Predictive Policing: A Mathematical Primer

Joseph Johnson, Theo McKenzie, and Tian An Wong

The opinions expressed here are not necessarily those of the Notices *or the AMS.*

1. Introduction

What is predictive policing? The RAND Corporation, for example, defines predictive policing as "the application of analytical techniques—particularly quantitative techniques—to identify likely targets for police intervention and prevent crime or solve past crimes by making statistical predictions." This brief survey serves as a mathematical supplement to recent discussions of predictive policing such in [20] and various investigative reports, and perhaps most importantly the boycott by several thousand mathematicians of collaboration with the police, appearing in the *Notices of the AMS* [3].[1] The foundations of predictive policing are highly mathematical, while modern usages also heavily incorporate machine learning to optimize models.

There are many surveys and explanations of predictive policing from various points of view, such as surveillance technology, algorithmic fairness, and criminal justice, but few specifically address its mathematical aspects. The purpose of this article is to present a selective overview of the mathematics underlying the model used by the company PredPol, followed by critiques and further developments of the model. We focus on PredPol as it is a well-known method of predictive policing and the methodology is arguably less oblique than other predictive policing algorithms that rely more heavily on black-boxed machine learning algorithms.

PredPol is often singled out in discussions surrounding predictive policing, in part because it is arguably the progenitor of the field, although it is far from the only player in the game. According to investigative reports, PredPol was the most widely used predictive policing company/software in the US as of 2019, having contracts in states including Utah, Washington, and California (including the University of California, Berkeley).The company was born out of a collaboration with the LAPD, the FBI, and UCLA in the early 2010s. In 2023, PredPol, which was rebranded as Geolitica, was effectively bought by SoundThinking, previously known as ShotSpotter. The same company also acquired HunchLab in 2018, another predictive policing software company. This growing consolidation shows that predictive policing remains a relevant method in modern policing, and deserves the attention of the mathematical community. Indeed, a recent Wired report called the acquisition of Geolitica its "latest step in becoming the Google of crime fighting."

2. Summary of the Basic PredPol Model

We begin by discussing two early models: the epidemic type aftershock (ETAS) model that the PredPol patent is built on (US Patent No. 8,949,164), and the reaction-diffusion model that ETAS in turn draws upon. Reaction-diffusion is modeled using partial differential equations, with original applications to the physical sciences. The

Joseph Johnson is an assistant professor of mathematics at Carleton College. His email address is jjohnson16@carleton.edu.

Theo McKenzie received his PhD in mathematics from Stanford University. His email address is theom@stanford.edu.

Tian An Wong is an assistant professor of mathematics at the University of Michigan-Dearborn. Their email address is tiananw@umich.edu.

McKenzie was partially supported by NSF Grant DMS-2212881. Wong was partially supported by NSF Grant DMS-2212924.

Communicated by Notices *Associate Editor Reza Malek-Madani.*

For permission to reprint this article, please contact: reprint-permission@ams.org.

DOI: https://doi.org/10.1090/noti2968

[1]*Disclosure: Two of the authors (Johnson and Wong) have signed this open letter.*

ETAS model, on the other hand, is a popular statistical model for earthquake occurrence. PredPol is thus based on equations that are used to model earthquake occurrences and chemical reactions which produce certain "hotspots," and where adding "hotspot policing" can be viewed as an inhibitor to the process. From a statistical point of view, the introduction of police to given hotspots is then viewed as applying "treatments" or interventions which are meant to reduce crime. Such models are also referred to as spatiotemporal models.

2.1. The reaction-diffusion model. We review the mathematics underlying [21, 22]. Reaction-diffusion models are typically used to describe chemical reactions, in which activators and inhibitors move, mix, and interact. In [22], the model describes houses and burglars, while [21] describes "motivated offenders," targets or victims as activators, and law enforcement as inhibitors. Their reaction-diffusion system involves "mobile criminal offenders" within a square environment with periodic boundary conditions. (See Section 4.1 for a discussion on the biases implicit in the language of crime.)

2.1.1. *Discrete model.* In the discrete model, houses are placed on a lattice in the plane with constant spacing ℓ. Within the plane $s = (x, y)$, the score in question $A_s(t)$ is interpreted either as the attractiveness of a house to a burglar, or the *risk of victimization*, "representing general environmental cues about the feasibility of committing a successful crime and/or specific knowledge offenders possess about target or victim vulnerability in the area." The risk is given by

$$A_s(t) = A_s^0 + B_s(t),$$

where A_s^0 is a fixed value and $B_s(t)$ is a dynamic value that models the idea that if a site has been attacked, it has a higher risk of being revictimized shortly after the first incident. The first approximation is

$$B_s(t+\delta t) = B_s(t)(1-\omega\delta t) + \theta E_s(t), \tag{1}$$

where ω sets a time scale over which repeat victimizations are most likely to occur, and θ is a multiplier of $E_s(t)$, the number of burglary events that occurred at site s since time t. The authors then modify this model to account for "near-repeat victimization, and the broken windows effect," by allowing the quantity $B_s(t)$ to spread spatially to its neighbors. Equation (1) is replaced with

$$B_s(t+\delta t) = \left[B_s(t) + \frac{\eta\ell^2}{z}\Delta B_s(t)\right](1-\omega\delta t) + \theta E_s(t),$$

where Δ is the discrete Laplacian, whereby

$$\Delta B_s(t) = \frac{1}{\ell^2}\left(\sum_{s'\sim s} B_{s'}(t) - zB_s(t)\right).$$

Here z is the number of sites s' neighboring s, and $0 \le \eta \le 1$ measures the significance of neighborhood effects. Computer simulations are then run to show that the model produces certain dynamic and stationary "hotspots."

