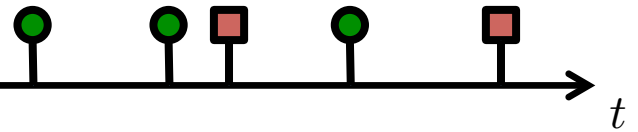


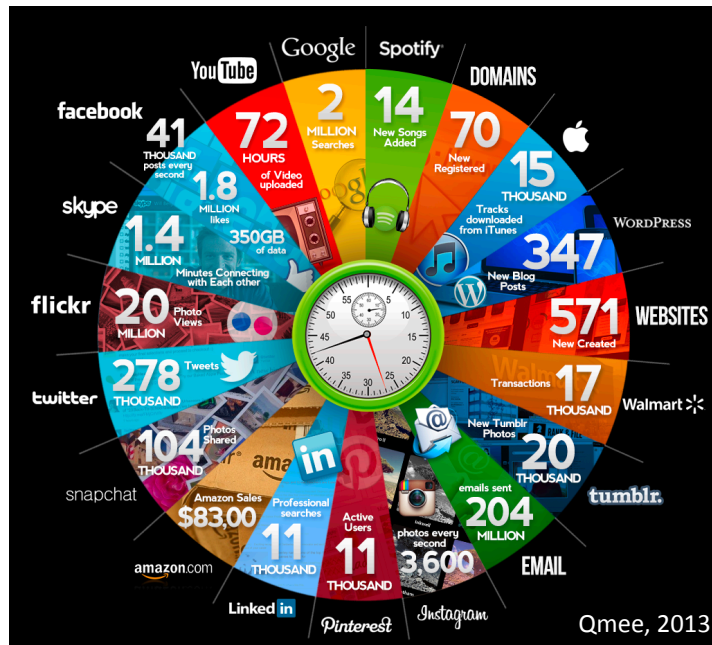
# Introduction and Applications of Temporal Point Processes



**Isabel Valera**

MPI for Intelligent Systems

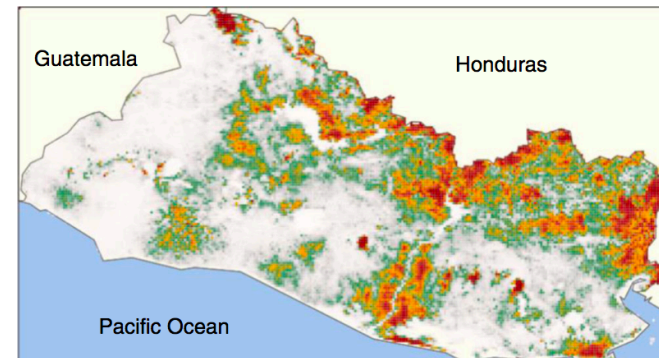
# Many discrete *events* in continuous time



