lewis and George Mohler, *A nonparametric EM algorithm for multiscale Hawkes processes*, Journal of Nonparametric Statistics (2011).

[5] G. O. Mohler, M. B. Short, Sean Malinowski, Mark Johnson, G. E. Tita, Andrea L. Bertozzi, and P. J. Brantingham, *Randomized controlled field trials of predictive policing*, J. Amer. Statist. Assoc. **110** (2015), no. 512, 1399–1411. MR3449035

[13] In mathematics, it is customary to end research articles "abruptly," i.e., without a "conclusion" section or any kind of reflection upon the results presented in the article. What is the purpose of evading this kind of critical self-reflection among mathematicians? Who benefits from standardizing this cut-and-dried approach?