The "criminal agent" is modeled as either committing a crime at site s or moving to a neighboring location based on a biased random walk so that site s^* is visited with probability

$$q_{s\to s^*}(t) = \frac{A_{s^*}(t)}{\sum_{s'\sim s} A_{s'}(t)}.$$

The probability of occurrence for each burglar located at site s between times t and $t+\delta t$ given by $p_s(t) = 1 - e^{-A_s(t)\delta t}$ in accordance with a standard Poisson process in which the expected number of events during the time interval of length δt is $A_s(t)\delta t$. In the discrete model, burglars are removed after committing a crime, and regenerated at each lattice site at a constant rate Γ. We write $E_s(t) = n_s(t)p_s(t)$, where $n_s(t)$ is the number of criminals at the site s at time t.

2.1.2. *Continuous model.* From the discrete model we form the difference quotient,

$$\frac{B_s(t+\delta t) - B(t)}{\delta t}$$

and take the limit as ℓ and δt approach 0 to arrive at the differential equation[2]

$$\frac{\partial B}{\partial t} = \frac{\eta D}{z}\nabla^2 B - \omega B + \epsilon D\rho A. \tag{2}$$

Here we have denoted

$$D = \frac{\ell^2}{\delta t}, \quad \epsilon = \theta\delta t, \quad \rho(s,t) = \frac{n_s(t)}{\ell^2},$$

where ρ is the density of criminal agents, and

$$\frac{\partial\rho}{\partial t} = \frac{D}{z}\nabla\cdot\left[\nabla\rho - \frac{2\rho}{A}\nabla A\right] - \rho A + \gamma, \tag{3}$$

where offenders exit the system at the rate ρA and are reintroduced at the constant rate $\gamma = \Gamma/\ell^2$. The PDE for ρ is obtained by a difference quotient for $n_s(t)$, using the equation

$$n_s(t+\delta t) = A_s \sum_{s'\sim s} \frac{n_{s'}(t)(1-p_{s'}(t))}{T_{s'}(t)} + \Gamma\delta t,$$

where

$$T_{s'}(t) = \sum_{s''\sim s'} A_{s''}(t),$$

which simply means that any agents that are present at s after one time step must have either arrived from a neighboring site after having not committed a crime there, or have been generated at s at rate Γ. The coupled differential equations (2) and (3) thus describe the continuous model.

[2] *This is essentially the patented algorithm displayed in* `https://www.predpol.com/technology/`.

In [21] the authors study these coupled PDEs to show that crime risk will form dense, well-spaced hotspots when the diffusion of risk by individual crimes is spatially broad enough. Police suppression is modeled by instantaneously setting the crime rate $\rho A_s(t) = 0$ at the locations of current crime hotspots and maintaining this suppression for a fixed time period. The authors then claim that subcritical crime hotspots may be permanently eradicated with police suppression.

2.2. Epidemic-type aftershock (ETAS). This section covers the mathematics underlying [14], which is the basis for the PredPol patent (US Patent No. 8,949,164). The authors treat the dynamic occurrence of crime as a continuous time, discrete space epidemic-type aftershock sequence point process. In seismology, point processes are used by considering a "parent earthquake" and subsequent background events or aftershocks. The ETAS model estimates long-term and short-term hotspots and systematically estimates the relative contribution to risk of each via an expectation-maximization (EM) algorithm.

The ETAS model can be intuitively understood as a branching process: first-generation events occur according to a Poisson process with constant rate μ, then events (from all generations) each give birth to N direct offspring events, where N is a Poisson random variable with parameter θ. As events occur, the rate of crime increases locally in space, leading to a contagious sequence of "aftershock" crimes that eventually dies out on its own or is interrupted by police intervention.

In this model, policing areas are discretized into square boxes. The probabilistic rate of events in box n at time t is defined to be

$$\lambda_n(t) = \mu_n + \sum_{t_n^i < t} \theta\omega e^{-\omega(t - t_n^i)}, \tag{4}$$

where t_n^i are the times of events in box n in the history of the process. The background rate μ is a (nonparametric histogram) estimate of a stationary Poisson process.

The expectation, or E-step in the EM algorithm, sets

$$p_n^{ij} = \frac{\theta\omega e^{-\omega(t_n^i - t_n^j)}}{\lambda_n(t_n^j)}, \qquad p_n^j = \frac{\mu_n}{\lambda_n(t_n^j)},$$

where $\theta\omega e^{-\omega t}$ is called the triggering kernel that models "near-repeat" or "contagion" effects in crime data. Here p_n^{ij} is the probability that event j is the offspring of event i and p_n^j is the probability that event j was generated by the Poisson process.