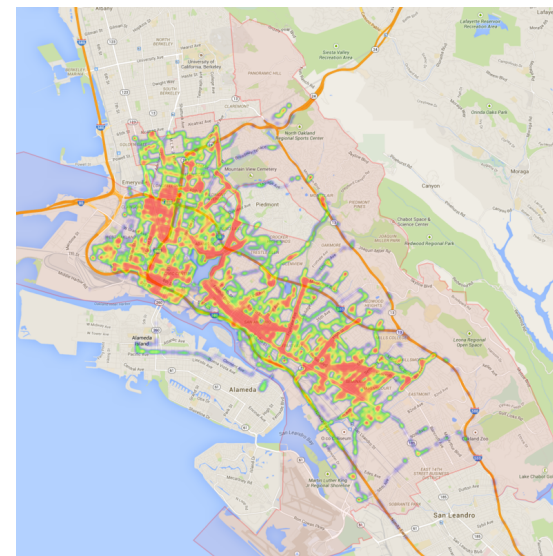
Online actions



Financial trading



Disease dynamics



Mobility dynamics

# Variety of processes behind these events

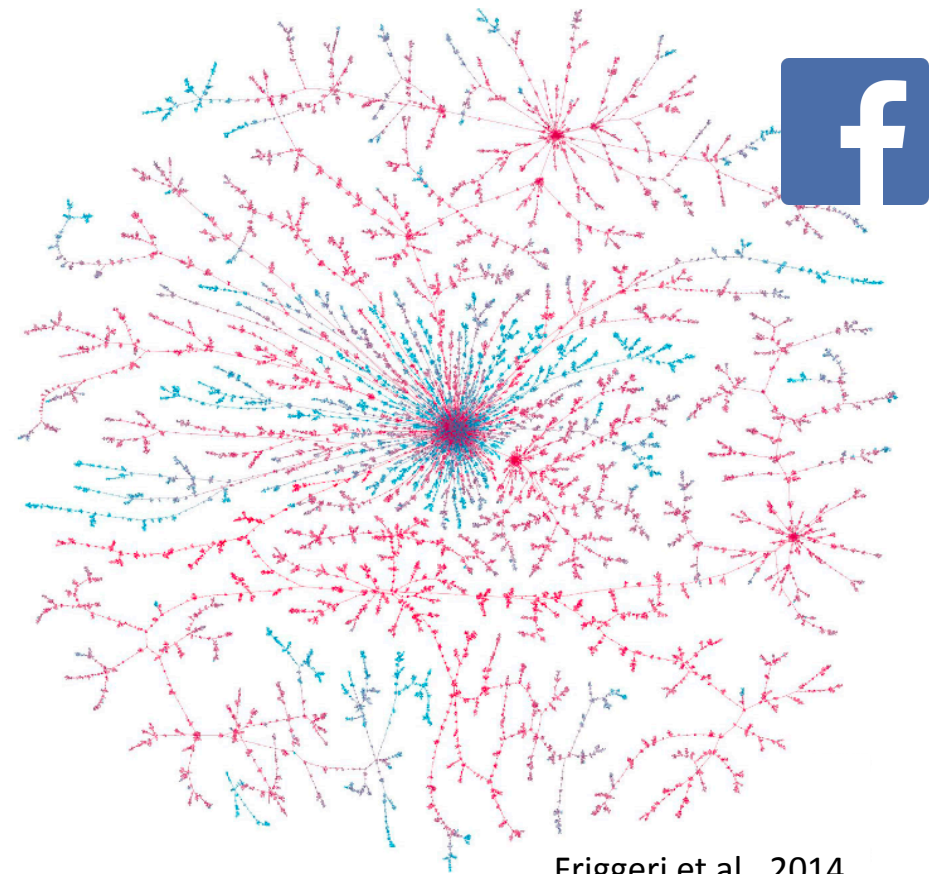
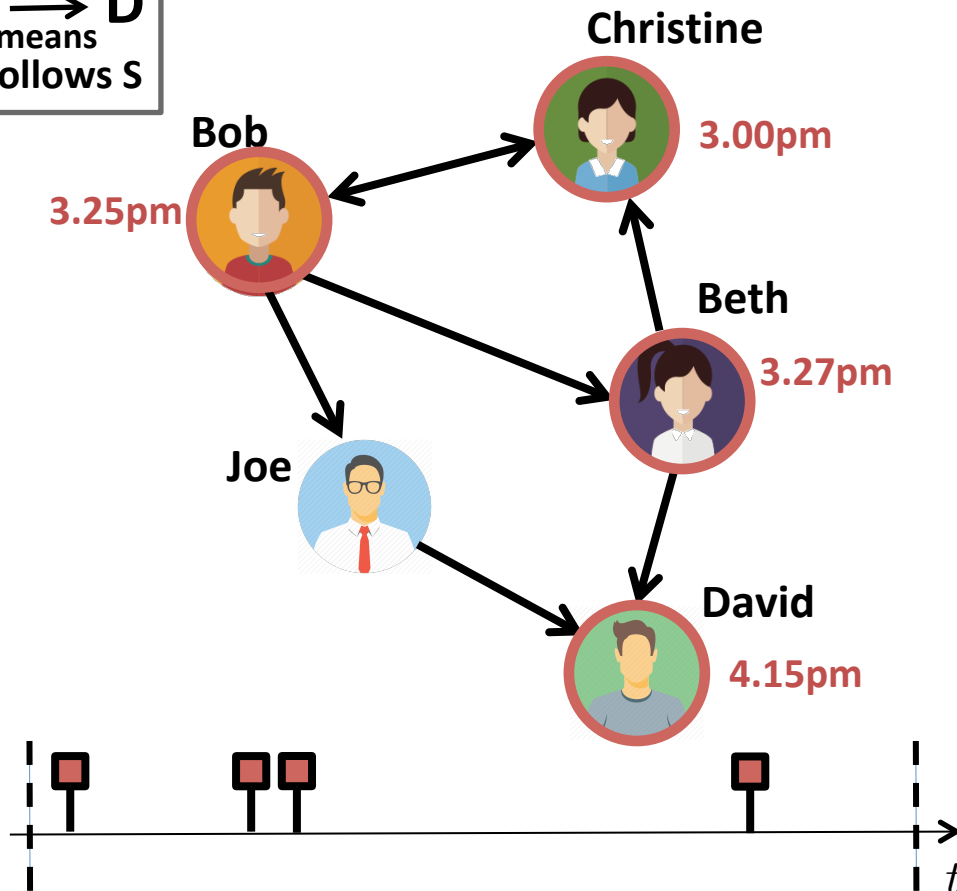
Events are (noisy) observations of a variety of complex dynamic processes...



...in a wide range of temporal scales. <sup>3</sup>

# Example I: Information propagation

S → D  
means  
D follows S



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



Barack Obama

From Wikipedia, the free encyclopedia

"Barack" and "Obama" redirect here. For his father, see Barack Obama Sr. For other uses of "Barack", see Barack (disambiguation). (disambiguation).

Barack Hussein Obama II (current President of the United States. He was president of the Harvard Law School, a civil rights attorney and taught at the University of Chicago, representing the 13th District of Columbia in the United States House of Representatives.

## Barack Obama: Revision history

03:41, 28 November 2016 Ranze (talk | contribs) .. (301,105 bytes) (+18) .. (E) (edit)

03:32, 28 November 2016 Xin Deui (talk | contribs) .. (301,087 bytes) (-68) .. (E) (edit)

00:57, 28 November 2016 SporkBot (talk | contribs) m .. (301,155 bytes) (-37) (edit)

07:03, 27 November 2016 Saiph121 (talk | contribs) .. (301,192 bytes) (+25) .. (E) (edit)

03:21, 20 September 2016

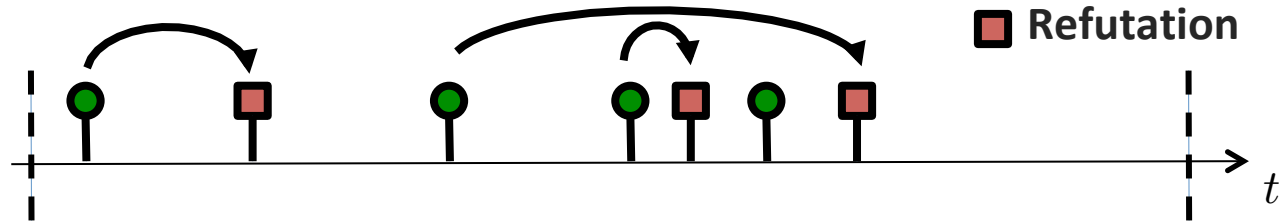
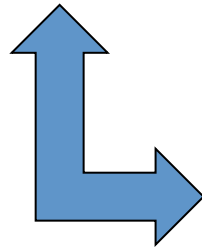
is a **Kenyan** politician



possible vandalism by MLM2016

is an American politician

● Addition  
■ Refutation



Moving to Australia Working in Australia Study abroad in Australia +4

## What are the pros and cons of living in Australia?

Answer Request Follow 109 Comment Share 9 Downvote

I have studied, worked and lived in Australia as an Intern employee, business owner and a citizen.

