Learning with
Temporal Point Processes

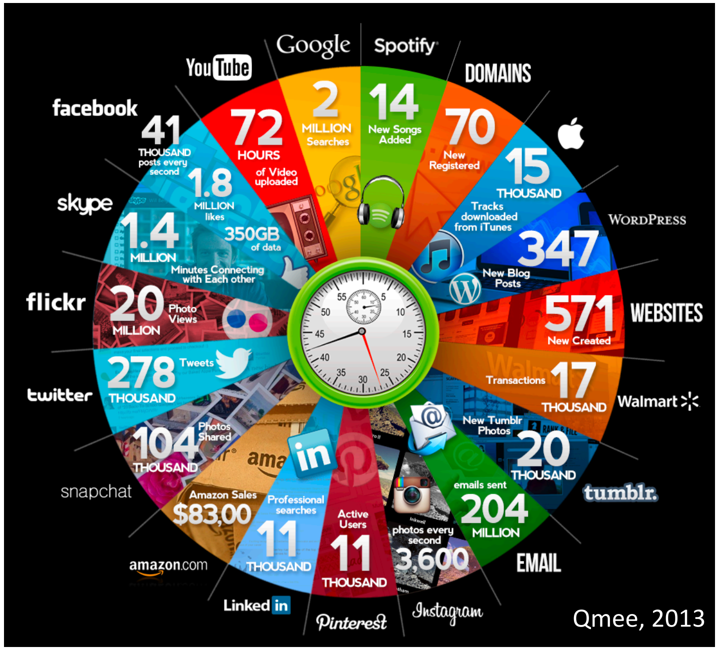
Manuel Gomez Rodriguez
MPI for Software Systems

Isabel Valera
MPI for Intelligent Systems

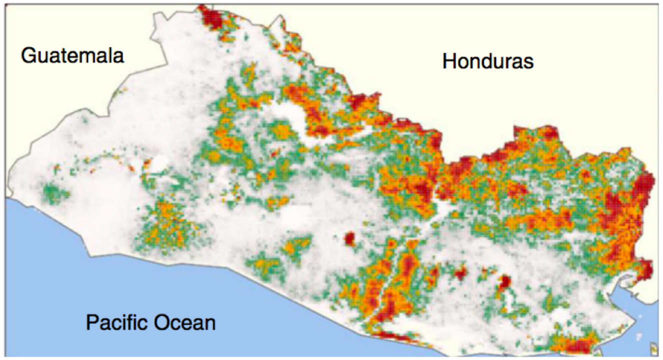
Slides/references: http://learning.mpi-sws.org/tpp-icml18

ICML TUTORIAL, JULY 2018

Many discrete *events* in continuous time

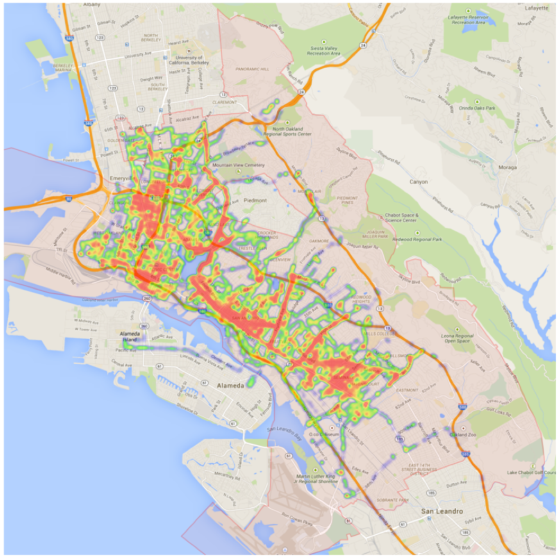


Online actions



Disease dynamics

Financial trading



Mobility dynamics

Variety of processes behind these events

Events are (noisy) observations of a variety of complex dynamic processes…

Stock trading

Flu spreading

Article creation in Wikipedia

News spread in Twitter

Reviews and sales in Amazon

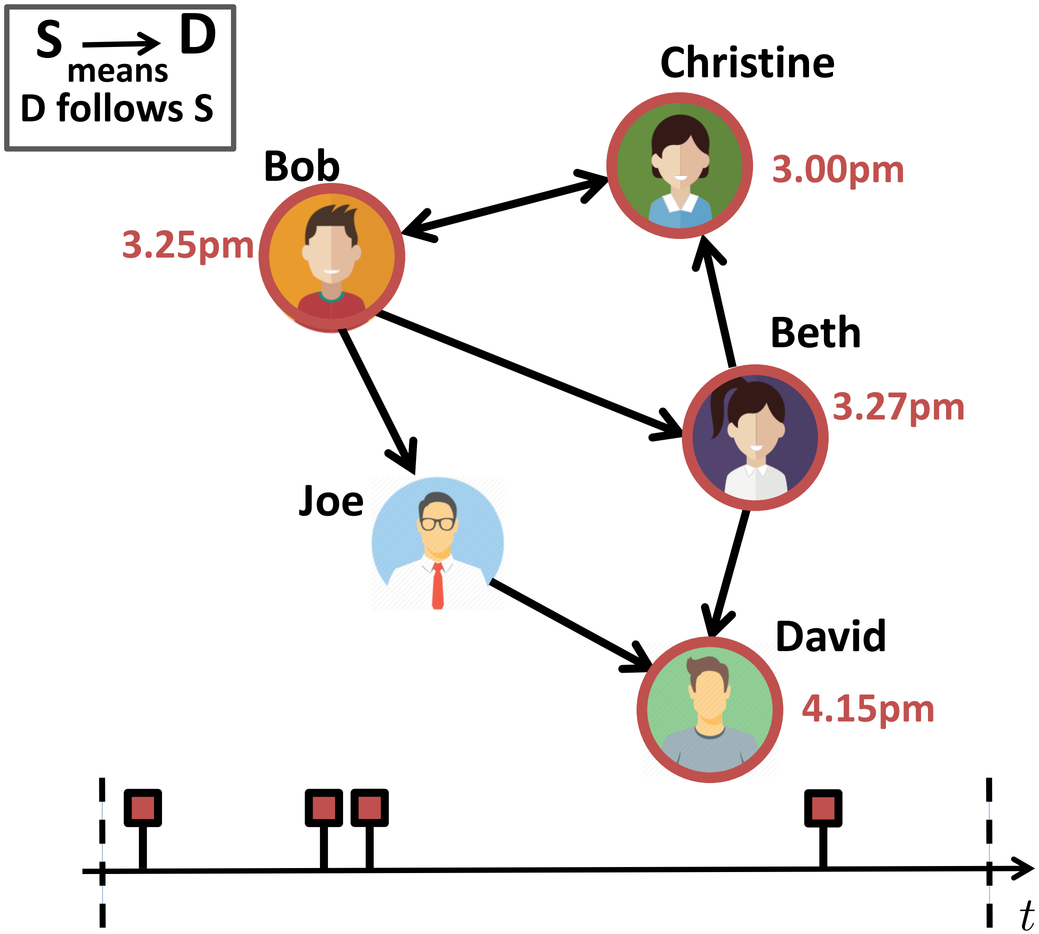
Ride-sharing requests

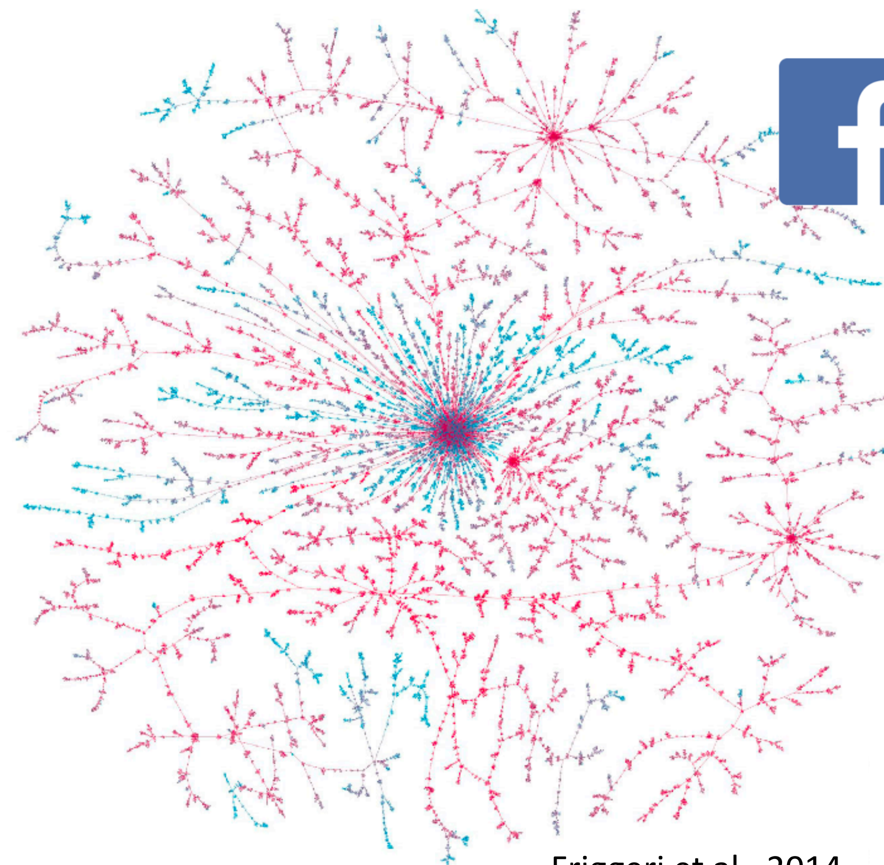
A user's reputation in Quora

FAST → **SLOW**

…in a wide range of temporal scales.