The maximization, or M-step, sets

$$\omega = \frac{\sum_n \sum_{i<j} p_n^{ij}}{\sum_n \sum_{i<j} p_n^{ij}(t_n^j - t_n^i)}, \tag{5}$$

$$\theta = \frac{\sum_n \sum_{i<j} p^{ij}}{\sum_n \sum_j 1}, \ \mu = \frac{\sum_n \sum_j p_n^j}{T}, \tag{6}$$

where T is the length of the time window of observation.

2.3. Field trials and case studies. The authors of the model undergirding the PredPol patent conducted a randomized controlled field trial to test the effectiveness of their model [14]. In their experiment, hotspots were generated daily by the ETAS model and a crime analyst. These hotspots were randomly assigned to foot patrols who then decided independently how to patrol the area as long as they remained within the prescribed area. The study was split across three LAPD divisions where each division studied effects for approximately six to eight months.

The authors compared how well their model predicted crime events against the predictions of crime analysts using standard methodologies of generating hotspot maps. Their ETAS model generated hotspots that successfully predicted crimes at a higher rate—ranging from 1.4 to 2.2 times better—than the hotspots generated by a crime analyst. They estimated a 7.4% decrease in crime (4.3 fewer crimes reported per week) at mean patrol levels when hotspots from their ETAS model were used compared to no patrol at all. When police used hotspots created by the crime analysts there was only a 3.5% decrease (2.0 fewer crimes per week) compared to no patrol at all. This corresponds to a decrease in crime that is 2.1 times larger in magnitude when comparing standard practice to the ETAS model, in line with the 1.4–2.2 times improvement in the prediction rate when comparing the analyst and the ETAS model.

Given this information, researchers asked whether this algorithm was susceptible to bias. Brantingham and Mohler, in collaboration with Valasik, expanded upon their previous work and analyzed the results of their field trial to see if the use of predictive policing results in arrest bias [4]. They compared the arrest data of police on patrol in algorithm-based hotspots to the arrest data of police on patrol in hotspots allocated by the crime analysts. After performing Cochran-Mantel-Haenszel and Woolf tests on their data, the authors found that the difference in arrest percentage with respect to ethnic group (i.e., Black, Latino, White) between the crime analyst hotspots and the ETAS generated hotspots was not statistically significant.

Although differences were not deemed statistically significant, they are worthy of further examination. There was a multiplicative increase of 1.8 in arrests in Foothill LAPD division on days where algorithms were used, but a 4.5

increase in arrests for Blacks compared to a 1.6 and 1.9 multiplicative increase in arrests for Latinos and Whites respectively. There was a 2.0 multiplicative increase in arrests on algorithm days in the Southwest division, but no increase in arrests for Whites compared to a 2.2 and a 1.8 increase in arrests for Blacks and Latinos respectively. The population distributions in the regions studied could have led to the statistical tests considering the differences observed in arrest rates as insignificant. For example, there were few arrests of Blacks in Foothill and few arrests of Whites in Southwest.

Furthermore, Brantingham, Mohler, and Valasik do not check if the arrest data is already biased. This would lead to their predictive policing model replicating the existing bias in policing and not removing the bias.

There have been many instances where the police were found to implement practices that would bias arrest data. For example, former New York City detective Stephen Anderson testified that he along with other police officers planted drugs on innocent civilians in order to boost their arrest numbers [13]. Sometimes even basic recording errors lead to a significant change in crime statistics, for example, a *Los Angeles Times* investigation found that from 2005 to 2012 the LAPD incorrectly classified 14,000 serious assaults as minor offenses [17].

It has been demonstrated that several predictive policing instruments (including PredPol) have been implemented in several jurisdictions where and when the police in that jurisdiction were either under government investigation into illegal policing practices or agreed to a federal settlement in response to illegal policing practices [19]. For example, police in Maricopa County have been found to conduct racially biased stops, searches, and arrests from 2007 to 2011 and from 2014 to 2017.

Of course, alterations to the algorithm have been theorized. Mohler included a fairness condition in order to control for bias due to group affiliation. An example of the impact of such a condition is given by Akpinar, De-Arteaga, and Chouldechova [1]. In their work they attempt to quantify and test the influence of crime reporting rates on predictive policing. A natural bias of PredPol is that it can only create hotspots based on crimes that are reported. Crime reporting rates are known to depend on factors such as age, gender, race, and socioeconomic class. Therefore the authors attempted to see how crime reporting rates would affect PredPol hotspots in Bogotá, where district-wide crime reporting rates are given.

By running various hotspot-creating models, the authors find that districts that report crime at higher rates will receive a number of hotspots disproportionately high compared to the amount of crime in the district, and districts that report crime at lower rates receive a disproportionately low number of hotspots.