Upvote | 150

I have experienced this country in all the ways possible, you know. However, I firmly believe that there are definitely more pros than cons in Australia but still I have mentioned below a few cons.

Hope it helps! :)

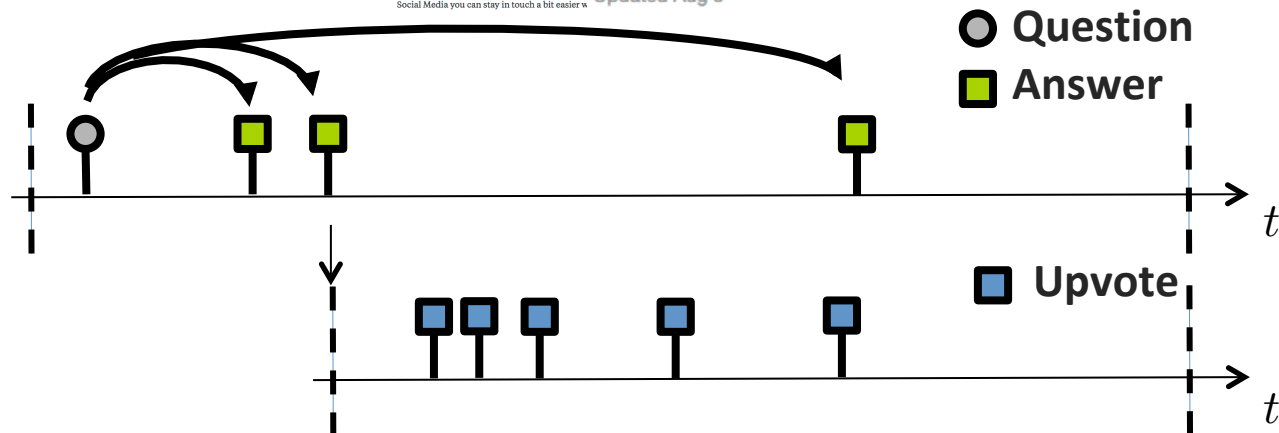
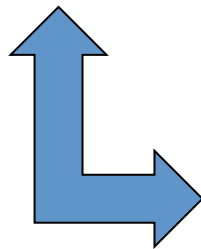
Possible Challenges

- Language problem for those who don't speak English
- Not having your family and friends around could be a challenge as society is more and more connected and thanks to social media you can stay in touch a bit easier with them.



M Sharma, Lived in Australia as Migrant, Student, Worker, Business Owner & Family Man