Example I: Information propagation





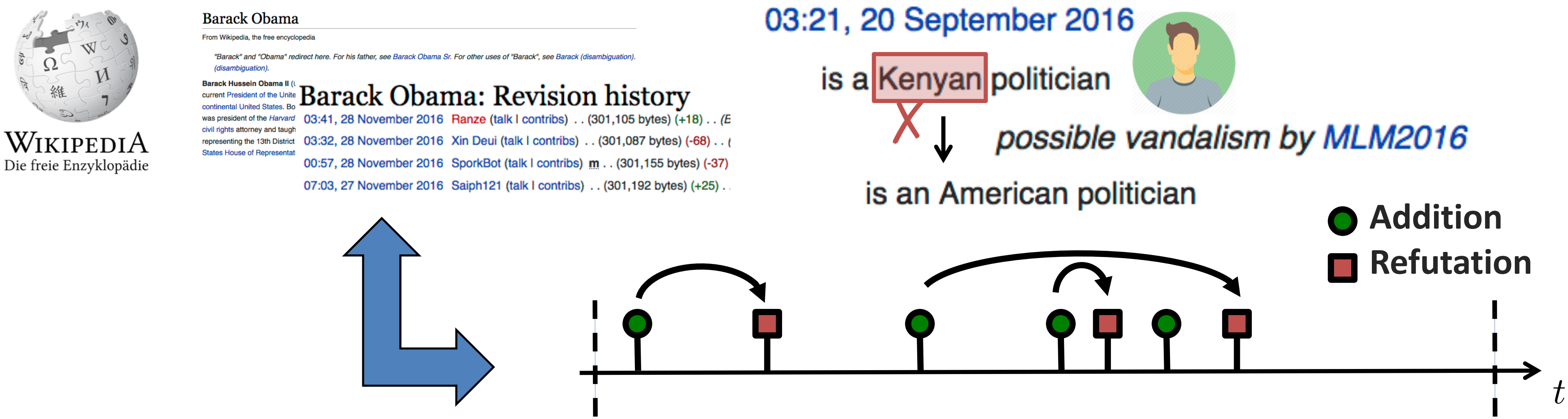
Friggeri et al., 2014

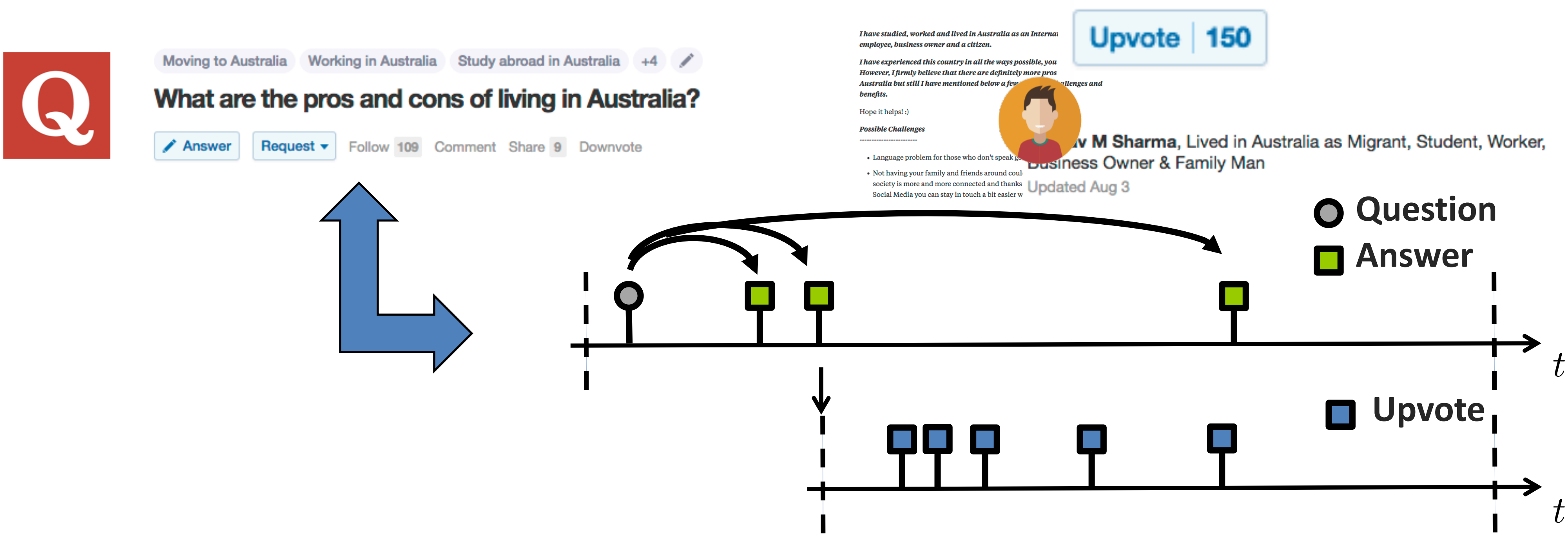
They can have an impact in the off-line world

theguardian

Click and elect: how fake news helped Donald Trump win a real election

Example II: Knowledge creation



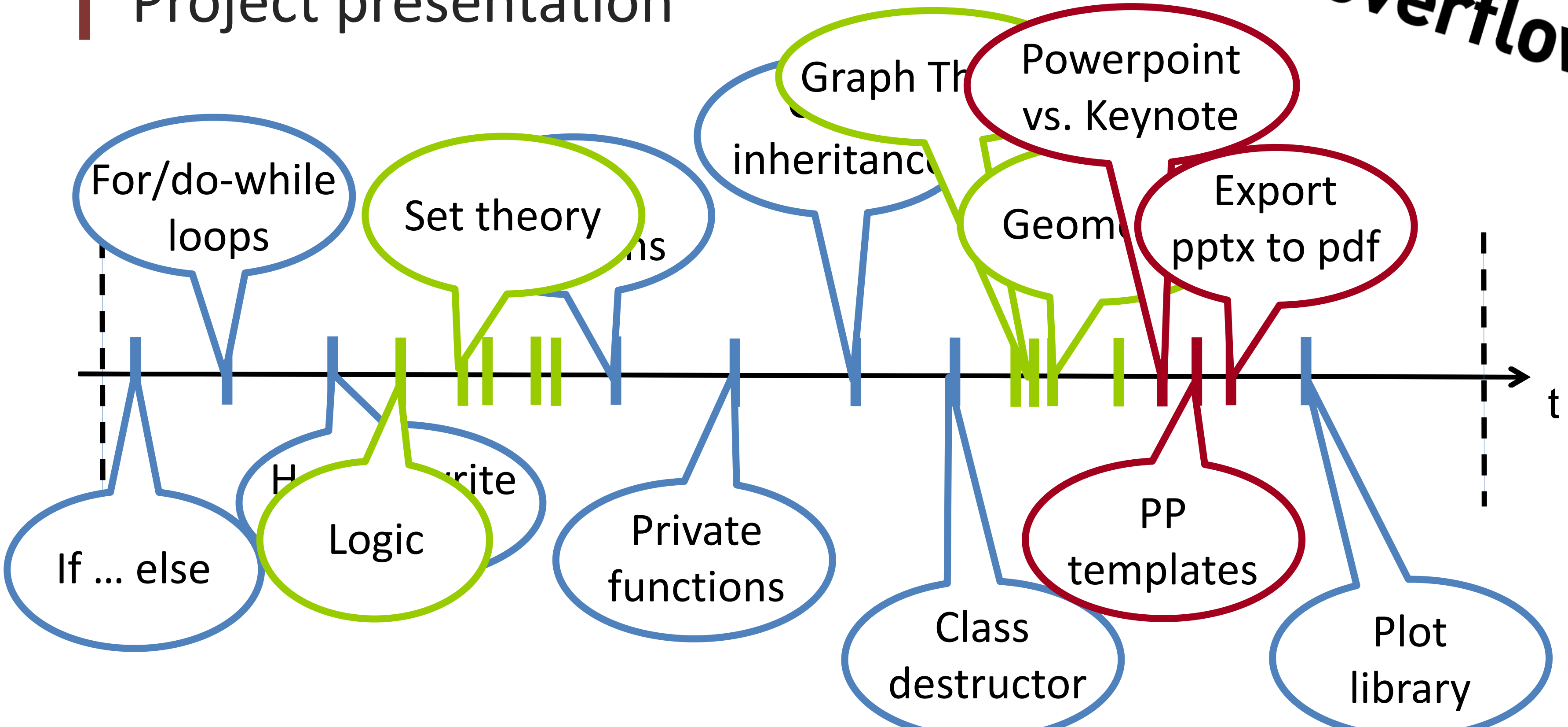


Example III: Human learning

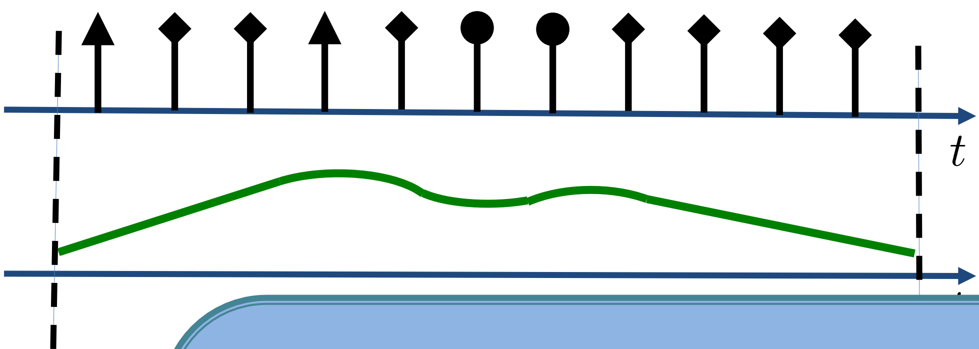
1st year computer science student

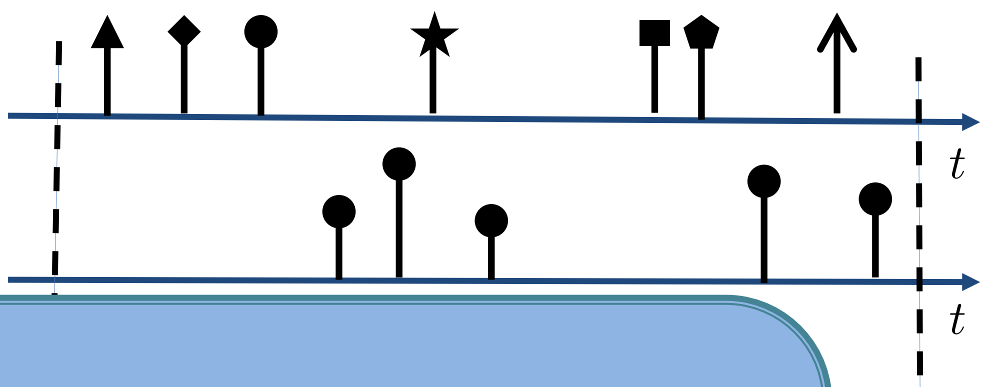
- Introduction to programming
- Discrete math
- Project presentation

stackoverflow



Aren’t these event traces just time series?





Di

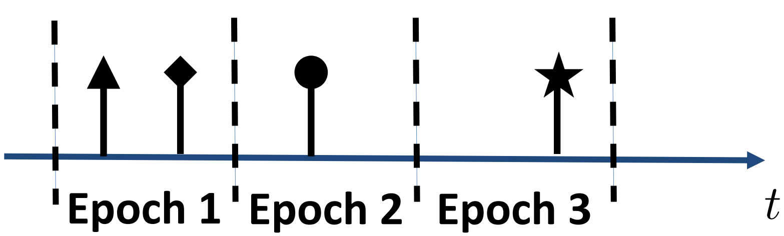
The framework of
temporal point processes
provides a ***native representation***

W

poch?

What if no event in one epoch?

What about time-related queries?