Moreover, considering Bogotá records crime reporting rates on a district-wide level, normalizing by crime reporting rates creates skewed data within districts, as all cells in a given district are normalized by the same factor rather than on a cell-wide basis. Therefore, they hypothesize that the only way to remove such bias is if we had cell-specific crime-reporting data.

3. Mathematical Critiques of Predictive Policing

There have been critiques of the use of predictive policing both from a social and a mathematical perspective. We focus on the mathematical critiques first, then turn to the broader social critiques, as academics frequently miss the social and political consequences of the topics studied.

3.1. Statistical bias using a synthetic population. The most notable quantitative study of the PredPol system is the study of Lum and Isaac [12], which simulates a synthetic population in Oakland, CA, based upon census data and applies a model based on data from the 2011 National Survey on Drug Use and Health (NSDUH) in order to predict an individual's probability of drug use within the past month based on their demographic characteristics. The resulting data set acts as a replacement for the "ground truth" of drug crime use data, giving estimates of illicit drug use from a noncriminal justice, population-based data source. Compared with police records, the authors find that drug crimes known to police are not a representative sample of all drug crimes.

Applying their reconstruction of the PredPol algorithm as outlined above, the authors conclude that rather than correcting for the apparent biases in the police data, the model reinforces these biases, suggesting that predictive policing of drug crimes results in increasingly disproportionate policing of historically overpoliced communities. In particular, this is despite PredPol's claim that they use "only three data points in making predictions: past type of crime, place of crime and time of crime. It uses no personal information about individuals or groups of individuals, eliminating any personal liberties and profiling concerns."

3.2. Runaway feedback loops via a generalized Pólya urn model. These concerns can be given a mathematical framework. Ensign et al. [6] used a generalized Pólya urn model to model a predictive policing algorithm. In their model, they assumed that police patrolled two areas A and B at a rate based on their prior beliefs—this is represented by initial ratio of balls in the urn. When the crime rates in the two areas were equal, the rate at which police were sent to a specific area does not converge to the actual rate of crime, rather the probability that police were sent to an area was a beta distribution dependent on the initial beliefs of the police. This means the police do not actually learn the crime rates are the same in each area. When the crime rates were not equal, the probability that the police visited the area with the higher crime rate asymptotically approached one (hence, the police will eventually ignore the area with a lower crime rate).

This is a general problem of what is called traditional batch-mode machine learning, and the theory of urns is a common framework in reinforcement learning. In the generalized Pólya urn model, an urn contains balls of two colors, say red and black. At each time step, a ball is drawn, and based on its color a number of balls are replaced. If red, we add a red and b black balls; and if black we add c red and d black balls. This is represented by the replacement matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, where the standard case is when $a = d = 1$ and $c = b = 0$. As a toy model for predictive policing, A and B are two policing blocks, and the goal is to distribute police officers according to the proportion of crime in each area. Let d_A be the rate at which police in A discover crimes, r_A the rate at which crimes are reported in A, and w_d, w_r the respective weights such that $w_d + w_r = 1$ and $w_d d_A + w_r r_A$ represents the total rate of incident data from A.

For the rest of the section we make the following assumptions:

1. (Predictive model) The officer decides where to go next with probability based on current statistics. This means that the model uses some form of statistical information to make predictions on crime.
2. (Context) The only information retained about a crime is a count.
3. (Truth in crime data) If an officer goes to a location A with an underlying ground truth crime rate of λ_A, the officer discovers crime at a rate of $d_A = \lambda_A$. Reported incidents are also reported at a rate that matches the underlying ground truth crime rate, i.e., $r_A = \lambda_A$.
4. (Discovery only) Incident data is only collected by an officer's presence in a neighborhood, i.e., $w_d = 1$ and $w_r = 0$.

3.2.1. *Uniform crime.* First assume that crime rate is uniform, so $\lambda = \lambda_A = \lambda_B$. In this case, sending police to area A or B is given by drawing a red or black ball. We then sample a Bernoulli λ distribution. If 1, we simulate one step of the standard Pólya urn, and if 0, we simply replace the ball that was drawn. In this case, the probability of drawing a red ball has a limiting distribution equal to the beta distribution with parameters $(\alpha, \beta) = (n_A, n_B)$ where (n_A, n_B) are the number of red balls and black balls initially in the urn (c.f. [18]). Recall that the beta distribution is the distribution on $[0, 1]$ given by probability distribution function

$$\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} x^{\alpha-1}(1-x)^{\beta-1},$$

where $\Gamma(x)$ is the usual gamma function. This means that as the number of draws increases, the probability of visiting A or B does not necessarily converge to $1/2$. Rather, this probability remains dependent on the initial data.

