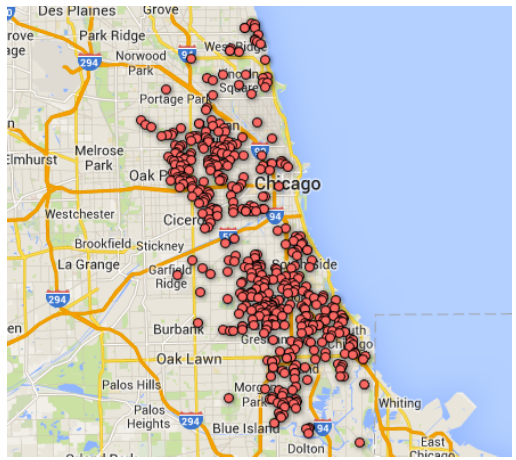
Learning Point Processes via Reinforcement Learning

Published at *Conference on in Neural Information Processing Systems (NeurIPS)*, 2018, spotlight (3.5%)

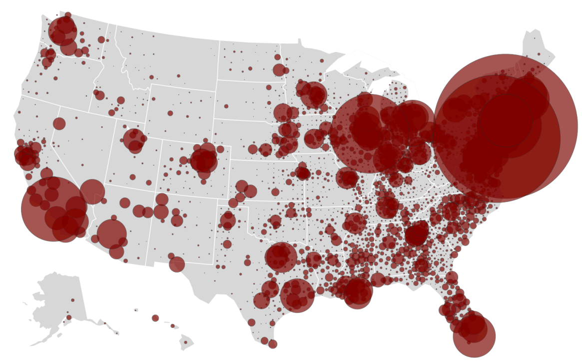
Motivating Examples

Develop efficient and stable algorithms for learning sophisticated (spatio-temporal) point processes

Bird migration



Chicago crimes



U.S. Confirmed Covid-19 Cases Up to May 2020

Challenges for Maximum-Likelihood

Specify conditional intensity

$$\lambda_\theta(t, u|\mathcal{H}_t)dt = \mu(u) + \sum_{i,\, t_i<t} g_\theta(u - u_i, t - t_i)$$

as a parametric/non-parametric/neural-based form

Learn model parameter θ by maximizing likelihood

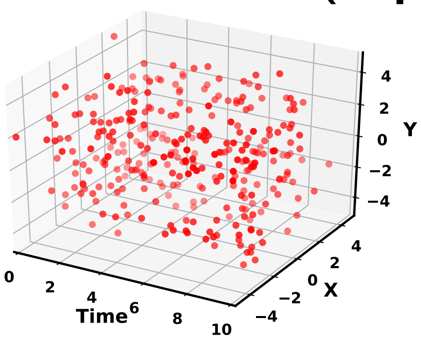
$$\mathcal{L}(\theta) = \exp\left(-\boxed{\int\int_{(0,T)\times S}} \boxed{\lambda_\theta(t, u|\mathcal{H}_t)} dt\, du \right) \prod_i \boxed{\lambda_\theta(t_i, u_i|\mathcal{H}_t)}$$

Computational challenge **Model-misspecification**

How to effectively learn point processes with complex intensity function?

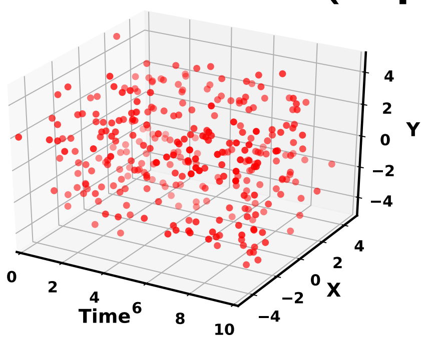
Our Reinforcement Learning Framework

Observations (expert π_E)