Outline of the Seminar

Temporal Point processes (TPPs): Intro

1. Intensity function
2. Basic building blocks
3. Superposition
4. Marks and SDEs with jumps

Next

Models & Inference

1. Modeling event sequences
2. Clustering event sequences
3. Capturing complex dynamics
4. Causal reasoning on event sequences

RL & Control

1. Marked TPPs: a new setting
2. Stochastic optimal control
3. Reinforcement learning

Slides/references: learning.mpi-sws.org/tpp-icml18

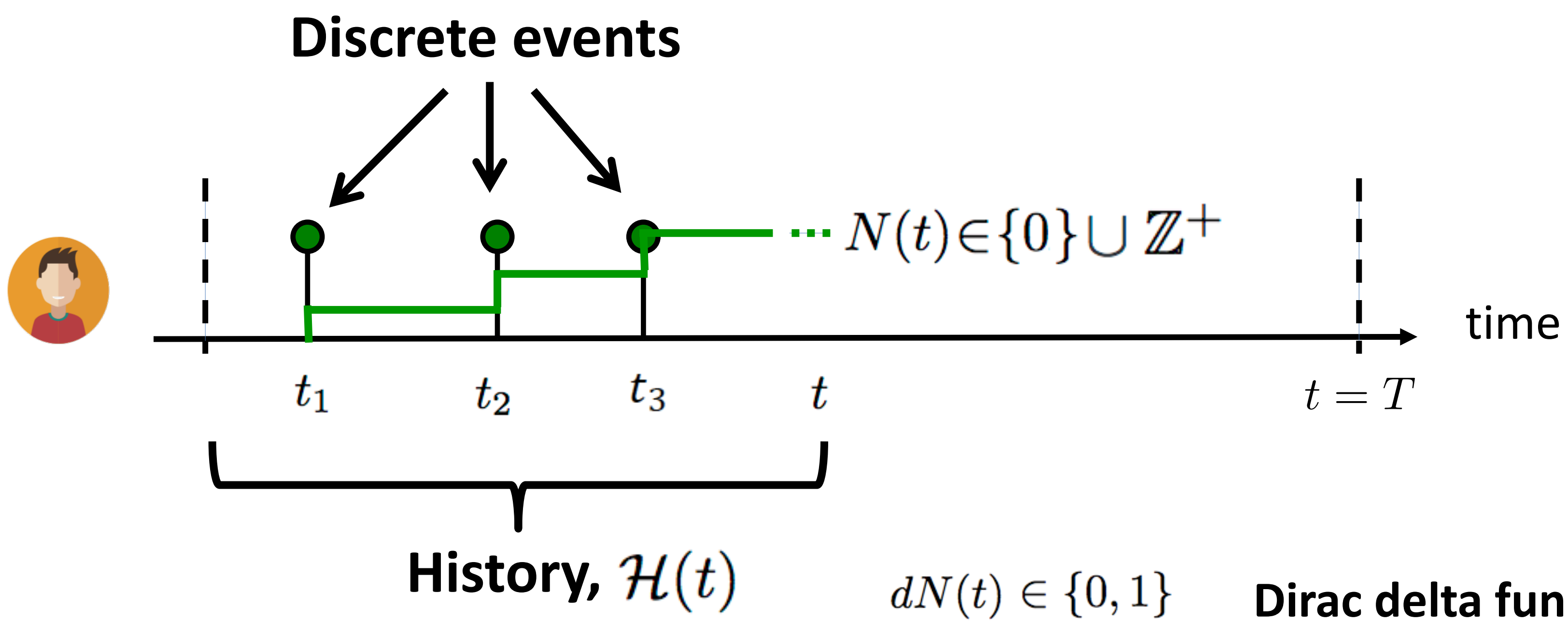
Temporal Point Processes (TPPs): Introduction

1. Intensity function
2. Basic building blocks
3. Superposition
4. Marks and SDEs with jumps

Temporal point processes

Temporal point process:

A random process whose realization consists of **discrete events localized in time** $\mathcal{H} = \{t_i\}$

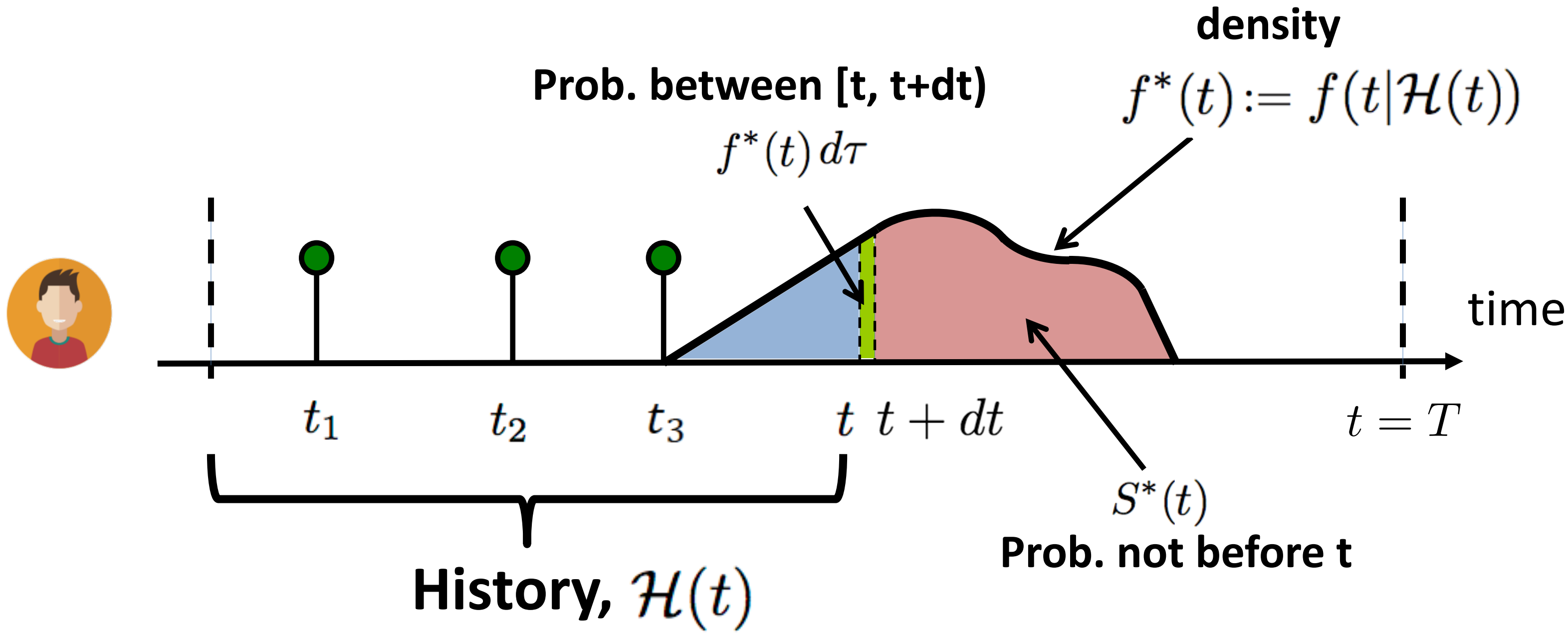


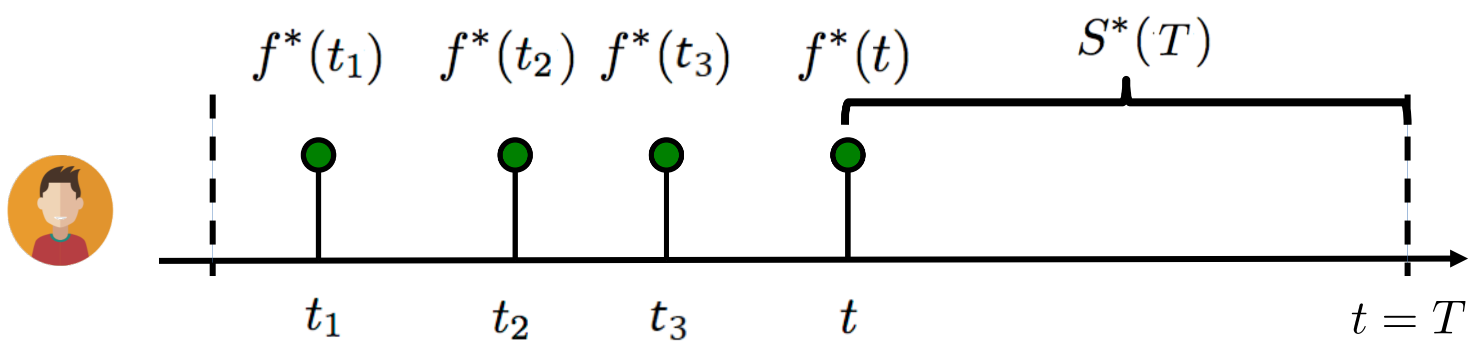
$dN(t) \in \{0, 1\}$

Dirac delta function

Formally: $N(t) = \int_0^t dN(s)$ ⇨ $dN(t) = \sum_{t_i \in \mathcal{H}} \delta(t - t_i)\, dt$

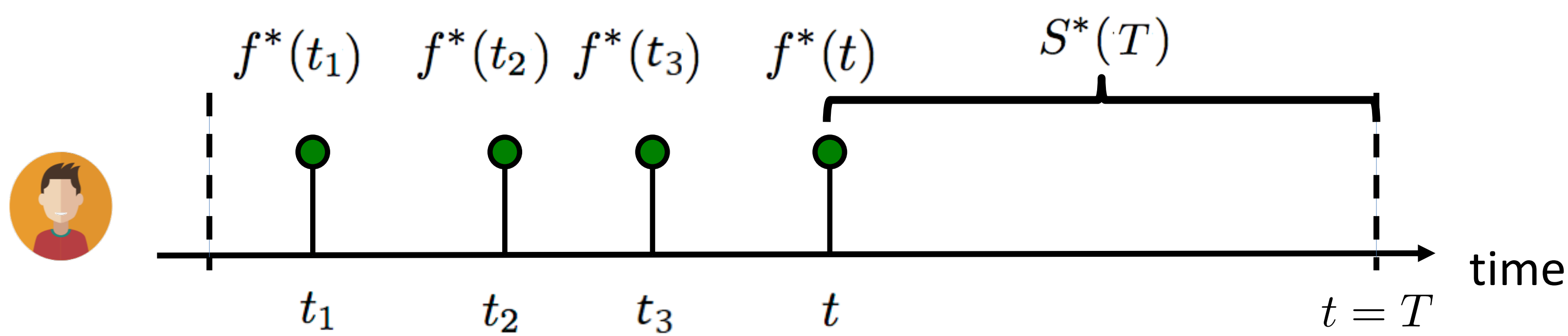
Model time as a random variable





Likelihood of a timeline: $f^*(t_1)\ f^*(t_2)\ f^*(t_3)\ f^*(t)\ S^*(T)$

Problems of density parametrization (I)