3.2.2. *Nonuniform crime.* Now we drop the uniformity assumption. The urn is now modelled by the stochastic addition matrix $\begin{pmatrix} X_A & 0 \\ 0 & X_B \end{pmatrix}$, where X_A, X_B are Bernoulli variables with parameter λ_A, λ_B, respectively. Let $n_A^{(t)}, n_B^{(t)}$ be the number of red and black balls respectively at time t. The probability of adding any ball to the urn is given by

$$P(\text{adding a ball}) = \frac{n_A^{(t)}\lambda_A + n_B^{(t)}\lambda_B}{n_A^{(t)} + n_B^{(t)}},$$

while the probability of adding a red ball conditioned on adding any ball, is given by

$$\frac{P(\text{adding a red ball})}{P(\text{adding a ball})} = \frac{n_A^{(t)}\lambda_A}{n_A^{(t)}\lambda_A + n_B^{(t)}\lambda_B}, \tag{7}$$

and similarly for a black ball. In particular, this is the same as the deterministic Pólya urn in which an i-colored ball is sampled, replaced, and then λ_i more balls of the same color are added. Thus the stochastic matrix reduces to $\begin{pmatrix} \lambda_A & 0 \\ 0 & \lambda_B \end{pmatrix}$. The authors then deduce a runaway feedback loop for this toy model.

Proposition 3.1 ([6, Lemma 4]). *The asymptotic probability of sampling a red ball is* 1 *if* $\lambda_A > \lambda_B$ *and* 0 *if* $\lambda_A < \lambda_B$.

This tells us that as long as A has a ground truth crime rate that is even slightly higher than that of B, the update process will lead to police being eventually completely sent to A.

3.2.3. *Incorporating feedback.* In order to learn the crime rate, the Pólya urn should contain balls in proportion to the relative probability of crime occurrence. The following update rule guarantees that the urn proportion will converge to the ratio of replacement (i.e., crime) rates. Consider the probabilities λ_A and λ_B now conditioned on a ball of the respective color having been sampled. This makes the probability of adding a red ball equal to

$$\frac{n_A^{(t)}\lambda_A}{n_A^{(t)} + n_B^{(t)}}$$

rather than λ, and the expected fraction of red balls being added to the urn after one step of the process equal to (7) instead of $\lambda_A/(\lambda_A + \lambda_B)$.

We introduce the following change: instead of always adding the new balls, we first sample another ball from the urn, and only add the new balls if the colors are different.

This makes the probability of adding a red or black ball

$$\frac{n_A^{(t)}\lambda_A}{n_A^{(t)}+n_B^{(t)}}\frac{n_B^{(t)}}{n_A^{(t)}+n_B^{(t)}} \quad \text{or} \quad \frac{n_B^{(t)}\lambda_B}{n_A^{(t)}+n_B^{(t)}}\frac{n_A^{(t)}}{n_A^{(t)}+n_B^{(t)}}$$

respectively, where we see that the probabilities are proportional to λ_A, λ_B up to the common factor $n_A^{(t)}n_B^{(t)}/(n_A^{(t)}+n_B^{(t)})^2$. This is an example of rejection sampling, where sampled values are dropped according to some probability scheme to affect the statistic collected. Another related scheme is importance sampling, where balls are weighted inversely proportional to the rate at which police are sent.

3.2.4. *Reported incidents.* It is also possible to remove Assumption (4), so that w_d and w_r can take on different values. That is, we allow both discovered and reported incidents to be used as input to the urn model, as is more typically the case in predictive policing systems. In this case, the total weight of incidents from A would be $w_d d_A + w_r r_A$ if A were visited and $w_r r_A$ otherwise. This yields the following urn replacement matrix

$$\begin{pmatrix} w_d d_A + w_r r_A & w_r r_B \\ w_r r_A & w_d d_B + w_r r_B \end{pmatrix}.$$

It is then possible to study the limiting fraction of balls in the urn and account for feedback again, which we refer to [6, §3.4–3.5] for details.

Crucially, the study [6] retains Assumption (3), namely, that reported and discovered incident rates track the true crime rates, an assumption that is empirically false in general.

Remark 3.2 (Application to proactive policing). In recent work it was shown by Kinsman and Wong [9] that the above framework of [6] is flexible enough such that the runaway feedback loops can be shown to occur for proactive policing in general — not just predictive policing reliant on algorithms or mathematical models. That is, as long as policing resources are allocated to detect crime according to historical crime incident data for a collection of regions, it is likely that feedback loops will occur. See Section 3.4 also for a broader discussion.

3.3. A fairness penalty for PredPol. In response to criticism of PredPol, Mohler et al. introduced a fairness penalty for the original model. As a variation on [14], following [15], we can adjust our model by now setting $\mu = \exp(a \cdot z_n)$ where a is a vector of coefficients, and z_n is the vector of populations of different groups in grid space n. The parameters of λ_n in (4) can be estimated by maximizing the log-likelihood function

$$L(a,\omega,\theta) = \sum_{i=1}^{N} \log(\lambda_{n_i}(t_i)) - \sum_{n\in G}\int_0^T \lambda_n(t)dt \qquad (8)$$

where n_i is the grid cell of event i and the integral is taken over the window of observation $[0, T]$. The Hawkes intensity process λ_n is used in practice by ranking all grid cells n at a given time t and then directing police patrols to the top k grid cells, called hotspots. We will denote the set of grid cells comprising the top k hotspots at time t as K_t. We will refer to λ_n as neutral if it is estimated by maximizing (8).