Updated Aug 3

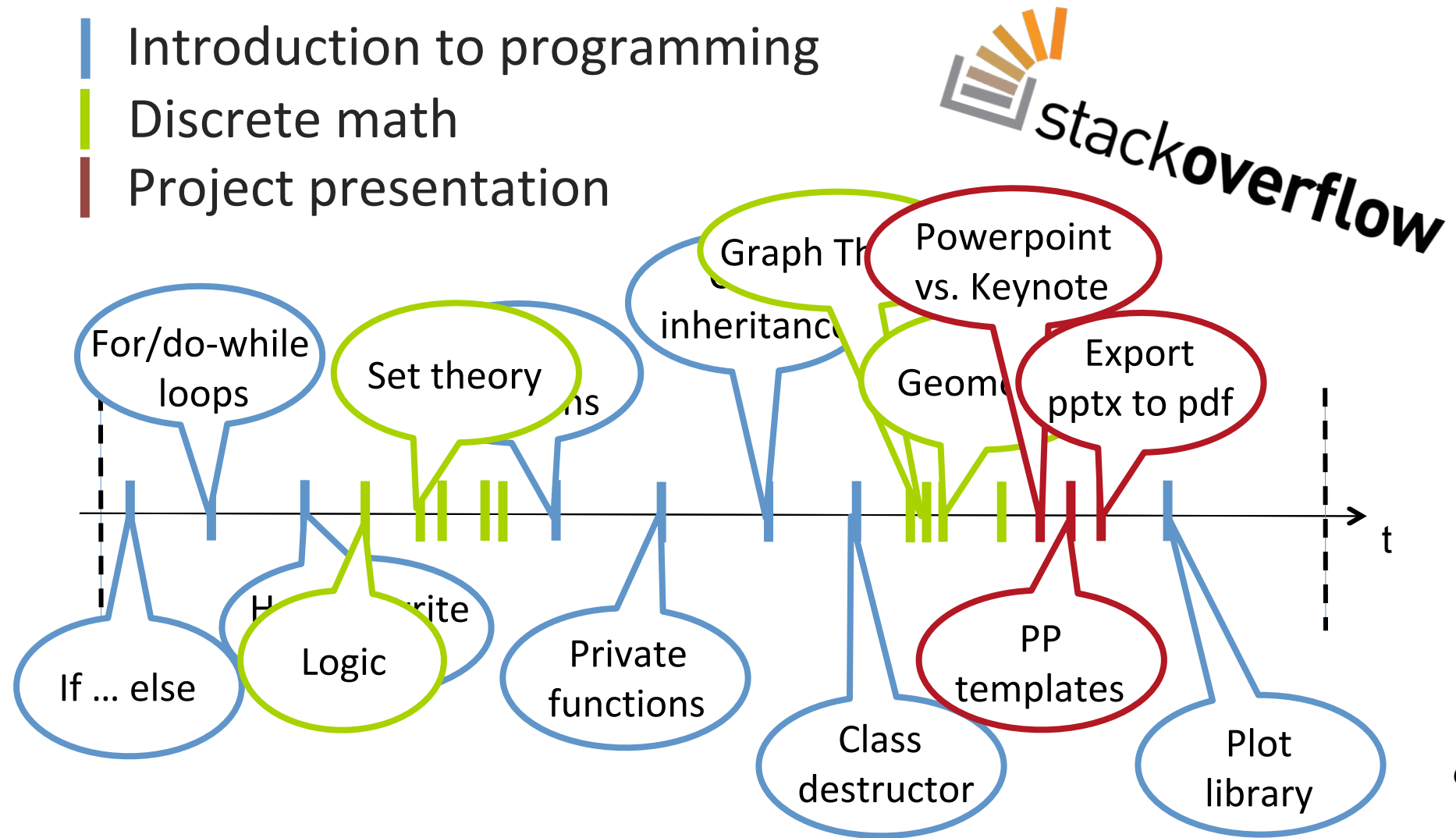


# Example III: Human learning

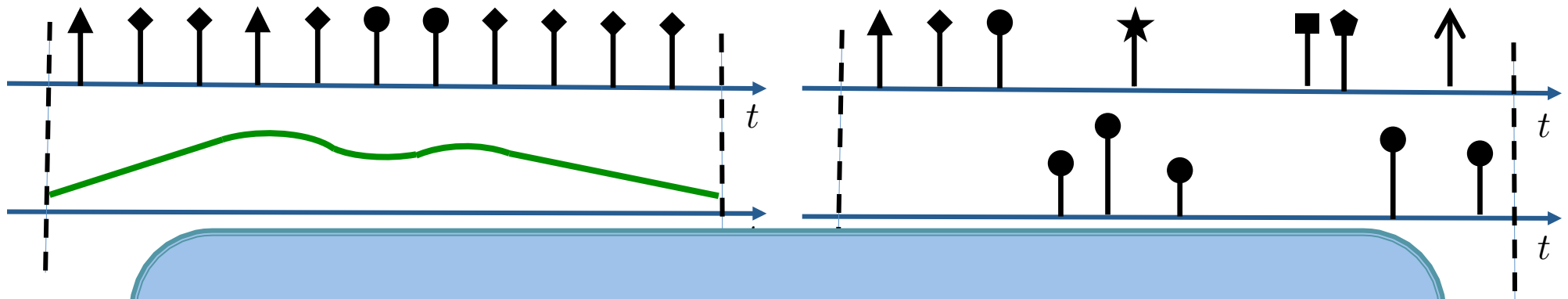


## 1st year computer science student

- Introduction to programming
- Discrete math
- Project presentation



# Aren't these event traces just time series?



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

Dis

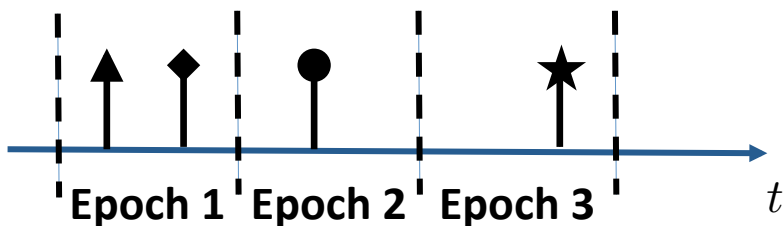
WI

epoch?

Events in epochs.

What if no event in one epoch?

What about time-related queries?



# Outline of the Lecture

## **INTRO TO TEMPORAL POINT PROCESSES (TPPs)**

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

## **APPLICATION: CLUSTERING EVENT SEQUENCES**

1. Problem Statement
2. Introduction to DPMM
3. CRP + HP (a.k.a. HDHP)
4. Generative process



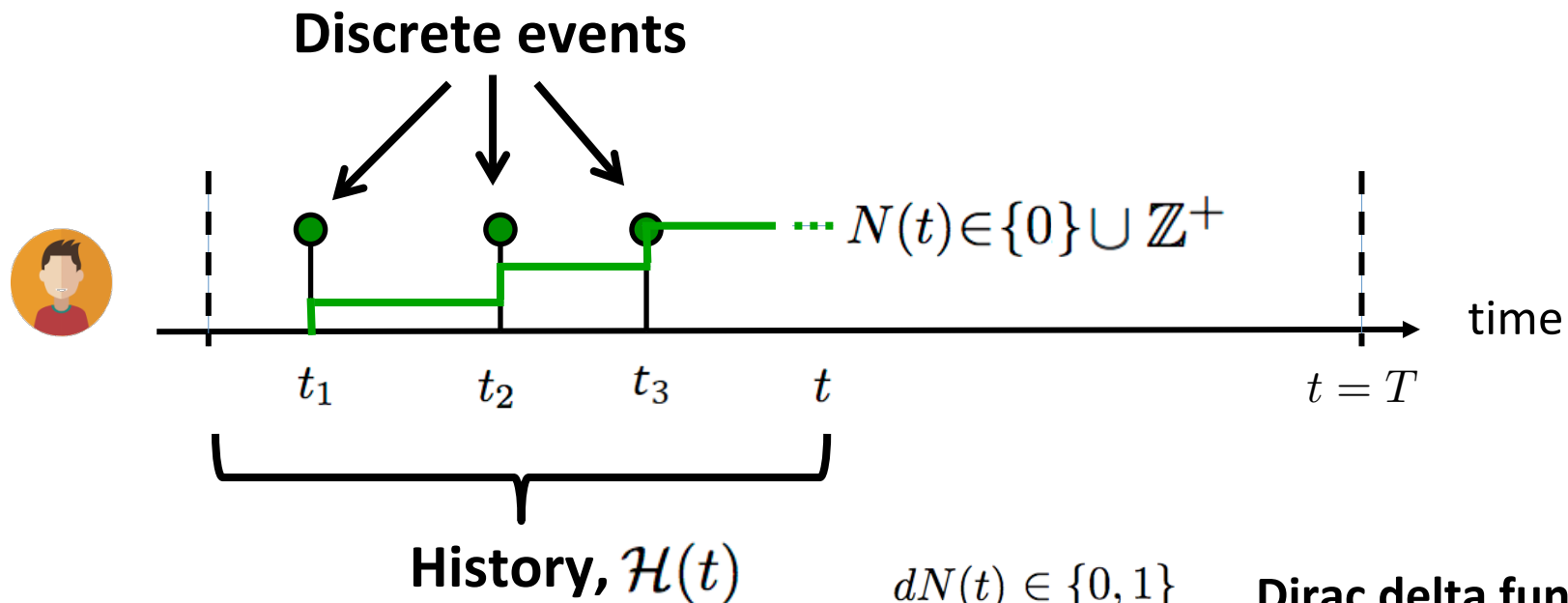
# Temporal Point Processes (TPPs): Introduction