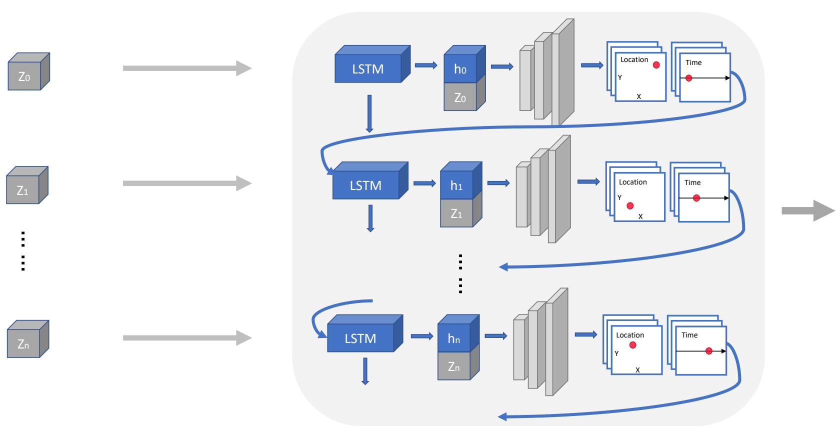


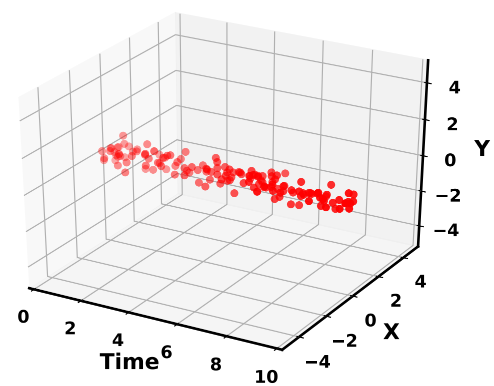
Our Reinforcement Learning Framework

Observations (expert π_E)