Let $m = 1, \ldots, M$ be an indexing set for the number of subgroups (e.g., racial groups) surveyed, and let z_n^m be the population count of racial group m in grid cell n. Assuming that each of the grid cells in K_t receives the same number of patrols, then the number of patrols a particular racial group receives, per individual of that group per day, is

$$p_m = \frac{\sum_{t=1}^T \sum_{n\in K_t} z_n^m}{T\sum_{n\in G} z_n^m}.$$

The time interval over which hotspots are defined can be taken to be days, so that t is a discrete sum over days. We then define a measure of fairness by comparing the patrol statistics p_m between pairs of groups

$$F(a,\omega,\theta) = \sum_{m>m'} (p_m - p_{m'})^2,$$

so that when $F = 0$, each group m receives the same number of patrols per individual.

We then add F to the log-likelihood (8) as a fairness penalty and maximize with respect to a, ω, θ the loss function $L(a,\omega,\theta) - \chi F$, where $\chi \in \mathbb{R}$ is a penalty parameter that controls the balance between accuracy and fairness in the point process model. Note that this loss function is not everywhere differentiable due to potential changes in the k hotspots K_t.

3.4. Impact of machine learning. One point not yet addressed is the influence of machine learning in modern predictive policing algorithms. For example, Fitzpatrick, Gorr, and Williamson compare the ability to predict crime in Pittsburgh of a perceptron-based model versus a nonperceptron predictive policing model, and show that perceptron-based models lead to a greater entropy in hotspot locations [8]. Nevertheless, these models still have 43.6% of all hotspots remain hotspots for at least 75% of the study period, and these seem to be less predictive than other, nonperceptron models.

In this study, researchers found a significant decrease in serious violent crime at both temporary and chronic hotspots with predictive policing. Moreover, rather than crime displacement, they found a weak spillover effect of crime prevention in adjacent areas. However, this study predicated on the adoption of nonaggressive tactics by police, and arrests for commonly overpoliced crimes did not increase. Moreover, these conclusions are based on a relatively small scale, with a reduction of 24 crimes over a 12 month period. It should also be noted that any results will be specific to the social context of the city and police department implementing policy.

The problem of guaranteeing fairness of algorithms is a highly active research area. See [2] for a survey of the implications of fairness in machine learning on predictive policing.

To give an idea of the fundamental algorithmic issue of fairness, we consider the work of Kleinberg, Mullainathan, and Raghavan [10]. Within a population, each individual is assigned a feature vector σ. Our goal is to classify individuals into positive and negative classes; therefore define p_σ to be the fraction of people with classification σ that are positive. To each population with a given set of characters σ, we assign a distribution of scores X_σ, so we say $X_{\sigma b}$ is the fraction of people in p_σ assigned to bin b, with associated score v_b.

The authors propose three goals of a fair algorithm between two groups, group 1 and group 2.

1. For each group and bin b, the fraction of people assigned to the positive class should be v_b. Namely, within each bin, the group should not affect classification.
2. The average score v_b of people in the positive class should not depend on group.
3. The average score of people in positive class should not change per group.

The authors show that any classification that satisfies all three of these conditions is in some way trivial.

Theorem 3.3. *Any assignment satisfying conditions (1), (2), (3), has either*

1. *Perfect prediction: for each feature vector σ, $p_\sigma \in \{0, 1\}$.*
2. *Equal base rates: The average does not depend on the group.*

Therefore for any nontrivial scenario, the classification could be classified as biased. Moreover, this theorem is *robust* in the sense that if approximate versions of the three fairness criteria are satisfied, then the selection must be approximately close to perfect prediction or equal base rates. Therefore,"any assignment of risk scores can in principle be subject to natural criticism on the grounds of bias." To give an idea of the proof, we follow the authors' proof sketch.

Sketch of proof. By assumption (1), for any fixed bin b, the total score given to those in group t in b is equal to the number of people in group t in b in the positive class. Summing over all bins gives that the sum of scores of people in group t is equal to the number of people in the positive class in group t. Call this number μ_t.

Define x to be the average score of someone in the negative class, and y the average score of someone in the positive class. By assumption (2) and (3), x and y do not depend on the group. Because the total score is μ_t, if there are N total individuals, then we must have

$$(N - \mu_t)x + \mu_t y = \mu_t.$$

This is a linear equation in x, y that always has a solution at $(0, 1)$. This corresponds to perfect prediction. If there is another solution, then μ_t cannot depend on t, and we must have equal base rates, and x, y can be arbitrary. □

This gives a strong suggestion that any nontrivial assignment of people to individual scores must be, in some sense, unfair. Liu et al. gave another example of this, considering a theoretical model based on banks assigning credit to potential loan applicants (such a model also is relevant to, for example, college admissions) [11]. They assign average score μ_j to group j and check the change in the average score $\Delta(\mu_j)$ under three different scenarios, and check whether group scores increase or decrease.

1. Maximum Utility: The bank makes no attempt at fairness and instead only tries to maximize the utility of giving out loans to people who would give them back and limiting giving loans that will not be repaid.
2. Democratic Parity: The bank attempts to optimize utility under the condition that the same fraction of loans are accepted from group A and group B.
3. Equal Opportunity: The bank optimizes utility under the condition that across groups, the probability of success if selected does not depend on the group.