- 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

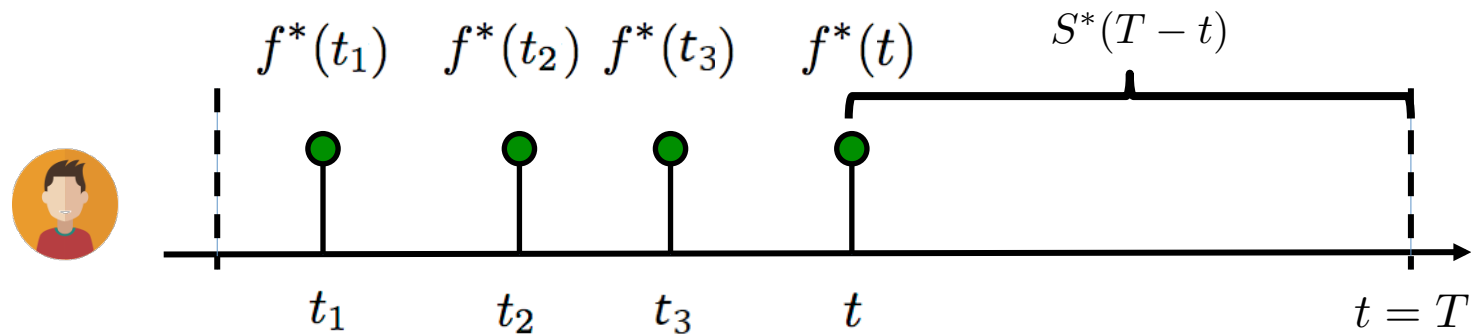
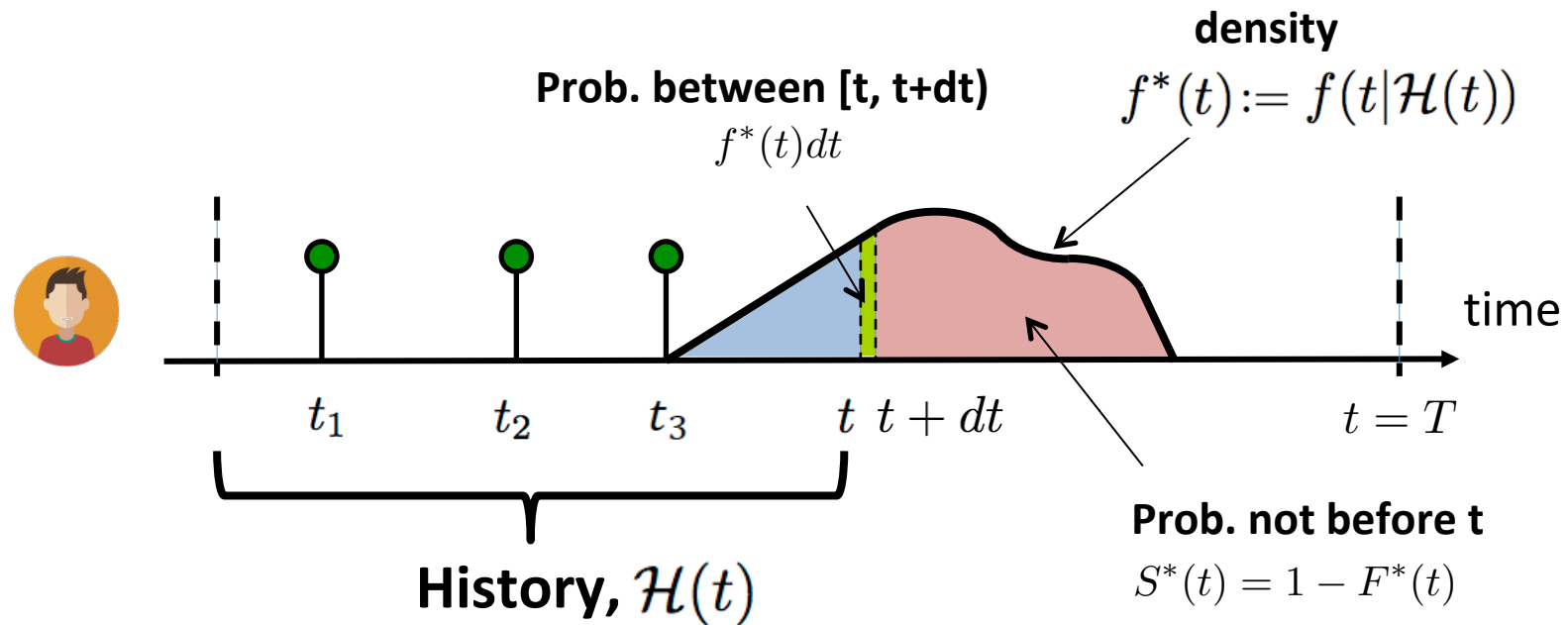