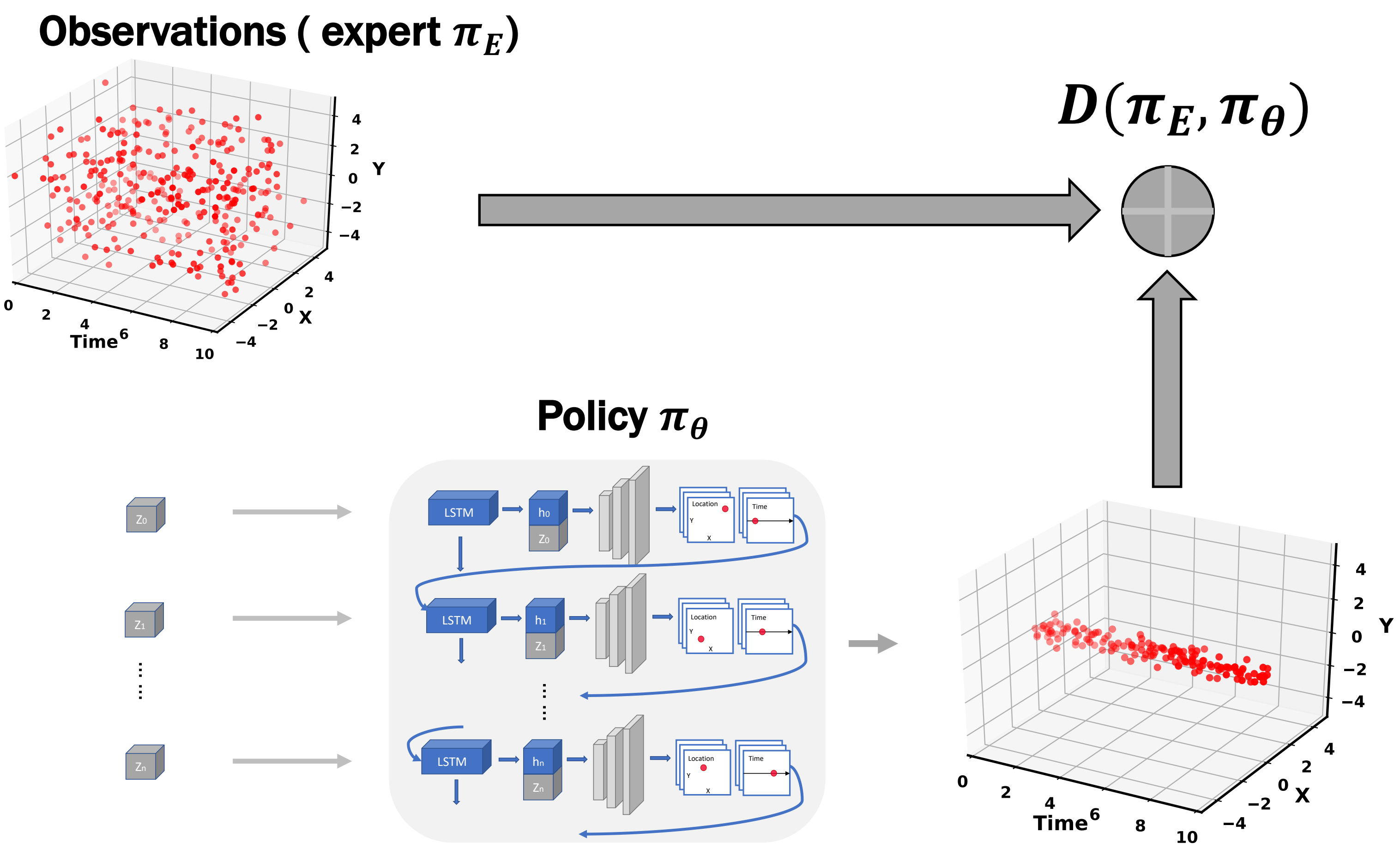


Policy π_θ

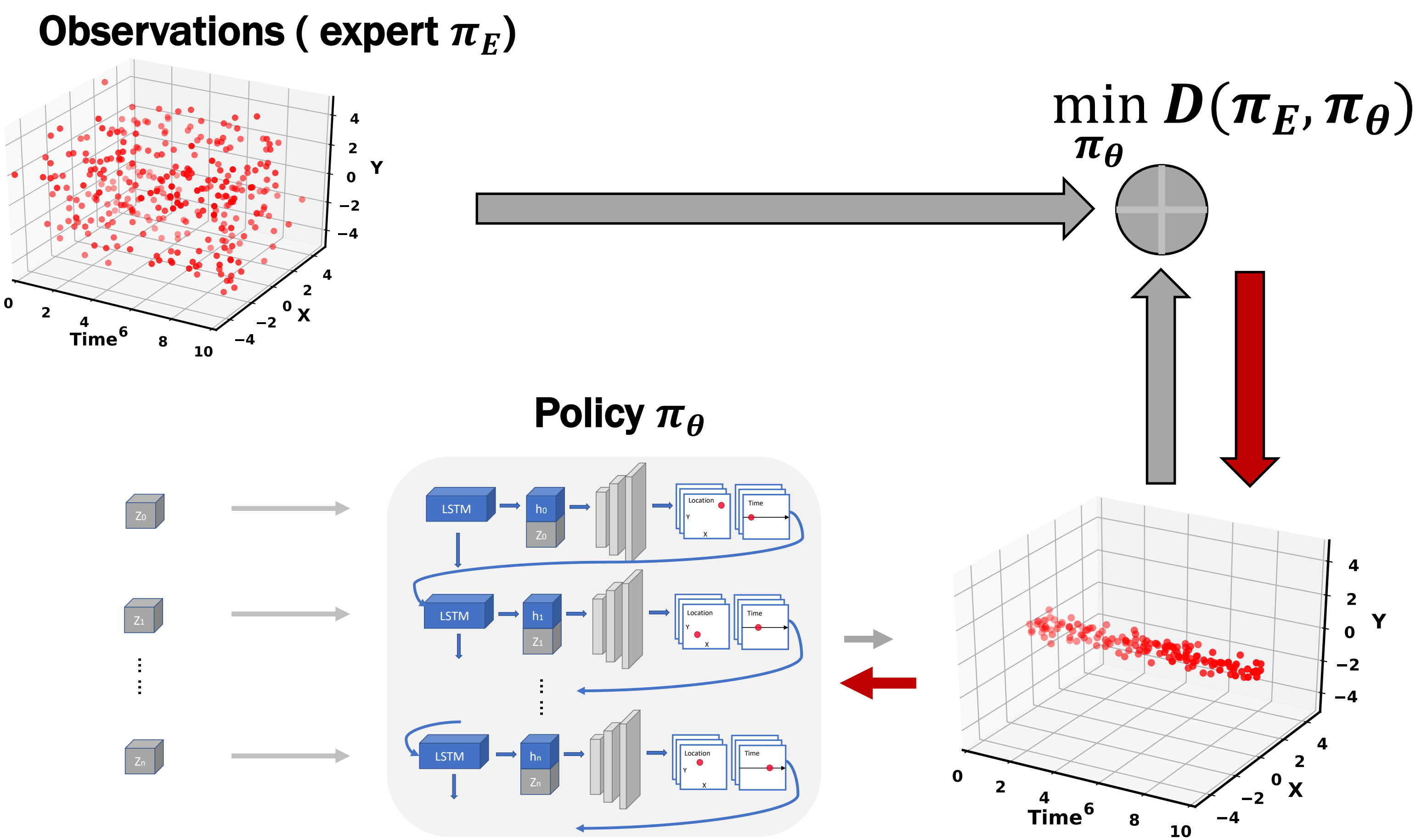




Our Reinforcement Learning Framework



Our Reinforcement Learning Framework

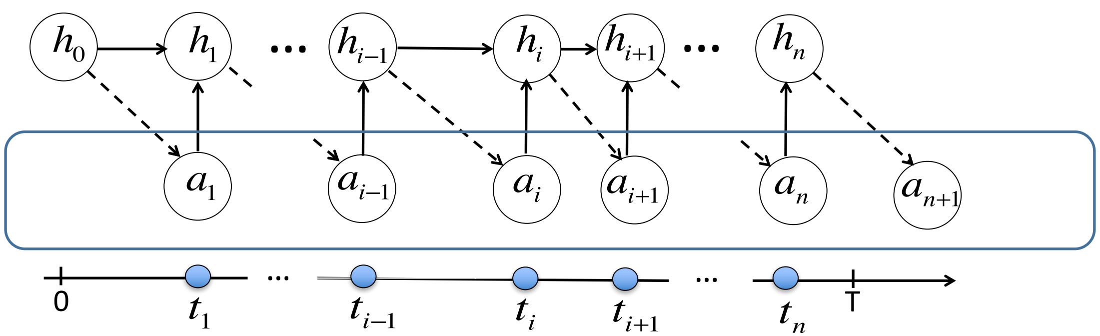


Policy Model

Treat $\pi_\theta(a|s_t)$ as the conditional density for the next time-to-event and location

$$\pi_\theta(a|s_t) = p(a_i \mid a_{i-1}, \ldots, a_1)$$

$\pi_\theta(a|s_t)$ examples: RNN, LSTM, Attention Model



Flexible model to capture nonlinear and long-range sequential dependency structure in events

Imitation Learning: Minimax Formulation

Given observed sequence of events

$$\xi := \{ e_1, e_2, \dots, e_{N_T^{\xi}} \} \qquad \xi \sim \pi_E$$

Generate sequence of events from $\pi_\theta(a|s_t)$

$$\eta := \{ a_1, a_2, \dots, a_{N_T^{\eta}} \} \qquad \eta \sim \pi_\theta$$

Imitation Learning requires:

Learn optimal reward function as (consider the worst case)

$$r^* = \arg\max_{r \in \mathcal{F}} \left(\mathbb{E}_{\xi \sim \pi_E} \left[\sum_{i=1}^{N_T^{\xi}} r(e_i) \right] - \max_{\pi_\theta \in \mathcal{G}} \mathbb{E}_{\eta \sim \pi_\theta} \left[\sum_{i=1}^{N_T^{\eta}} r(a_i) \right] \right)$$

Obtain optimal policy as

Time-consuming!

$$\pi_{\theta^*} = \arg \max_{\pi_\theta \in \mathcal{G}} \mathbb{E}_{\eta \sim \pi_\theta} \left[\sum_{i=1}^{N_T^{\eta}} r^*(a_i) \right]$$