The authors show that in their model, there is no guarantee that adding a fairness condition will help a disadvantaged group. A summary of their various theorems is as follows.

Theorem 3.4. *There are population proportions for which for an underrepresented group j, $\Delta(\mu_j^{DemParity}) > \Delta(\mu_j^{MaxUtil})$ and $\Delta(\mu_j^{EqOpt}) > \Delta(\mu_j^{MaxUtil})$.*

Therefore these fairness criteria promote the typical score associated with group j at a faster rate than an algorithm that is not conditioned for fairness. However, the opposite is also possible.

Theorem 3.5. *There are population proportions for which for an underrepresented group j, $\Delta(\mu_j^{DemParity}) < \Delta(\mu_j^{MaxUtil})$ and $\Delta(\mu_j^{EqOpt}) < \Delta(\mu_j^{MaxUtil})$.*

Namely, the condition of fairness unintentionally harms the underrepresented group j. This is analyzed by creating a linear program with the given conditions, then optimizing utility by taking the derivative across parameters.

These results suggest that although the promise of better and stronger algorithms is appealing, there are general issues with any algorithm that assigns individuals scores based on available data. These potential pitfalls in creating a fair algorithm have been considered real world reviews of machine learning algorithms, such as concerning racial bias in medical diagnoses [16].

4. Social Critiques of Predictive Policing

In this section, we bring forth questions from a social prespective about the theoretical understanding of crime that predictive policing embodies and the use of police to deal with crime within the predicted hotspots. There is an extensive literature, both academic and nonacademic, regarding the social critiques and impacts of predictive policing that only focusing on mathematical models do not capture. We mention only several aspects below, deferring to social scientists and theorists for these critiques, and encourage the reader to explore the question more deeply and broadly on their own.

4.1. On the notion of crime. Social scientists and community advocates have long argued that the notion of "crime" is itself a political matter (e.g., [5]). To label a specific act as a crime assumes a particular understanding of the social contract, whereby certain crimes are policed and others are not. Following this line of critique, *The New Inquiry* published a tool called White Collar Crime Risk Zones[3] predicting where financial crimes will happen in the US, trained on incidents of financial malfeasance. As the authors note, "unlike typical predictive policing apps which criminalize poverty, White Collar Crime Risk Zones criminalizes wealth." As argued by Proposition 3.1, due to the focus of predictive policing on crime data that is skewed towards communities of color in poor neighborhoods, the biases that we have discussed above in prediction will lead to overpolicing in those areas.

4.2. Policing as a response to prediction. Researchers argue that predicting crimes—or rather, particular undesirable incidents or activities—is not problematic in and of itself, but rather it is the policing response, known as hotspot policing, or informally as cops on dots, that is problematic and what actually leads to overpolicing (see, for example references cited in Section 4.1 of [2]).

4.3. The broader surveillance complex. Perhaps most importantly, predictive policing is only a single aspect of a much larger phenomenon at the intersection of policing and surveillance. Predictive policing is only a small part of a large network of surveillance systems such as gunshot detectors, automated license plate readers, facial recognition systems, geofencing, and CCTV cameras, as well as predictive systems such as risk terrain modeling, recidivism risk scores, and pretrial detention risk assessments [7]. As such, in order to holistically assess the relevance and impact of predictive policing, it is necessary to consider it in the context of this broader system, which will take us beyond simple mathematical models. Indeed, thinking at this systemic level will require collaboration with social scientists, lawyers, and activists who are familiar with crime and the effects of policing on communities to inform our modeling decision.

5. Conclusion

Predictive policing comes from an interesting mathematical background. However, analyzing the mathematical framework shows fundamental theoretical and empirical issues that have yet to be properly addressed (principally expressed in Section 3). Moreover, there are mathematical results that caution against *any* algorithmic model that inputs a set of features and outputs a risk score (see Theorem 3.3). Furthermore, possibility/impossibility theorems such as the latter suggest that broad theoretical work on algorithmic fairness can have major implications for the application of such prediction systems, which have immense impact in society.

Persistent questions surround the practice of predictive policing: If it replicates the same outcomes as conventional police methods, should it also replicate the biases inherent in those methods? Why do marked disparities persist in patrolled areas between predictive policing and traditional approaches? Should our response to predicted crime prioritize alternatives like health responders or financial relief over police intervention?

While the mathematics and scientific communities aspire to maintain neutrality, the reality is that we introduce our own biases through the fundamental assumptions of the models we create (Proposition 3.1). We must be acutely aware of the potential applications of our mathematical innovations. It is not surprising when after mathematical research is handed over to law enforcement, it is then used to further oppress those already victimized by police violence.

We believe that our awareness of the broader societal context should inform the ethical standards guiding our decisions of which projects we choose to undertake. We hope that this examination of predictive policing prompts a reflection on the projects that the mathematical and scientific broadly opt to pursue.