**Formally:**  $N(t) = \int_0^t dN(s) \Rightarrow dN(t) = \sum_{t_i \in \mathcal{H}(t)} \delta(t - t_i) dt$

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

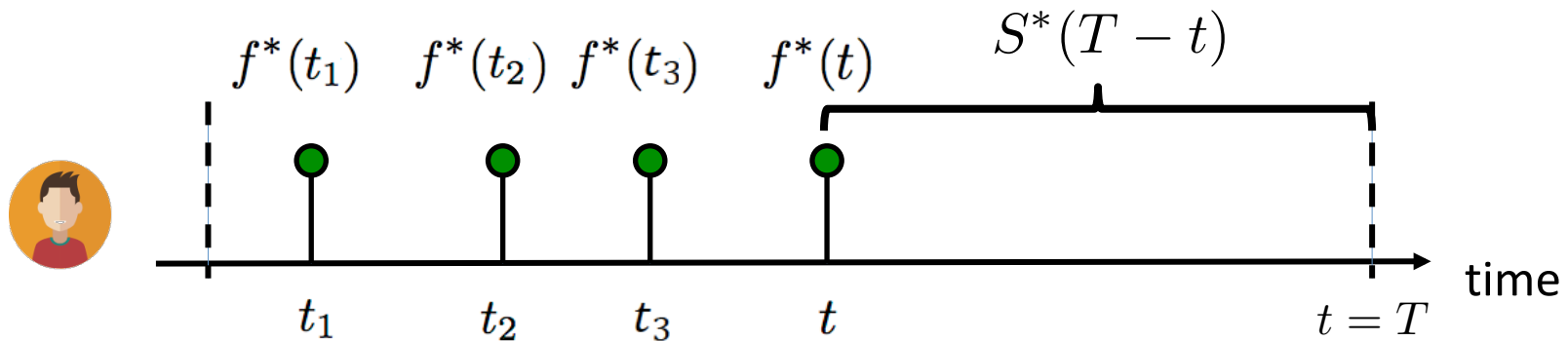
↓ ↓

# Model time as a random variable



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

# Problems of density parametrization

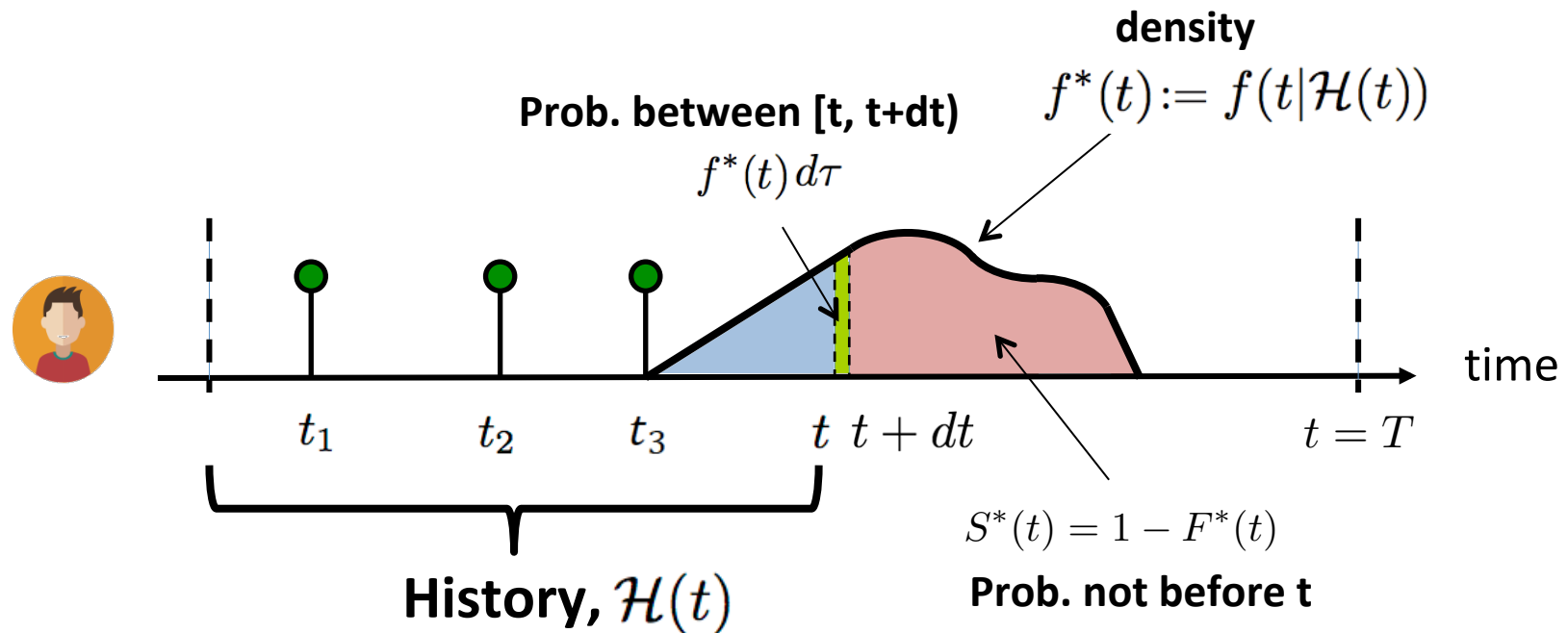


$$\begin{array}{cccccc}
 f^*(t_1) & f^*(t_2) & f^*(t_3) & f^*(t) & S^*(T) & \\
 \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \\
 \frac{\exp\langle w, \psi^*(t_1) \rangle}{Z} & & \frac{\exp\langle w, \psi^*(t_3) \rangle}{Z} & & & 1 - \int_t^T \frac{\exp\langle w, \psi^*(\tau) \rangle}{Z} d\tau \\
 & \frac{\exp\langle w, \psi^*(t_2) \rangle}{Z} & & \frac{\exp\langle w, \psi^*(t) \rangle}{Z} & & 
 \end{array}$$