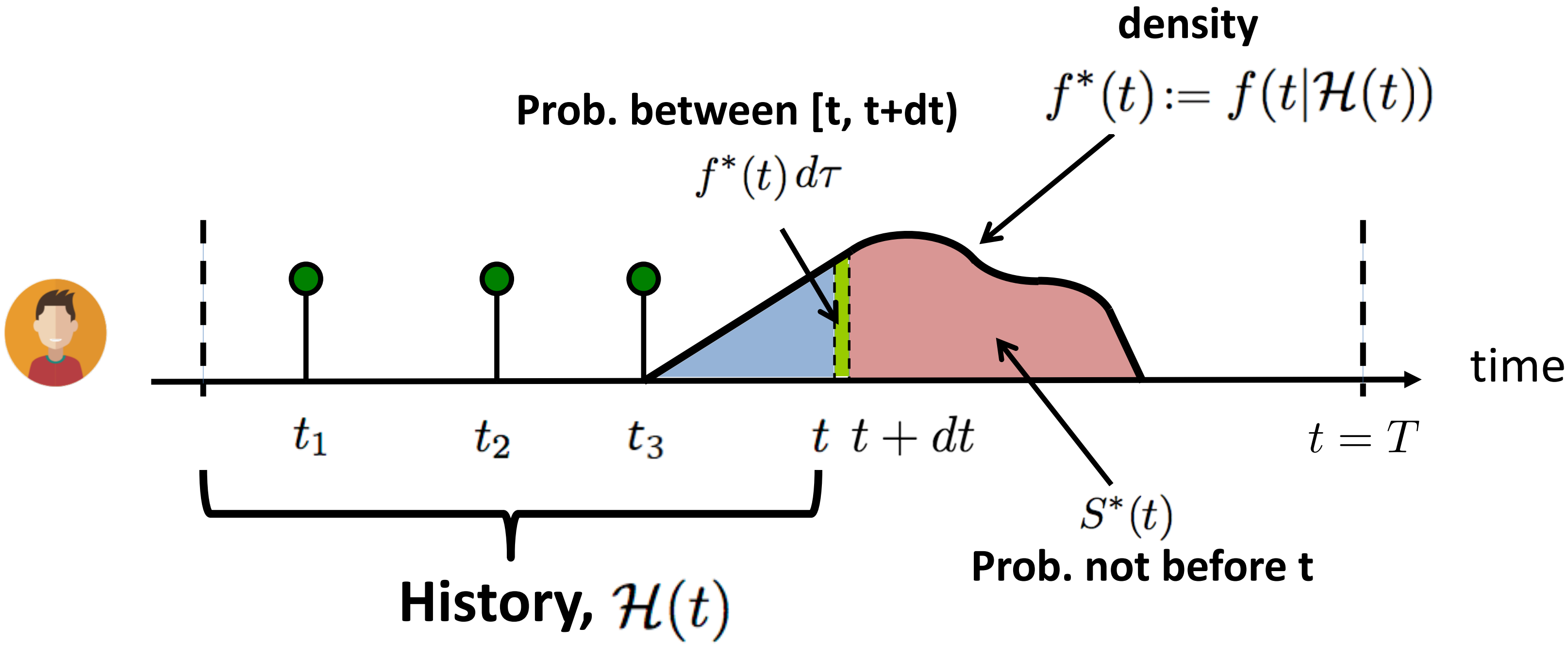
$$f^*(t_1) \quad f^*(t_2) \quad f^*(t_3) \quad f^*(t) \quad S^*(T)$$

$$\frac{\exp\langle w, \psi^*(t_1)\rangle}{Z} \quad \frac{\exp\langle w, \psi^*(t_2)\rangle}{Z} \quad \frac{\exp\langle w, \psi^*(t_3)\rangle}{Z} \quad \frac{\exp\langle w, \psi^*(t)\rangle}{Z} \quad 1 - \int_t^T \frac{\exp\langle w, \psi^*(\tau)\rangle}{Z} d\tau$$

It is difficult for model design and interpretability:

1. **Densities need to integrate to 1 (i.e., partition function)**
2. **Difficult to combine timelines**

Intensity function



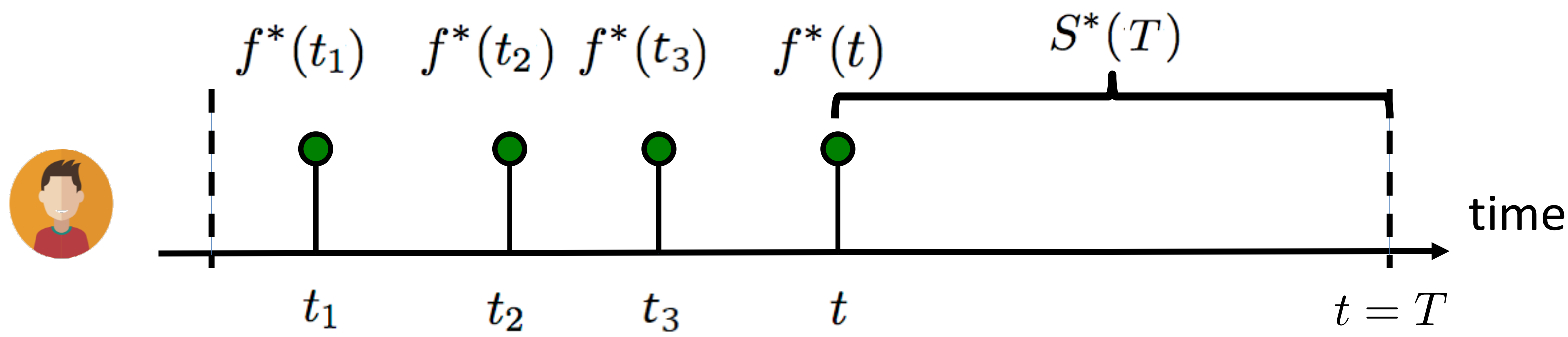
Intensity:

Probability between [t, t+dt) but not before t

$$\lambda^*(t)dt = \frac{f^*(t)dt}{S^*(t)} \geq 0 \quad \Rightarrow \quad \lambda^*(t)dt = \mathbb{E}[dN(t)|\mathcal{H}(t)]$$

Observation: $\lambda^*(t)$ **It is a rate = # of events / unit of time**

Advantages of intensity parametrization (I)



$$\lambda^*(t_1)\ \lambda^*(t_2)\ \lambda^*(t_3)\ \lambda^*(t)\ \exp\left(-\int_0^T \lambda^*(\tau)\, d\tau\right)$$

$\langle w, \phi^*(t_1)\rangle$

$\langle w, \phi^*(t_2)\rangle$

$\langle w, \phi^*(t_3)\rangle$

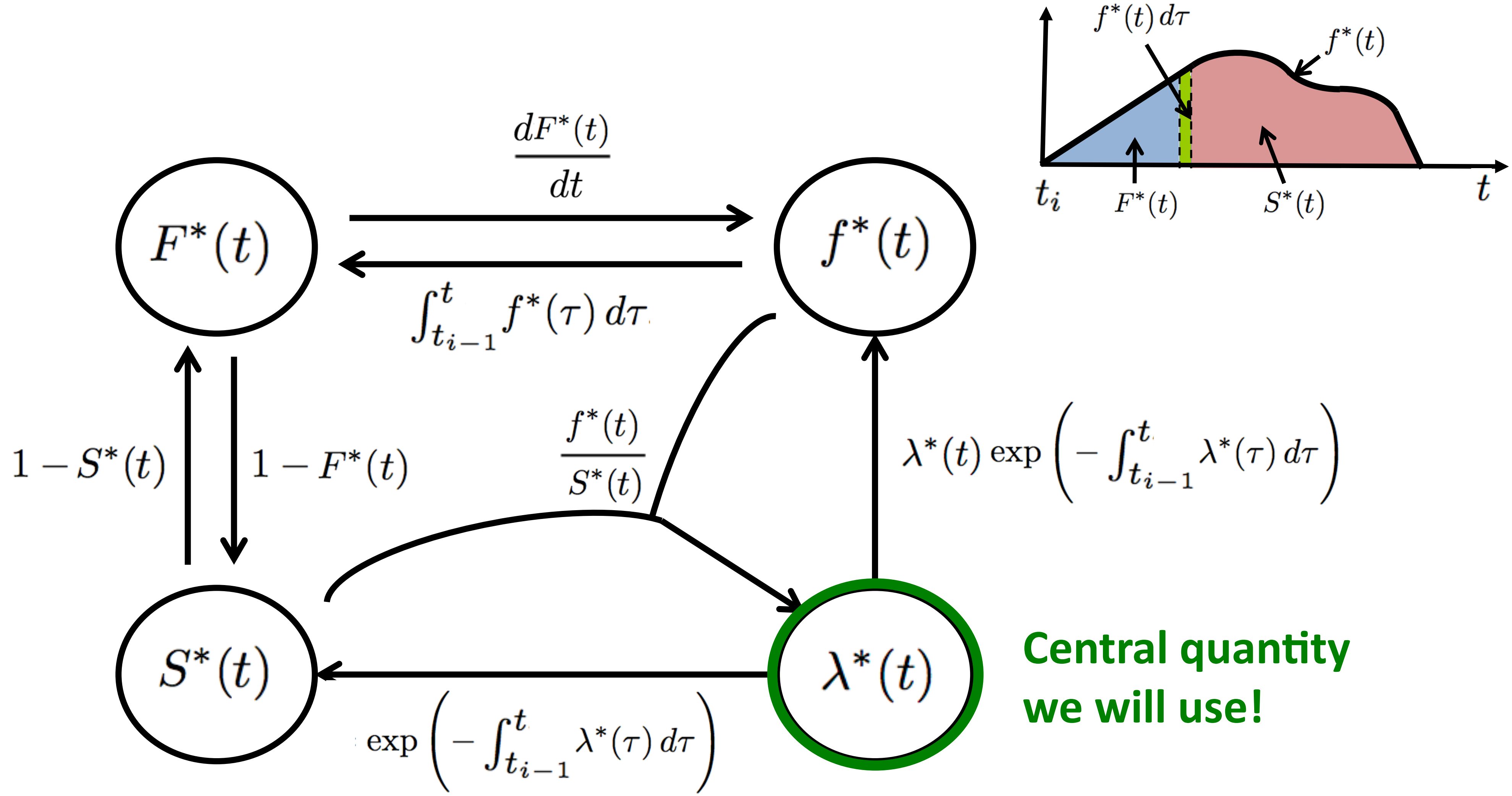
$\langle w, \phi^*(t)\rangle$

$\exp\left(-\int_0^T \langle w, \phi^*(\tau)\rangle d\tau\right)$

Suitable for model design and interpretable:

1. **Intensities only need to be nonnegative**
2. **Easy to combine timelines**

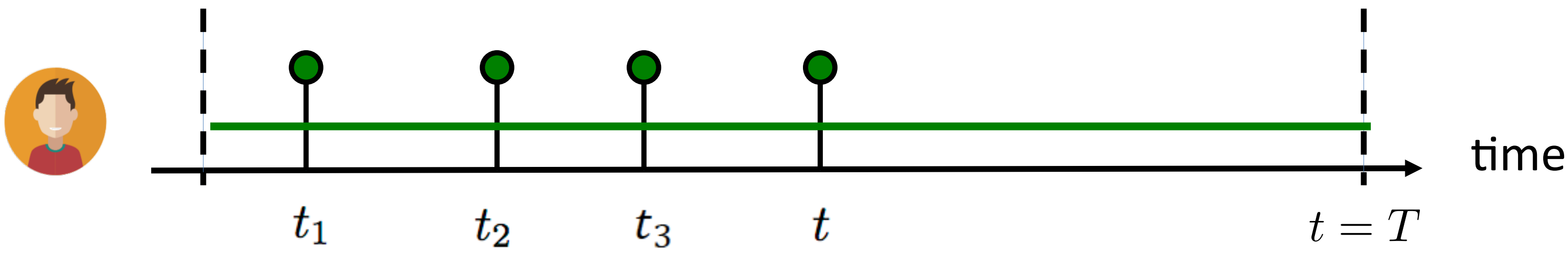
Relation between f*, F*, S*, λ*



Representation:
Temporal Point Processes

1. Intensity function
2. Basic building blocks
3. Superposition
4. Marks and SDEs with jumps

Poisson process



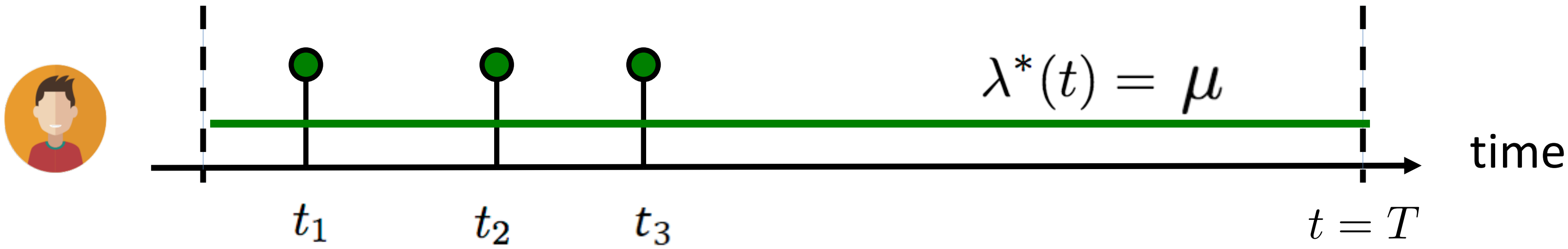
Intensity of a Poisson process

$$\lambda^*(t) = \mu$$

Observations:

1. **Intensity independent of history**
2. **Uniformly random occurrence**
3. **Time interval follows exponential distribution**

Fitting & sampling from a Poisson



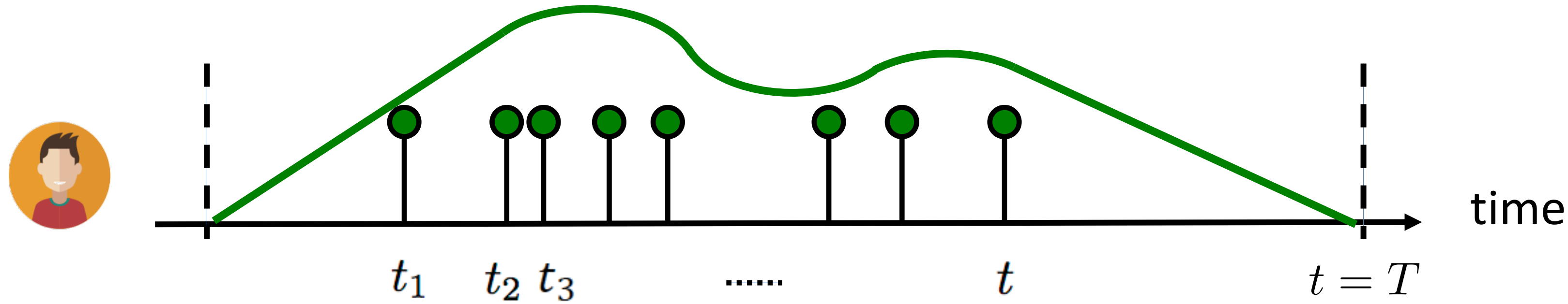
Fitting by maximum likelihood:

$$\mu^* = \underset{\mu}{\operatorname{argmax}} \ 3 \log \mu - \mu T = \frac{3}{T}$$

Sampling using inversion sampling:

$$\underbrace{t \sim \mu \exp(-\mu(t - t_3))}_{f_t^*(t)} \Rightarrow t = \underbrace{-\frac{1}{\mu} \log(1 - \overset{Uniform(0,1)}{\overset{\downarrow}{u}}) + t_3}_{F_t^{-1}(u)}$$

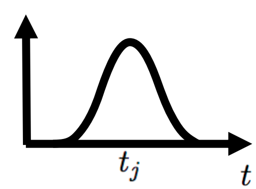
Inhomogeneous Poisson process



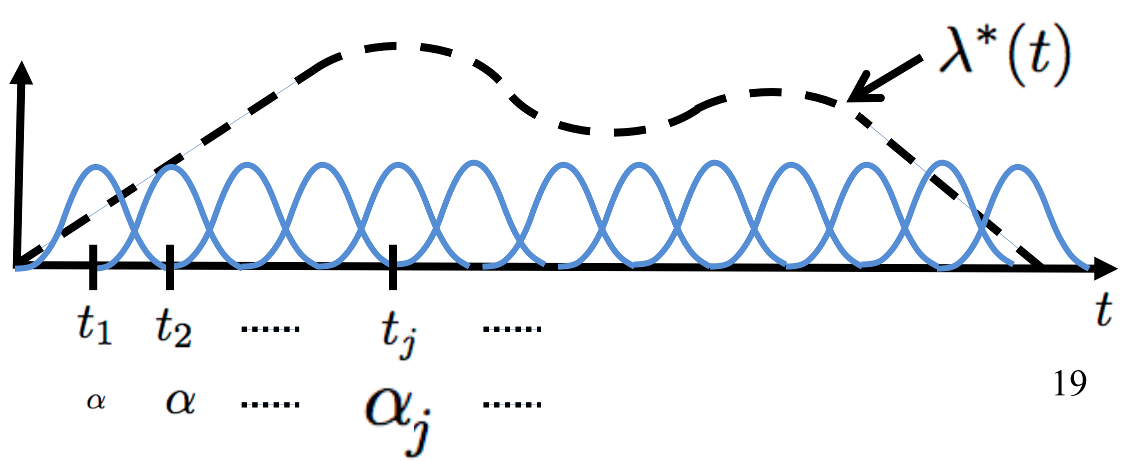
Intensity of an inhomogeneous Poisson process

$$\lambda^*(t) = g(t) \geqslant 0$$ **(Independent of history)**

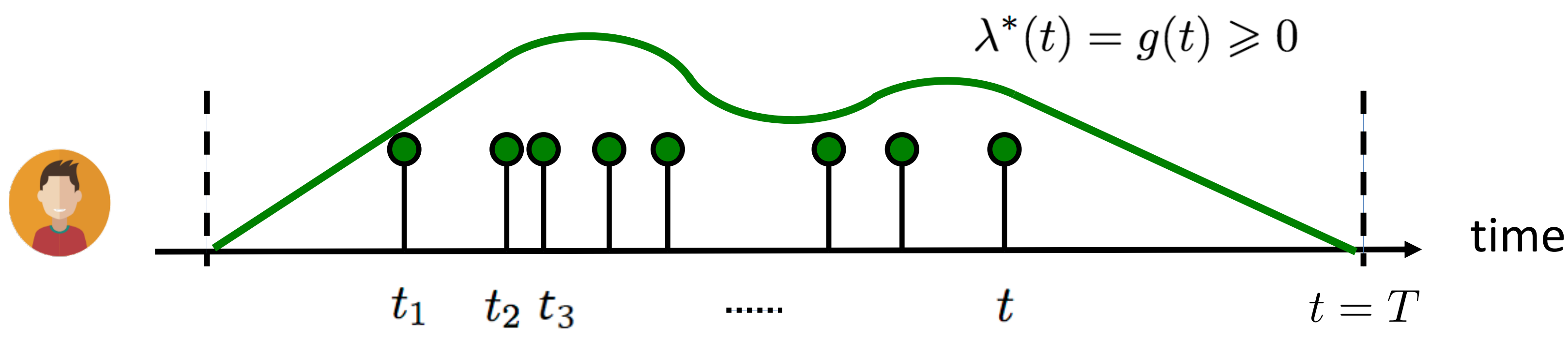
Example:



$$\lambda^*(t) = \sum_j \alpha_j k(t - t_j)$$



Fitting & sampling from inhomogeneous Poisson



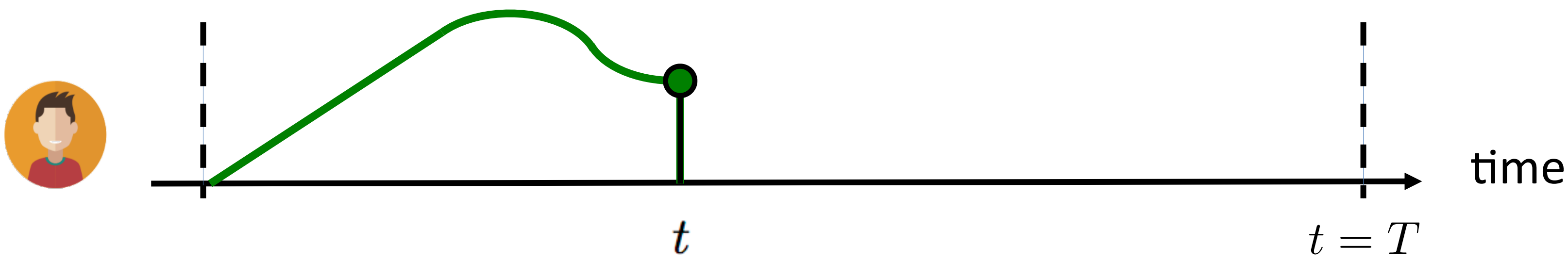
Fitting by maximum likelihood: $\underset{g(t)}{\text{maximize}} \sum_{i=1}^{n} \log g(t_i) - \int_0^T g(\tau)\, d\tau$

Sampling using thinning (reject. sampling) + inverse sampling:

1. **Sample** t **from Poisson process with intensity** μ **using inverse sampling**
2. **Generate** $u_2 \sim Uniform(0,1)$
3. **Keep the sample if** $u_2 \leq g(t) / \mu$

Keep sample with prob. $g(t)/\mu$

Terminating (or survival) process



Intensity of a terminating (or survival) process

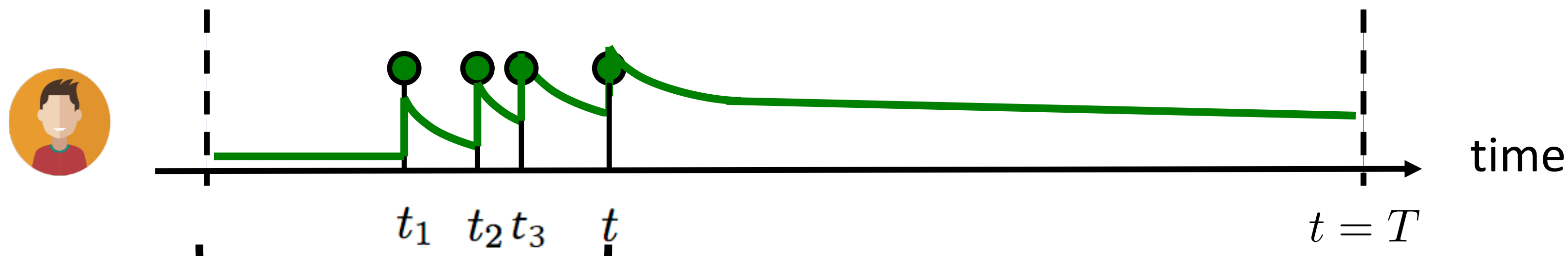
$$\lambda^*(t) = g^*(t)(1 - N(t)) \geqslant 0$$

Observations:

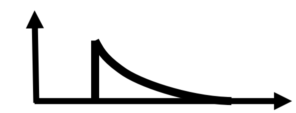
1. **Limited number of occurrences**

Try sampling and fitting!

