A spatio-temporal point processes (STPP) is a random process whose realization consists of an ordered sequence of events localized in time and space, i.e.,

\[ \mathcal{H}_t := \{ \mathbf{H}_t = (u, t) \mid \mathbf{H}_t \in \mathbb{R}^n \times S, u \in S \} \]

where \( t \in \mathbb{R}^+ \) is the time of the occurrence of event \( t \in \mathbb{R}^+ \), and \( u \in S \) is the location of the event.

### Intensity Function

The STPP can be characterized by a conditional intensity function, denoted as

\[ \lambda(t, u \mid H_t) \]

which means the current intensity depends on the history \( H_t \) up to time \( t \), where \( H_t \) is the \( \sigma \)-algebra.

- Denote \( N(A) \) the number of \( (t, u) \) falling in a set \( A \subset \mathbb{R}^+ \times S \).

\[ \lambda(t, u \mid H_t) dt du = \mathbb{E}[N(dt \times du) \mid H_t] \]

- Use temporal point process as an example, \( \lambda(t \mid H_t) = \frac{\text{count}(t)}{\text{area}(t)} \).

### Classic Maximum-Likelihood-Based Methods

- Handcraft the intensity function \( \lambda(t, u \mid H_t) \) to capture the potentially complex triggering and clustering pattern of events.
- Estimate \( \theta \) by maximizing likelihood of a realization of \( \{ \mathbf{H}_1, \ldots, \mathbf{H}_n \} \)

\[ \rho(\theta_1, \ldots, \theta_n) = \exp \left\{ -\int \int \lambda(t, u \mid H_t) dt du \right\} \prod_{i=1}^n \lambda(t_i, u_i \mid H_{t_i}) \]

- Challenges: (1) trade-off between model flexibility and model complexity; (2) limited prior knowledge and model misspecification.

### New Imitation Learning Method

- **Main idea**: Learn a generative model \( \pi_\theta(a|s_t) \), where \( a \in \mathbb{R}^+ \times S \), to mimic the behaviors of observed STPP. In this way, \( \pi_\theta(a|s_t) \) will be able to capture the dynamic structures of the observed STPP.

- **Highlights**:
  1. Directly 
     - match the counting measure: of the point processes generated by \( \pi_\theta(a|s_t) \) to the observed STPP.
  2. Avoid alternating minimax optimization in traditional imitation learning, and
     - is simply a minimization problem.
  3. Monitor the quality of learning via a proposed metric tailored for point processes defined in the reproducing kernel Hilbert Space (RKHS).

### Generative Model

- Treat \( \pi_\theta(a|s_t) \), where \( a = (t, u) \in \mathbb{R}^+ \times S \), as the conditional density for event \( \mathbf{H}_t \):

\[ \pi_\theta(a|s_t) = \pi(t, u | \mathbf{H}_t) \]

- Model: Long short term memory networks (LSTM), which is flexible to capture the nonlinear and long range sequential dependency structure.

- History information \( \{ \mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_t \} \) is recorded in the last hidden state \( \mathbf{h}_t \), where \( h \in \mathbb{R}^m \).

- Generating mechanism:

\[ a_i \sim \pi(a_t \mid a_0 = \Theta(h_{t-1})), \quad u_i \sim \mathcal{N}(\mathbf{V}_a + W_{\mathbf{h}_t}, 1) \quad h_0 = 0, \]

where \( \mathbf{h}_t \) is a **nonlinear** mapping from \( \mathbb{R}^m \) to the parameter space of probability density \( \pi_\theta \).

- For example, let \( \pi_\theta \) be exponential distribution to produce \( t \) and bivariate Gaussian distribution to produce \( u \).

### Learning

- **Traditional Imitation Learning**: Minimax Problem

\[ r^* = \max_r \left( \mathbb{E}_{\pi} \left[ \sum_{i=1}^{N_0} \epsilon_i(r) \right] - \max_{\pi \not= \pi_\theta} \mathbb{E}_{\pi} \left[ \sum_{i=1}^{N_0} \epsilon_i(r) \right] \right) \tag{1} \]

- Time-consuming in that it requires to solve the inner maximization problem repeatedly when estimate \( r^* \).

### Learning (to be continued)

- **Our Imitation Learning**: Minimization Problem

Main idea: Choose reward function \( r(a) \), where \( a \in \mathbb{R}^+ \times S \), as a unit ball in RKHS denoted as \( \mathcal{F} \), and this leads to a nonparametric closed-form estimation for \( r^* \).

- Details: For short notation, we denote \( \mathcal{N}(\cdot, \cdot) \) as the counting process associated with sample path \( \eta \), and \( k(a, a') \) is a universal RKHS kernel.

\[ r_\pi := \mathbb{E}_{\pi} \left[ \phi(\eta) \right] \]

where \( a := (t, u) \), \( dN(a) := dN(a) \) is the counting process associated with sample path \( \eta \), and \( k(a, a') \) is a universal RKHS kernel.

- From (1), \( r^* \) is obtained by

\[ \max_{\pi \not= \pi_\theta} \mathbb{E}_{\pi} - \min_{\pi \not= \pi_\theta} \mathbb{E}_{\pi} - \min_{\pi \not= \pi_\theta} \mathbb{E}_{\pi} \]

where the first equality is guaranteed by the minimax theorem, and

\[ r^* = \mathbb{E}_{\pi} - \min_{\pi \not= \pi_\theta} \mathbb{E}_{\pi} - \min_{\pi \not= \pi_\theta} \mathbb{E}_{\pi} \]

In this way, we change the original minimax problem to simply a minimization problem.

### Algorithm

**Algorithm 1 RLPP**: Mini-batch Reinforcement Learning for Learning Point Processes

- **Initialize model parameters \( \theta \)**

  - for number of mini-batch \( \text{batch} \)

  - Sample minibatch of \( L \) trajectory of events \( \{ (t^1, u^1), \ldots, (t^L, u^L) \} \) from expert, where \( t^i = (t^i, \ldots, t^i) \).

  - Sample minibatch of \( M \) trajectory of events \( \{ (t^1, u^1), \ldots, (t^M, u^M) \} \) from policy \( \pi_{\pi_\theta} \), where \( \pi(t, u) \)

- **Update \( \pi_\theta \)** by policy gradient:

\[ \theta = \theta + \frac{\Delta \theta}{|| \Delta \theta ||} \]

where \( \Delta \theta \) can be computed, and \( \frac{\Delta \theta}{|| \Delta \theta ||} \) can be estimated by 1 expert trajectories and \( (M-1) \) roll-out samples without \( \gamma \).

\[ r_\pi(\theta) = \mathbb{E}_{\pi} \left[ \sum_{i=1}^{N_0} \epsilon_i(r) \right] \]

**Numerical Results**

(a): 911 dataset

(b): MIMIC dataset

Figure 2: KS test results: CDF of p-values.