Analytical Nonparametric Reward

Choose reward from the unit ball in Reproducing Kernel Hilbert Space (RKHS)

$$r \in \mathcal{F} \qquad\qquad \mathcal{F} = \{\, r \mid ||r||_{\mathcal{H}} \leq 1 \}$$

Analytical Nonparametric Reward

Choose reward from the unit ball in Reproducing Kernel Hilbert Space (RKHS)

$$r \in \mathcal{F} \qquad \mathcal{F} = \{\, r \mid \|r\|_{\mathcal{H}} \leq 1 \}$$

Then

$$\begin{aligned}
J(\pi_E) &:= \mathbb{E}_{\xi\sim\pi_E}\left[\sum_{i=1}^{N_T^{\xi}} r(e_i)\right] \\
&= \mathbb{E}_{\xi\sim\pi_E}\left[\iint_{[0,T]\times S} r(t,u)\, dN_{t\times u}^{\xi}\right] \\
&= \mathbb{E}_{\xi\sim\pi_E}\left[\iint_{[0,T]\times S} \langle r, k((t,u),\cdot\,)\rangle\, dN_{t\times u}^{\xi}\right] \\
&= \left\langle r, \boxed{\mathbb{E}_{\xi\sim\pi_E}\left[\iint_{[0,T]\times S} k\big((t,u),\cdot\,\big)\, dN_{t\times u}^{\xi}\right]}\right\rangle \longrightarrow \mu_{\pi_E} \\
&= \langle r, \mu_{\pi_E} \rangle
\end{aligned}$$

Analytical Nonparametric Reward

Imitation Learning

$$r^* = \arg \boxed{\max_{||r||_{\mathcal{H}} \leq 1} \left(\mathbb{E}_{\xi \sim \pi_E} \left[\sum_{i=1}^{N_T^{\xi}} r(e_i) \right] - \max_{\pi_\theta \in \mathcal{G}} \mathbb{E}_{\eta \sim \pi_\theta} \left[\sum_{i=1}^{N_T^{\eta}} r(a_i) \right] \right)}$$

$$\max_{||r||_{\mathcal{H}} \leq 1} \left(J(\pi_E) - \max_{\pi_\theta \in \mathcal{G}} J(\pi_\theta) \right)$$

Analytical Nonparametric Reward

Imitation Learning

$$r^* = \arg \boxed{\max_{||r||_{\mathcal{H}} \leq 1} \left(\mathbb{E}_{\xi \sim \pi_E}\left[\sum_{i=1}^{N_T^{\xi}} r(e_i)\right] - \max_{\pi_\theta \in \mathcal{G}} \mathbb{E}_{\eta \sim \pi_\theta}\left[\sum_{i=1}^{N_T^{\eta}} r(a_i)\right] \right)}$$

$$\max_{||r||_{\mathcal{H}} \leq 1} \left(J(\pi_E) - \max_{\pi_\theta \in \mathcal{G}} J(\pi_\theta) \right)$$

$$= \max_{||r||_{\mathcal{H}} \leq 1} \min_{\pi_\theta \in \mathcal{G}} \left(J(\pi_E) - J(\pi_\theta) \right)$$

$$= \max_{||r||_{\mathcal{H}} \leq 1} \min_{\pi_\theta \in \mathcal{G}} \left(\langle r, \mu_{\pi_E} \rangle - \langle r, \mu_{\pi_\theta} \rangle \right)$$

$$= \max_{||r||_{\mathcal{H}} \leq 1} \min_{\pi_\theta \in \mathcal{G}} \langle r, \mu_{\pi_E} - \mu_{\pi_\theta} \rangle$$

$$= \min_{\pi_\theta \in \mathcal{G}} \boxed{\max_{||r||_{\mathcal{H}} \leq 1} \langle r, \mu_{\pi_E} - \mu_{\pi_\theta} \rangle}$$