References

[1] Nil-Jana Akpinar, Maria De-Arteaga, and Alexandra Chouldechova, *The effect of differential victim crime reporting on predictive policing systems*, In *Proceedings of the 2021*

[3] `https://whitecollar.thenewinquiry.com/`

ACM Conference on Fairness, Accountability, and Transparency, pages 838–849, 2021.

[2] Kiana Alikhademi, Emma Drobina, Diandra Prioleau, Brianna Richardson, Duncan Purves, and Juan E. Gilbert, *A review of predictive policing from the perspective of fairness*, Artificial Intelligence and Law, pages 1–17, 2022.

[3] Tarik Aougab, Federico Ardila, Jayadev Athreya, Edray Goins, Christopher Hoffman, Autumn Kent, Lily Khadjavi, Cathy O'Neil, Priyam Patel, and Katrin Wehrheim, *Boycott collaboration with police*, Notices Amer. Math. Soc. **67** (2020), no. 9, 1293.

[4] P. Jeffrey Brantingham, Matthew Valasik, and George O. Mohler, *Does predictive policing lead to biased arrests? Results from a randomized controlled trial*, Statistics and Public Policy **5** (2018), no. 1, 1–6.

[5] Stanley Cohen, *Crime and politics: spot the difference*, The British Journal of Sociology **47** (1996), no. 1, 1–21.

[6] Danielle Ensign, Sorelle A. Friedler, Scott Neville, Carlos Scheidegger, and Suresh Venkatasubramanian, *Runaway feedback loops in predictive policing*, In *Conference on Fairness, Accountability, and Transparency*, pages 160–171, PMLR, 2018.

[7] Andrew Guthrie Ferguson, *The rise of big data policing*, In *The Rise of Big Data Policing*, New York University Press, 2017.

[8] Dylan J. Fitzpatrick, Wilpen L. Gorr, and Daniel B. Neill, *Policing chronic and temporary hot spots of violent crime: A controlled field experiment*, Preprint, `arXiv:2011.06019`, 2020.

[9] Dawson Kinsman and Tian An Wong, *Proactive policing as reinforcement learning*, In Krystal Maughan, Rosanne Liu, and Thomas F. Burns, editors, *The First Tiny Papers Track at ICLR 2023, Tiny Papers @ ICLR 2023, Kigali, Rwanda, May 5, 2023*, OpenReview.net, 2023.

[10] Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan, *Inherent trade-offs in the fair determination of risk scores*, 8th Innovations in Theoretical Computer Science Conference, LIPIcs. Leibniz Int. Proc. Inform., vol. 67, Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2017, pp. Art. No. 43, 23. MR3754967

[11] Lydia T. Liu, Sarah Dean, Esther Rolf, Max Simchowitz, and Moritz Hardt, *Delayed impact of fair machine learning*, In *International Conference on Machine Learning*, pages 3150–3158, PMLR, 2018.

[12] Kristian Lum and William Isaac, *To predict and serve?*, Significance **13** (2016), no. 5, 14–19.

[13] John Marzulli, *We fabricated drug charges against innocent people to meet arrest quotas, former detective testifies*, NY Daily News (2011).

[14] 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, DOI 10.1080/01621459.2015.1077710. MR3449035

[15] George Mohler, Rajeev Raje, Jeremy Carter, Matthew Valasik, and Jeffrey Brantingham, *A penalized likelihood method for balancing accuracy and fairness in predictive policing*, In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 2454–2459, IEEE, 2018.

[16] Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan, *Dissecting racial bias in an algorithm used to manage the health of populations*, Science **366** (2019), no. 6464, 447–453.

[17] Ben Poston, Joel Rubin, and Anthony Pesce, *LAPD underreported serious assaults, skewing crime stats for 8 years*, Los Angeles Times (2015).

[18] Henrik Renlund, *Generalized Pólya urns via stochastic approximation*, Preprint, `arXiv:1002.3716`, 2010.

[19] Rashida Richardson, Jason M. Schultz, and Kate Crawford, *Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice*, NYUL Rev. Online **94** (2019), no. 15.

[20] Aaron Shapiro, *Reform predictive policing*, Nature **541** (2017), no. 7638, 458–460.

[21] M. B. Short, A. L. Bertozzi, and P. J. Brantingham, *Nonlinear patterns in urban crime: hotspots, bifurcations, and suppression*, SIAM J. Appl. Dyn. Syst. **9** (2010), no. 2, 462–483, DOI 10.1137/090759069. MR2670006

[22] M. B. Short, M. R. D'Orsogna, V. B. Pasour, G. E. Tita, P. J. Brantingham, A. L. Bertozzi, and L. B. Chayes, *A statistical model of criminal behavior*, Math. Models Methods Appl. Sci. **18** (2008), no. suppl., 1249–1267, DOI 10.1142/S0218202508003029. MR2438215

Joseph Johnson

Theo McKenzie

Tian An Wong

Credits

Photo of Joseph Johnson is courtesy of Adeline Jurrs.
Photo of Theo McKenzie is courtesy of Saumya Karki.
Photo of Tian An Wong is courtesy of Jim Gipe.