Self-exciting (or Hawkes) process



History, $\mathcal{H}(t)$



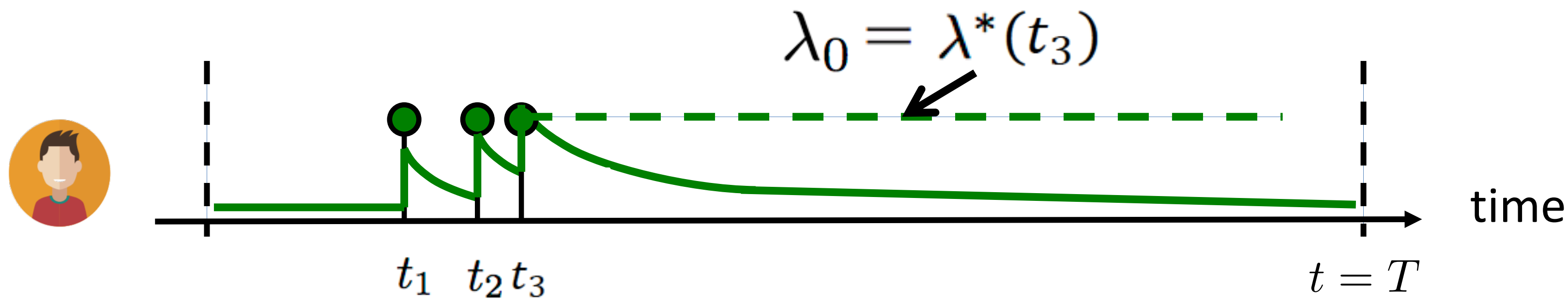
Intensity of self-exciting (or Hawkes) process:

$$\lambda^*(t) = \mu + \alpha \sum_{t_i \in \mathcal{H}(t)} \overbrace{\kappa_\omega(t - t_i)}^{\text{Triggering kernel}}$$
$$= \mu + \alpha \kappa_\omega(t) \star dN(t)$$

Observations:

1. **Clustered (or bursty) occurrence of events**
2. **Intensity is stochastic and history dependent**

Fitting a Hawkes process from a recorded timeline



Fitting by maximum likelihood:

$$\underset{\mu, \alpha}{\text{maximize}} \sum_{i=1}^{n} \log \lambda^*(t_i) - \int_0^T \lambda^*(\tau)\, d\tau$$

The max. likelihood is jointly convex in μ and α

Sampling using thinning (reject. sampling) + inverse sampling:

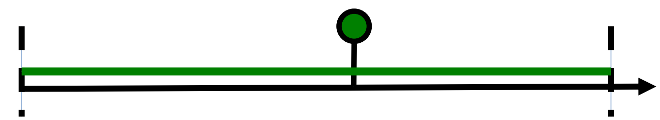
Key idea: the maximum of the intensity λ_0 changes over time

Summary

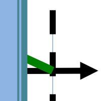
Building blocks to represent **different dynamic processes:**

Poisson processes:

$$\lambda^*(t) = \lambda$$



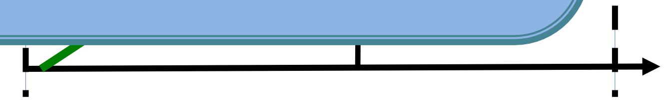
Inho



We know **how to fit** them
and **how to sample** from them

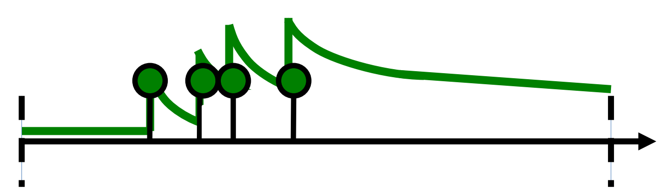
Term

$$\lambda^*(t) = g^*(t)(1 - N(t))$$



Self-exciting point processes:

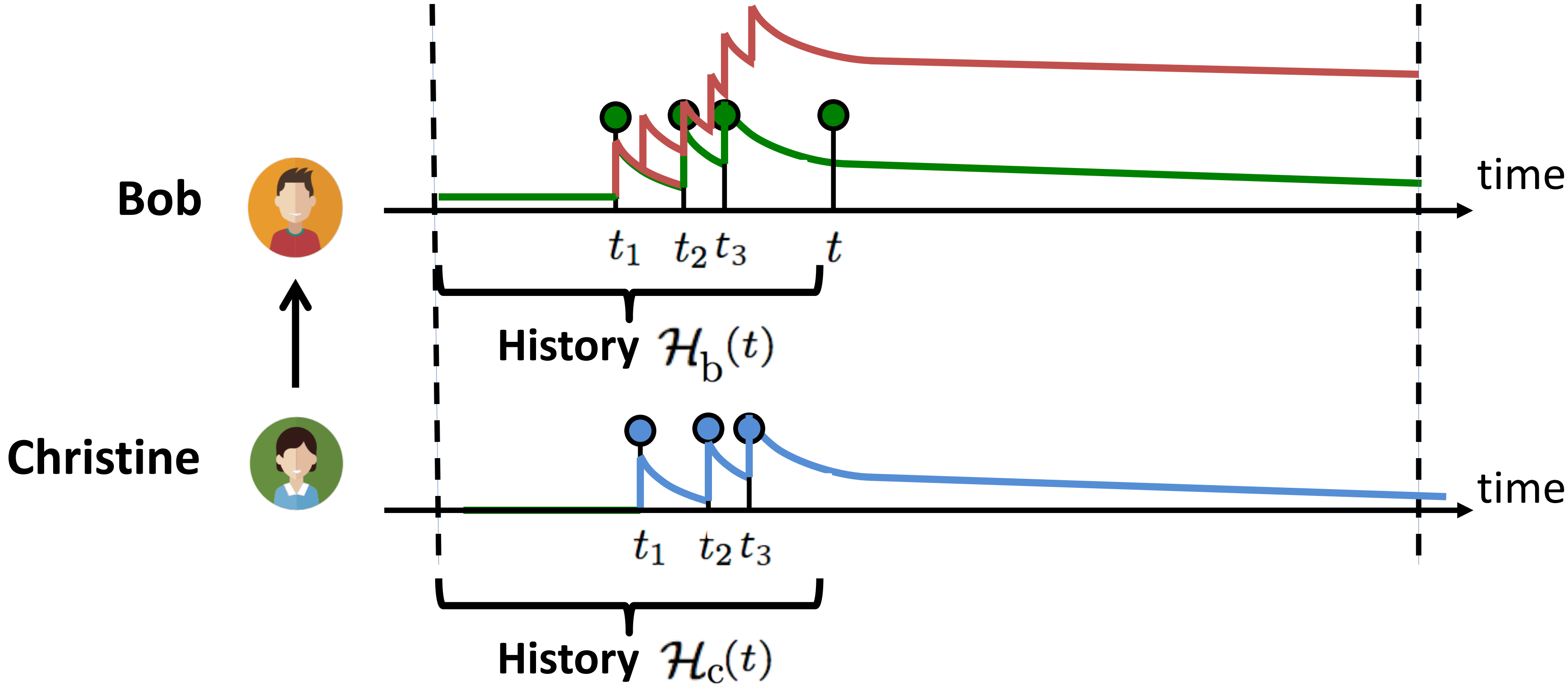
$$\lambda^*(t) = \mu + \alpha \sum_{t_i \in \mathcal{H}(t)} \kappa_\omega(t - t_i)$$



Representation:
Temporal Point Processes

1. Intensity function
2. Basic building blocks
3. **Superposition**
4. Marks and SDEs with jumps

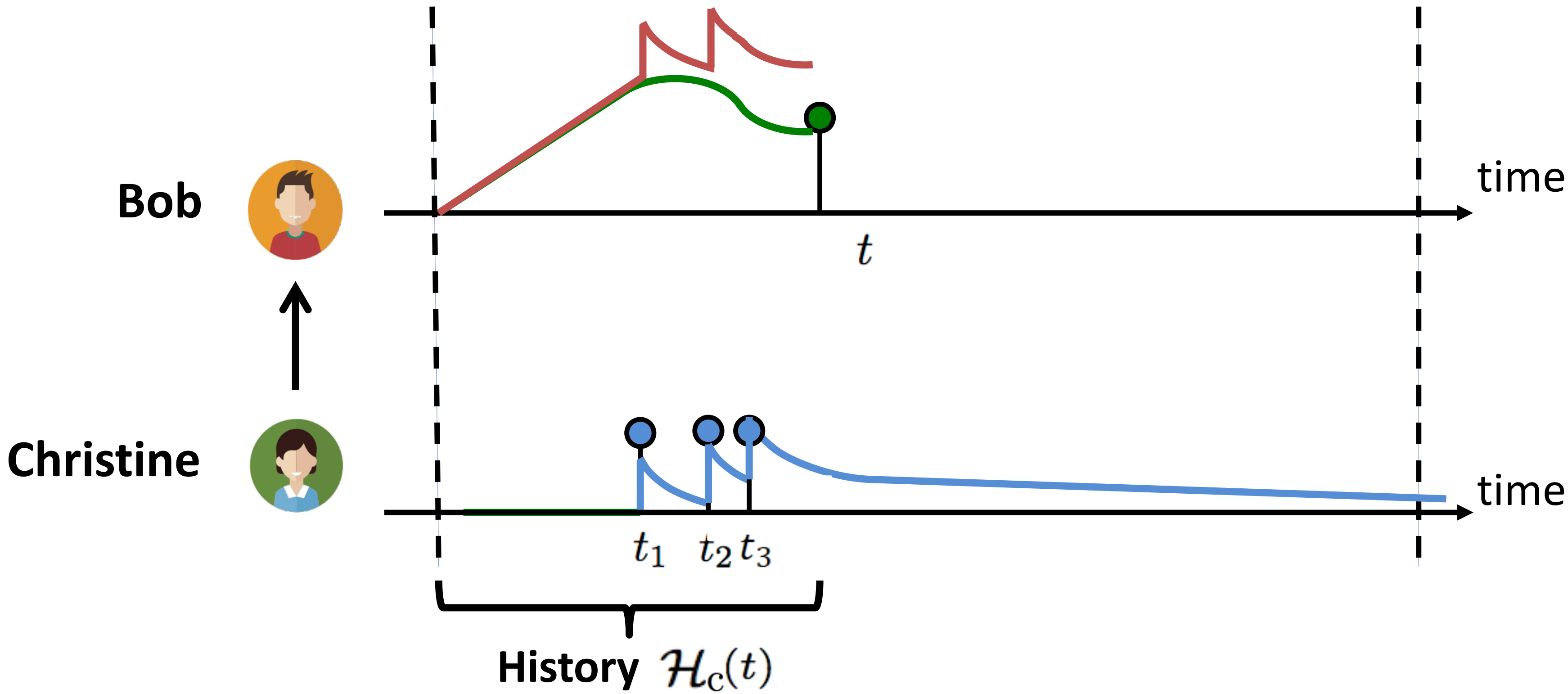
Mutually exciting process



Clustered occurrence affected by neighbors

$$\lambda^*(t) = \mu + \alpha \sum_{t_i \in \mathcal{H}_{\mathrm{b}}(t)} \kappa_\omega(t - t_i) + \beta \sum_{t_i \in \mathcal{H}_{\mathrm{c}}(t)} \kappa_\omega(t - t_i)$$