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



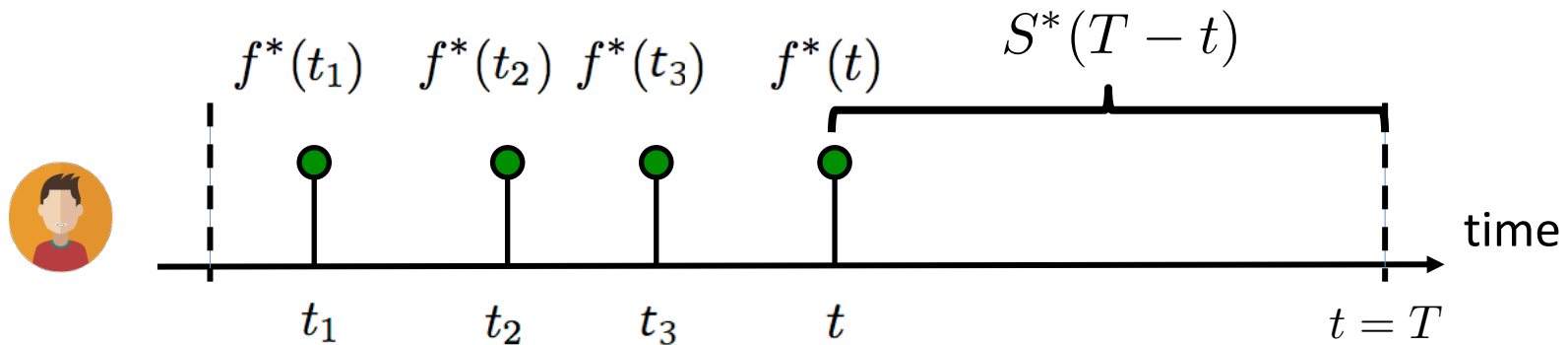
**Intensity:**

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)



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

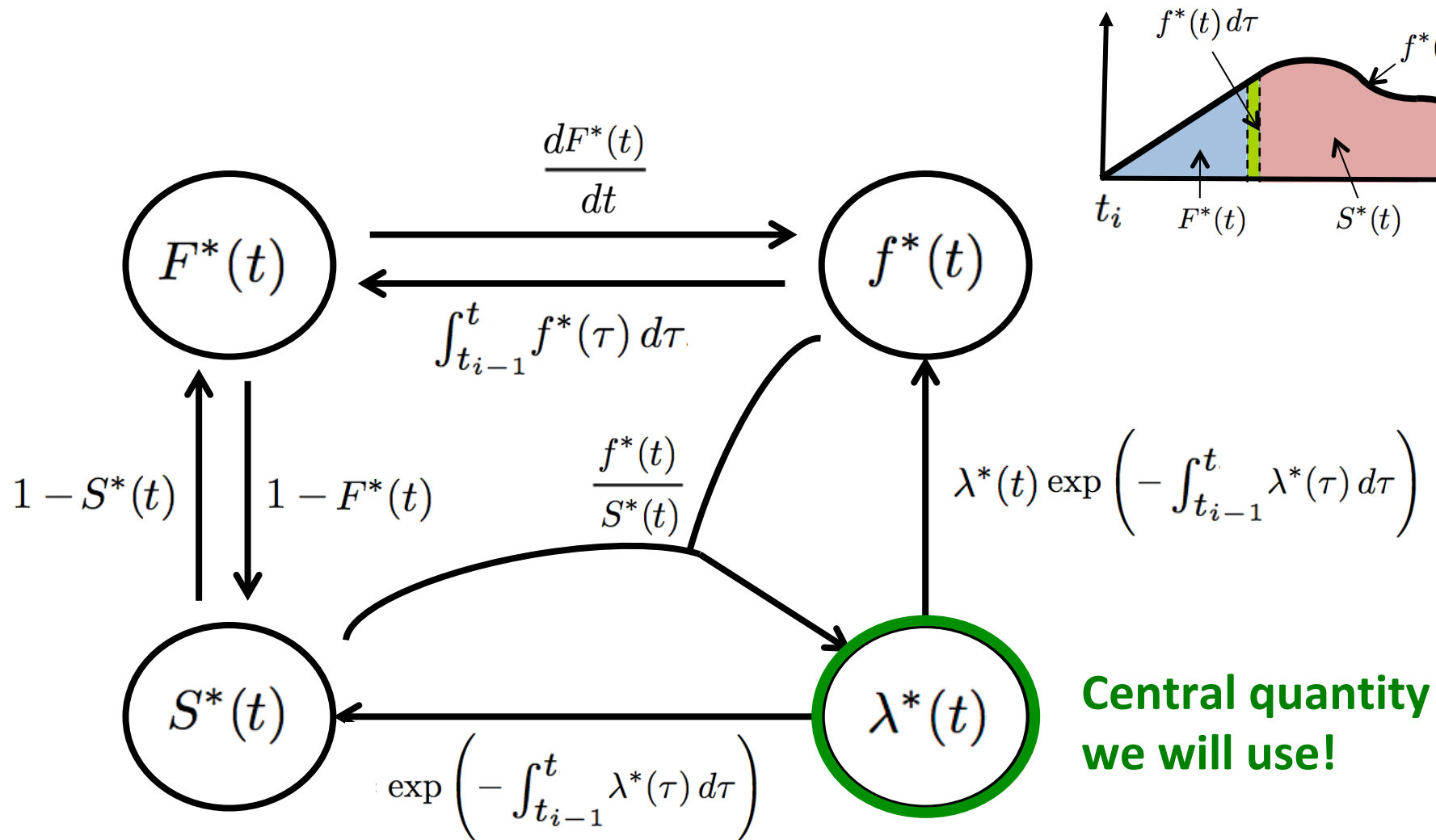
$\langle w, \phi^*(t_1) \rangle$        $\langle w, \phi^*(t_2) \rangle$        $\langle w, \phi^*(t_3) \rangle$        $\langle w, \phi^*(t) \rangle$        $\exp\left(-\int_0^T \langle w, \phi^*(\tau) \rangle d\tau\right)$