Analytical Nonparametric Reward

Imitation Learning

$$r^* = \arg\ \max_{||r||_{\mathcal{H}} \le 1} \left(\mathbb{E}_{\xi \sim \pi_E}\left[\sum_{i=1}^{N_T^{\xi}} r(e_i)\right] - \max_{\pi_\theta \in \mathcal{G}} \mathbb{E}_{\eta \sim \pi_\theta}\left[\sum_{i=1}^{N_T^{\eta}} r(a_i)\right] \right)$$

$$\max_{||r||_{\mathcal{H}} \le 1} \left(J(\pi_E) - \max_{\pi_\theta \in \mathcal{G}} J(\pi_\theta) \right)$$

$$= \max_{||r||_{\mathcal{H}} \le 1} \min_{\pi_\theta \in \mathcal{G}} \left(J(\pi_E) - J(\pi_\theta) \right)$$

$$= \max_{||r||_{\mathcal{H}} \le 1} \min_{\pi_\theta \in \mathcal{G}} \left(\langle r, \mu_{\pi_E} \rangle - \langle r, \mu_{\pi_\theta} \rangle \right)$$

$$= \max_{||r||_{\mathcal{H}} \le 1} \min_{\pi_\theta \in \mathcal{G}} \langle r, \mu_{\pi_E} - \mu_{\pi_\theta} \rangle$$

$$= \min_{\pi_\theta \in \mathcal{G}} \boxed{\max_{||r||_{\mathcal{H}} \le 1} \langle r, \mu_{\pi_E} - \mu_{\pi_\theta} \rangle}$$

Finite sample estimate

$$= \min_{\pi_\theta \in \mathcal{G}} \left\| \mu_{\pi_E} - \mu_{\pi_\theta} \right\|_{\mathcal{H}}$$

Minimization

where $r^* = \dfrac{\mu_{\pi_E} - \mu_{\pi_\theta}}{\left\| \mu_{\pi_E} - \mu_{\pi_\theta} \right\|_{\mathcal{H}}}$

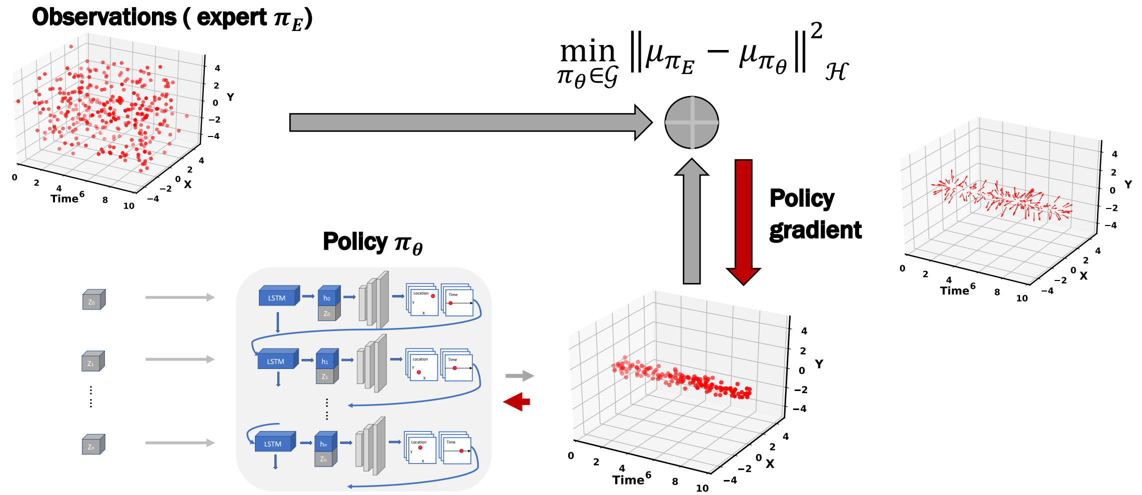
Policy Learning

Our Reinforcement Learning Framework

$$\pi_{\theta^*} = \arg \min_{\pi_\theta \in \mathcal{G}} \left\| \mu_{\pi_E} - \mu_{\pi_\theta} \right\|_{\mathcal{H}}$$