Mutually exciting terminating process



Clustered occurrence affected by neighbors

$$\lambda^*(t) = (1 - N(t))\Big(g(t) + \beta \sum_{t_i \in \mathcal{H}_c(t)} \kappa_\omega(t - t_i)\Big)$$

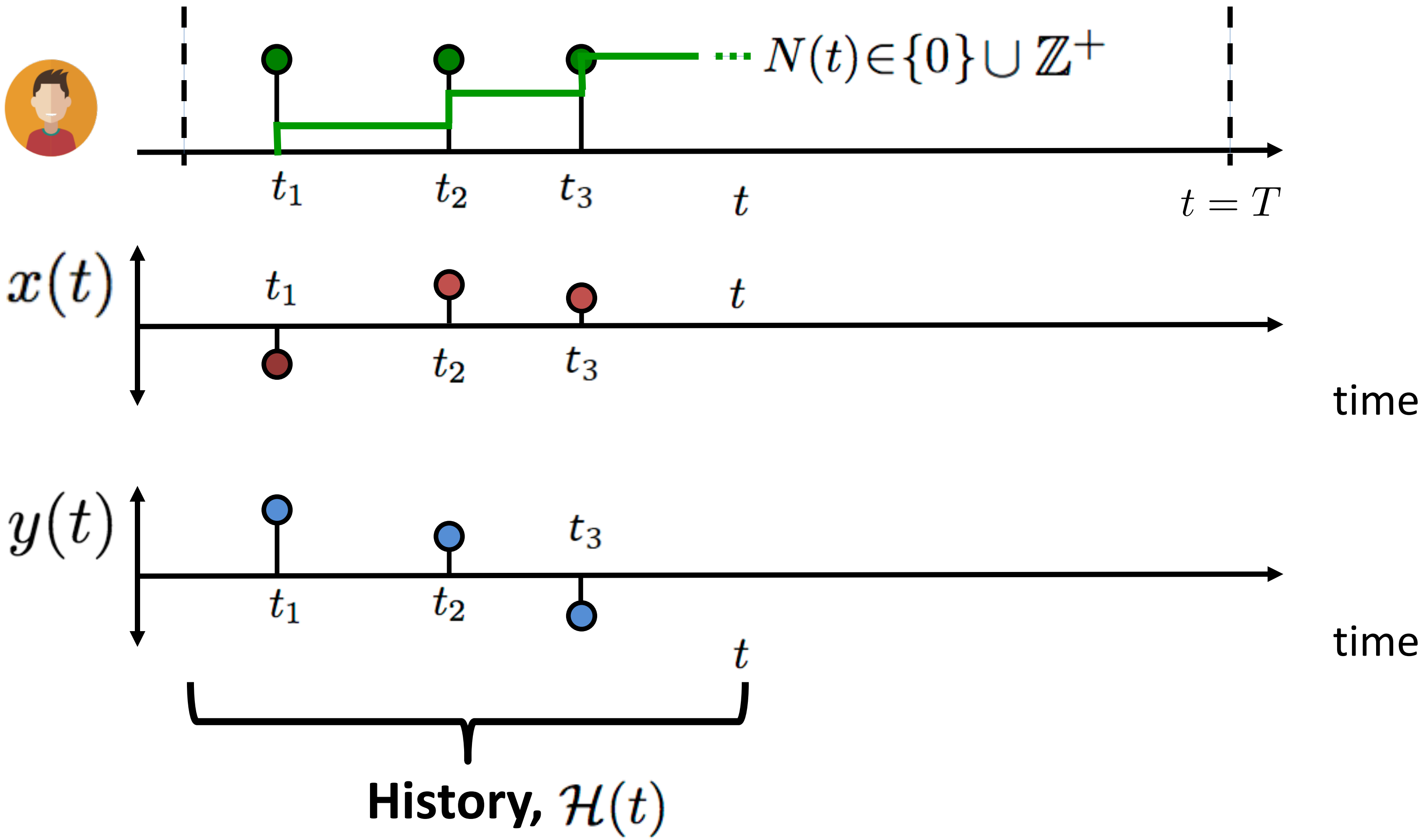
Representation:
Temporal Point Processes

1. Intensity function
2. Basic building blocks
3. Superposition
4. Marks and SDEs with jumps

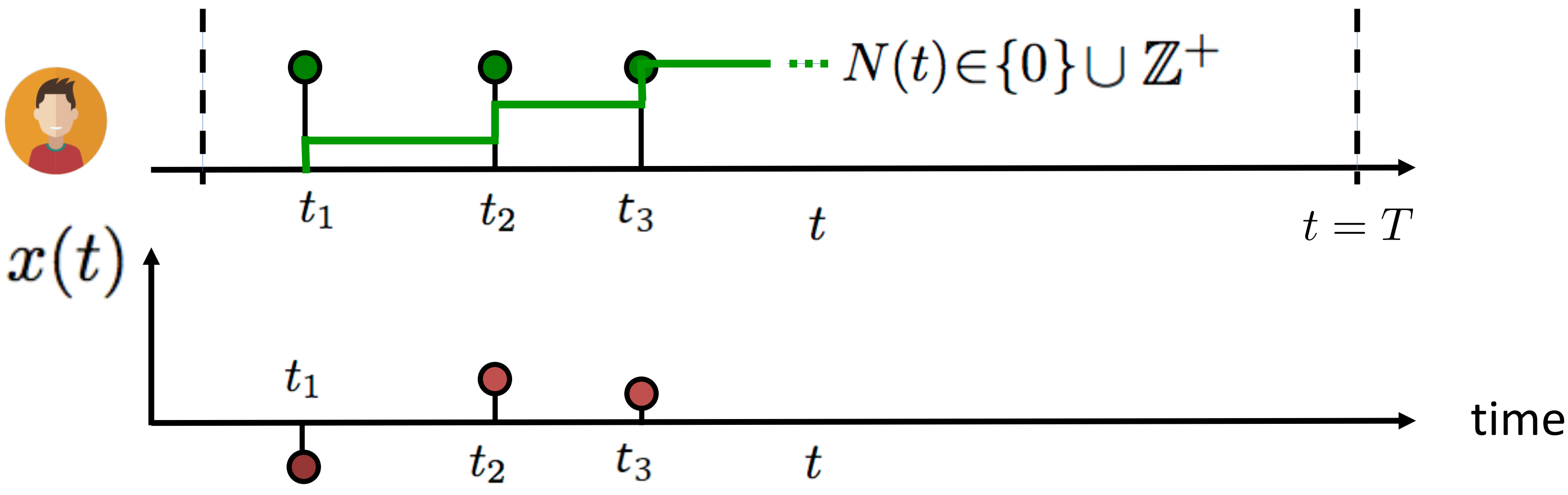
Marked temporal point processes

Marked temporal point process:

A random process whose realization consists of **discrete *marked* events localized in time**



Independent identically distributed marks



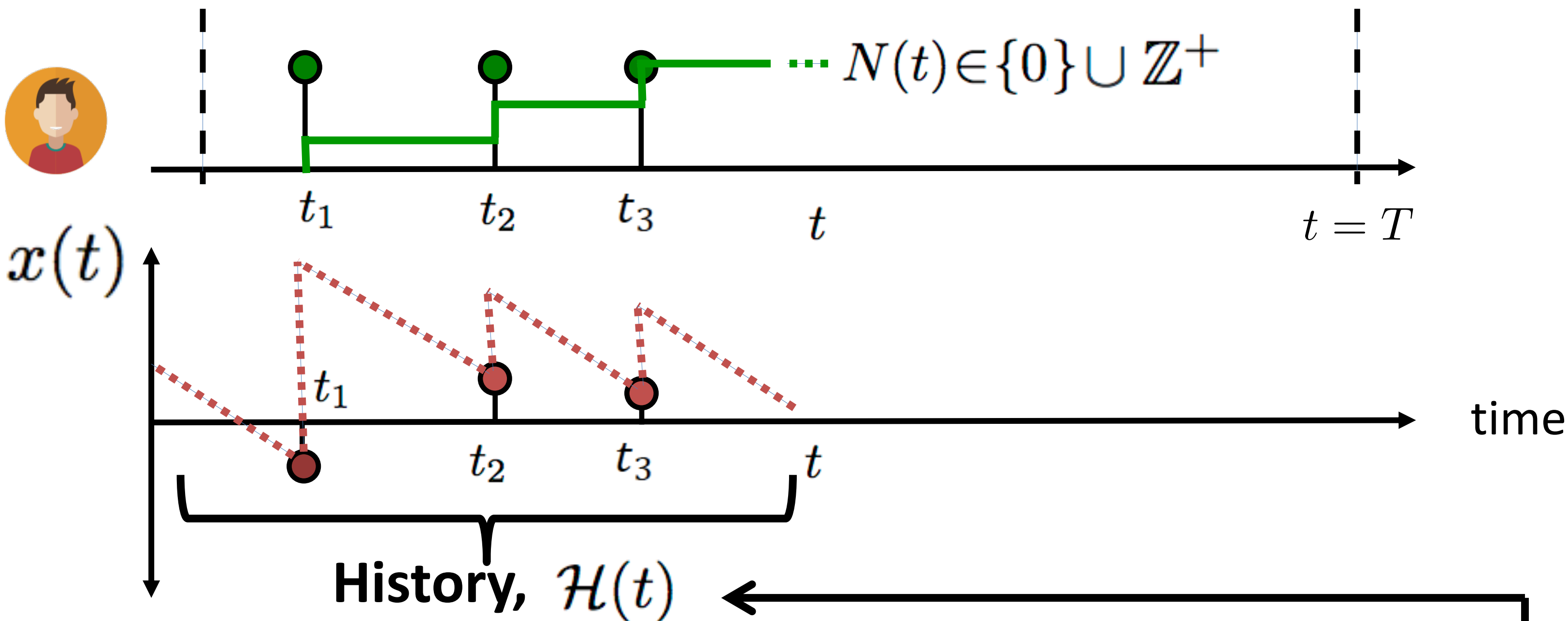
Distribution for the marks:

$$x^*(t_i) \sim p(x)$$

Observations:

1. **Marks independent of the temporal dynamics**
2. **Independent identically distributed (I.I.D.)**

Dependent marks: SDEs with jumps



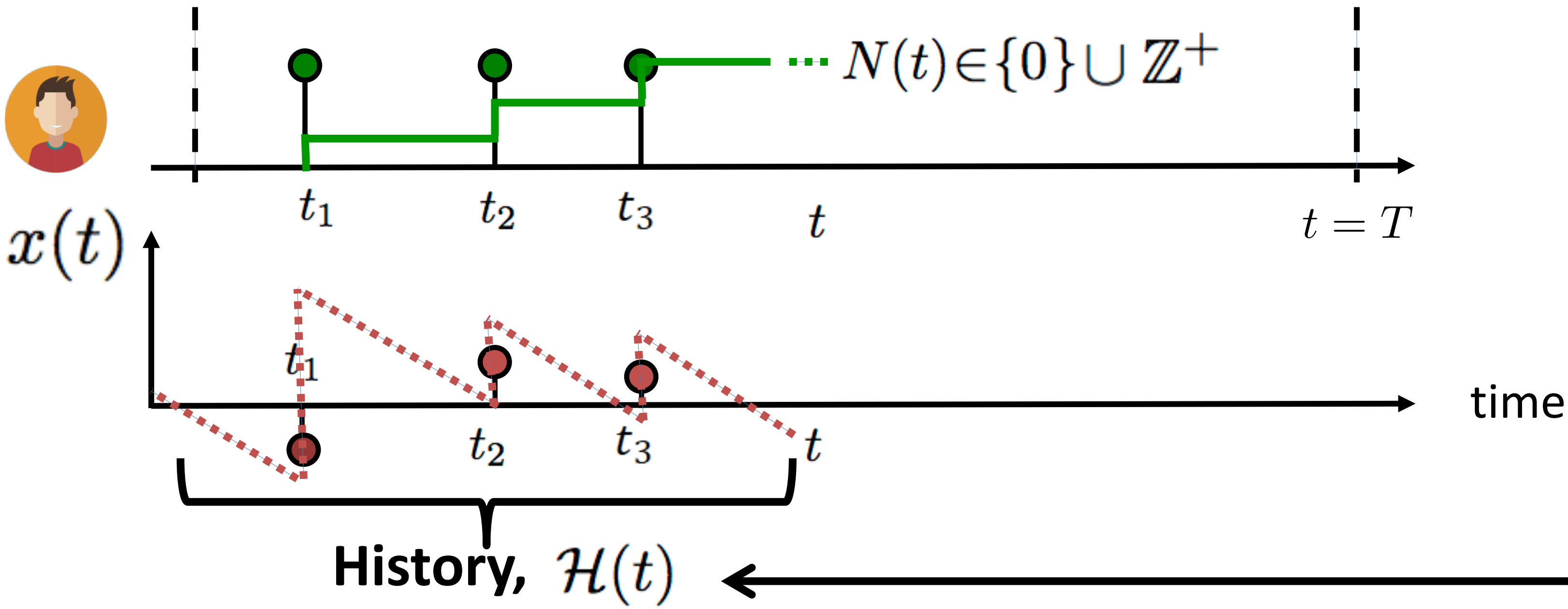
Marks given by stochastic differential equation with jumps:

$$x(t+dt) - x(t) = dx(t) = \underbrace{f(x(t),t)dt}_{\textbf{Drift}} + \underbrace{h(x(t),t)dN(t)}_{\textbf{Event influence}}$$

Observations:

1. **Marks dependent of the temporal dynamics**
2. **Defined for all values of t**

Dependent marks: distribution + SDE with jumps



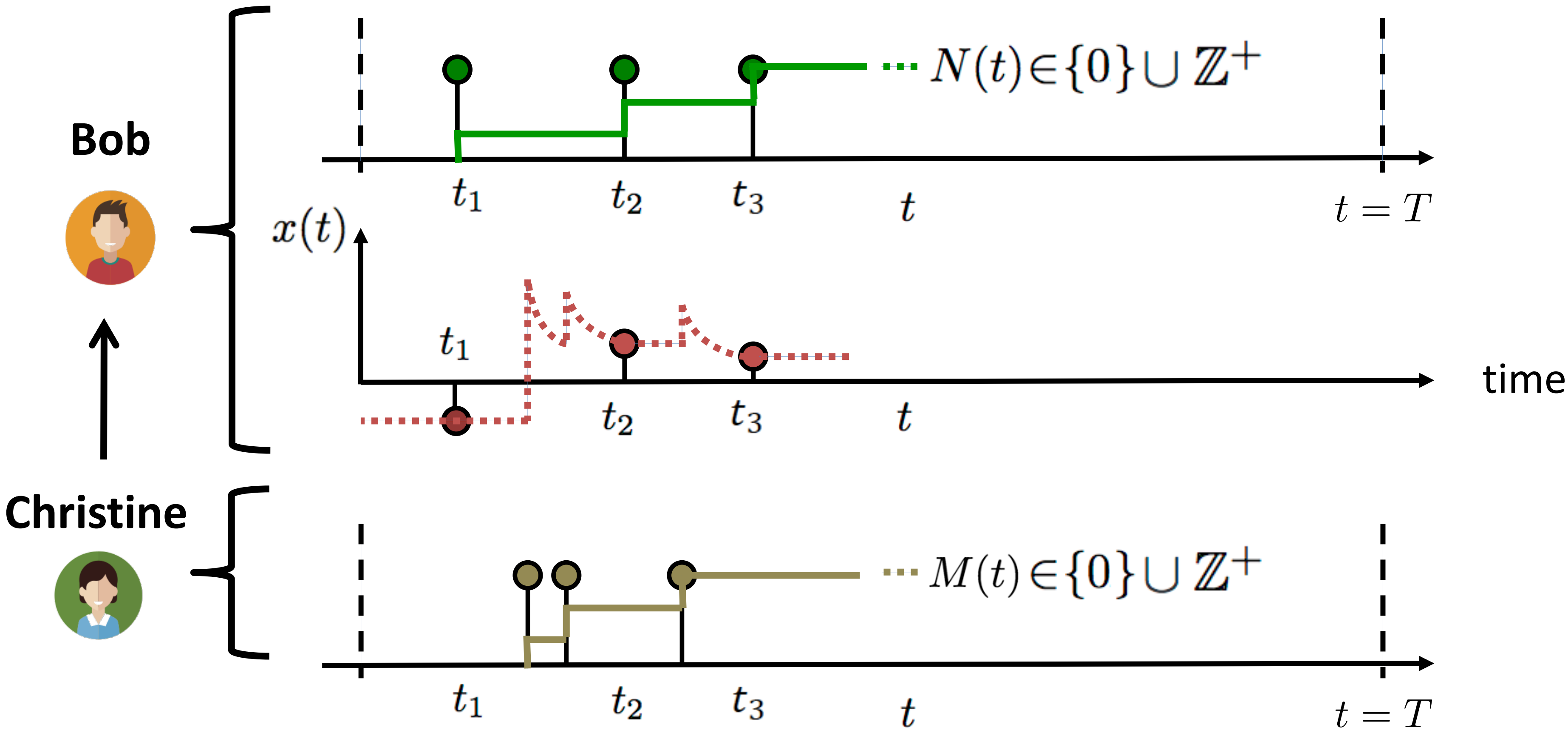
Distribution for the marks:

$$x^*(t_i) \sim p(x^* | x(t)) \Rightarrow dx(t) = \underbrace{f(x(t),t)dt}_{\text{Drift}} + \underbrace{h(x(t),t)dN(t)}_{\text{Event influence}}$$

Observations:

1. **Marks dependent on the temporal dynamics**
2. **Distribution represents additional source of uncertainty**

Mutually exciting + marks



Marks affected by neighbors

$$dx(t) = \underbrace{f(x(t), t)dt}_{\text{Drift}} + \underbrace{g(x(t), t)dM(t)}_{\text{Neighbor influence}}$$

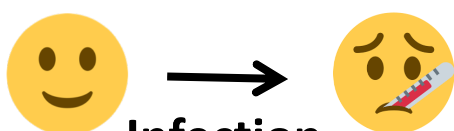
Marked TPPs as stochastic dynamical systems

Example: Susceptible-Infected-Susceptible (SIS)

$X_i(t) = 0$ **Susceptible** $\quad$ $X_i(t) = 1$ **Infected** $\quad$ $X_i(t) = 0$ **Susceptible**

SDE with jumps

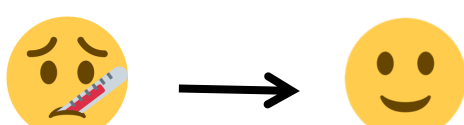
$$\overbrace{dX_i(t) = \underbrace{dY_i(t)}_{\text{It gets infected}} - \underbrace{dW_i(t)}_{\text{It recovers}}}^{\text{SDE with jumps}}$$



Infection rate

$$\mathbb{E}\left[dY_i(t)\right] = \lambda_{Y_i}(t)dt$$

$$\lambda_{Y_i}(t)dt = \overbrace{(1 - X_i(t))}^{\text{Node is susceptible}}\beta \underbrace{\textstyle\sum_{j \in \mathcal{N}(i)} X_j(t)}_{\text{If friends are infected, higher infection rate}} dt$$



Recovery rate

$$\mathbb{E}\left[dW_i(t)\right] = \lambda_{W_i}(t)dt$$

$$\overbrace{d\lambda_{W_i}(t) = \underbrace{\delta\, dY_i(t)}_{\text{Self-recovery rate when node gets infected}} \underbrace{- \lambda_{W_i}(t)dW_i(t)}_{\text{If node recovers, rate to zero}} + \underbrace{\rho\, dN_i(t)}_{\text{Rate increases if node gets treated}}}^{\text{SDE with jumps}}$$

Outline of the Seminar

Temporal Point processes (TPPs): Intro

1. Intensity function
2. Basic building blocks
3. Superposition
4. Marks and SDEs with jumps

This section

Models & Inference

1. Modeling event sequences
2. Clustering event sequences
3. Capturing complex dynamics
4. Causal reasoning on event sequences

Next section

RL & Control

1. Marked TPPs: a new setting
2. Stochastic optimal control
3. Reinforcement learning

Slides/references: learning.mpi-sws.org/tpp-icml18