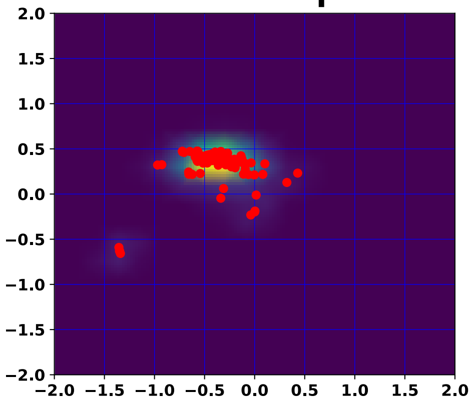
Learn policy π_θ

- Policy gradient
- Reparameterization trick (end-to-end, reduce gradient variance)

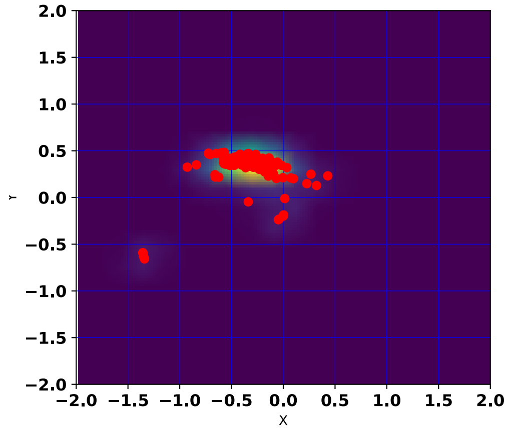


Numerical Results: Data Description

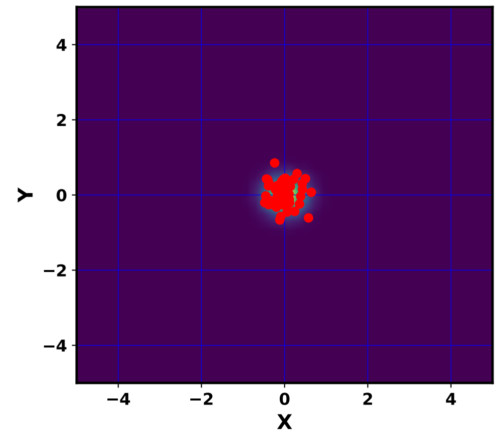
New York Taxi Trips



Observation

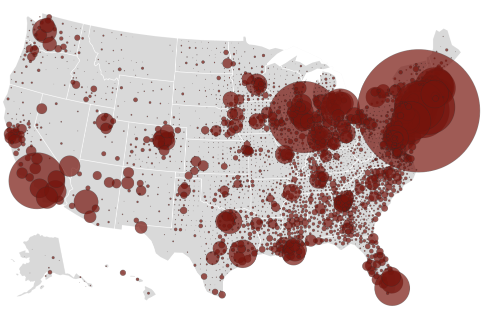


Generated by learner π_{θ^*}

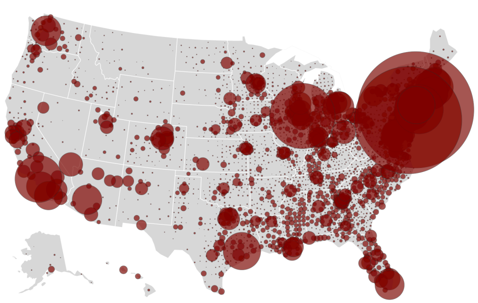


Baseline (MLE, Triggering function with decomposed spatial and temporal components)

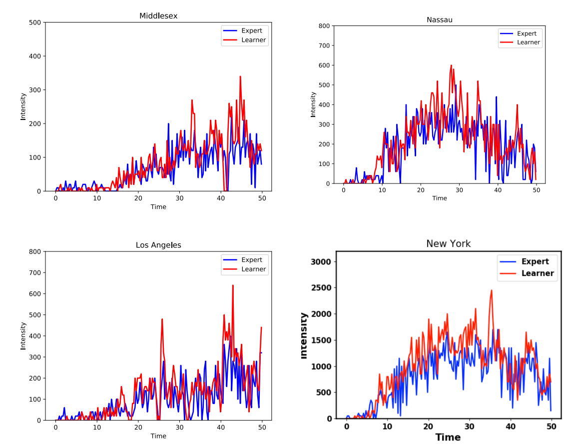
Confirmed COVID Cases until May 20th, 2020



Observation

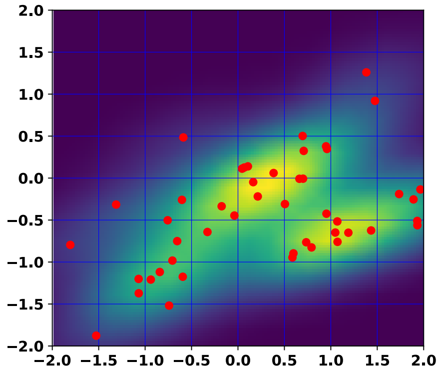


Generated by learner π_{θ^*}

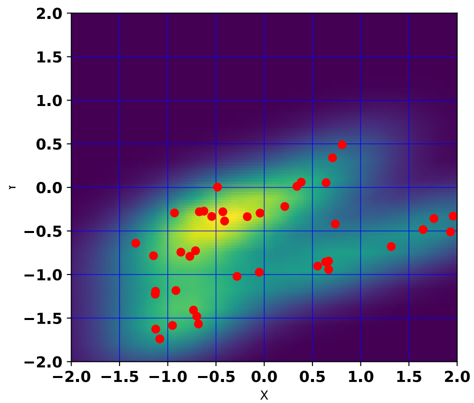


Numerical Results: Data Prediction

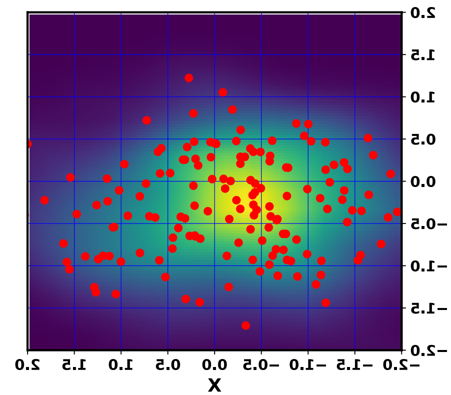
Crime Events



Observation

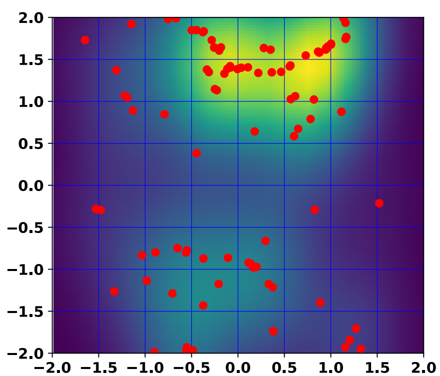


Predicted by learner π_{θ^*}

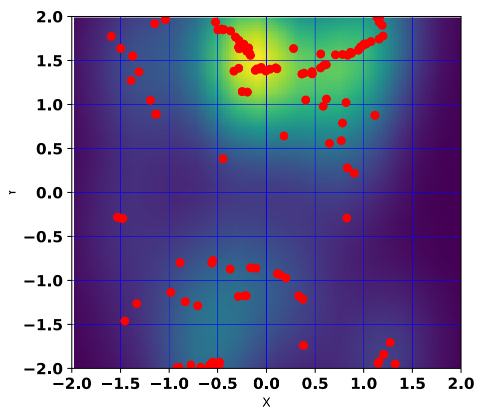


Baseline

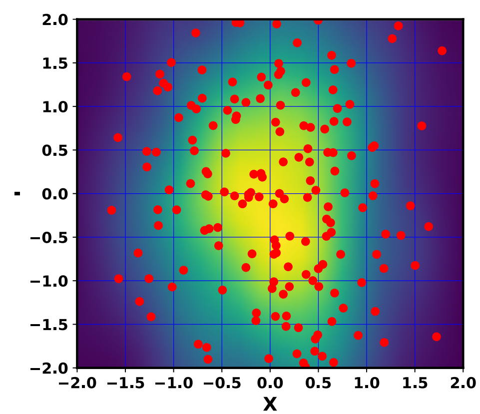
Earthquakes



Observation



Predicted by learner π_{θ^*}



Baseline