Arrows point from the inner product terms to the corresponding  $\lambda^*$  terms in the equation above.

**Suitable for model design and interpretable:**

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

# Relation between $f^*$ , $F^*$ , $S^*$ , $\lambda^*$



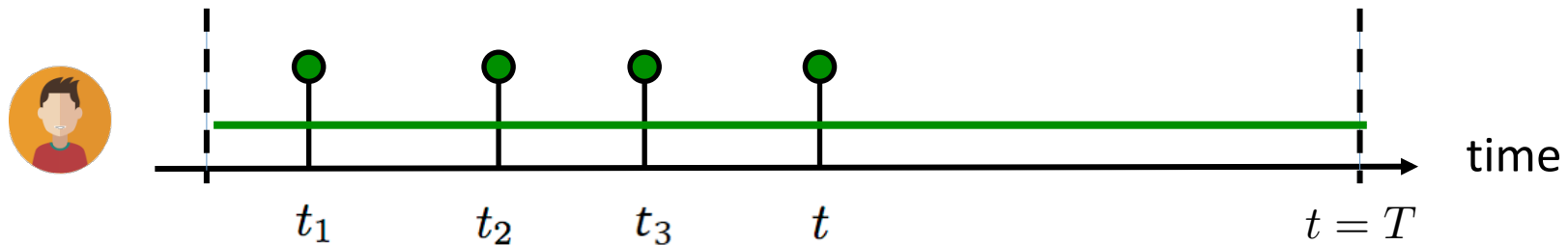
# Representation:

## Temporal Point Processes

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



# Poisson process



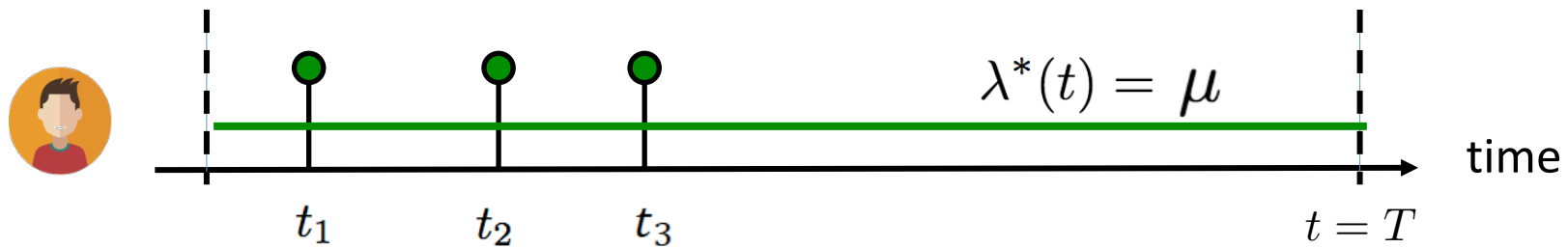
## Intensity of a Poisson process

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

## 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:**

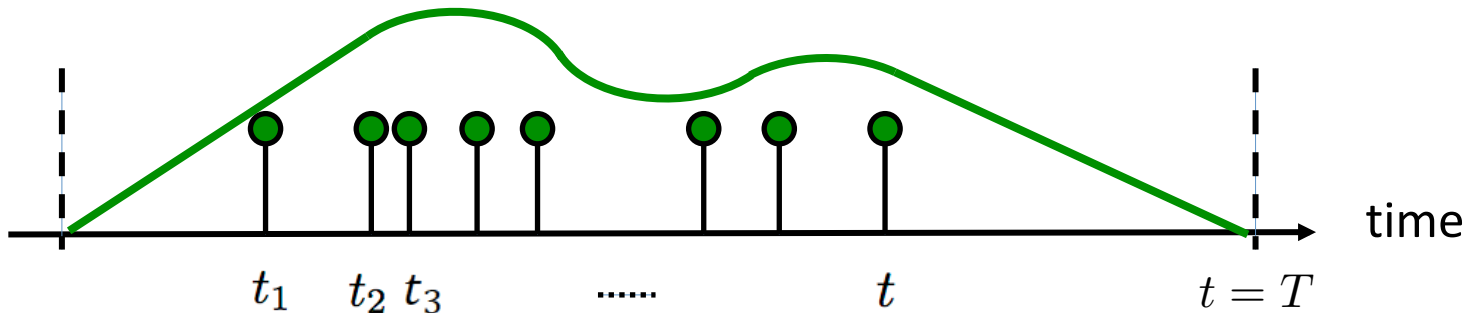
$$\mu^* = \operatorname{argmax}_{\mu} 3 \log \mu - \mu T = \frac{3}{T}$$

**Sampling using inversion sampling:**

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

*Uniform(0, 1)*  
↓  
 $u$

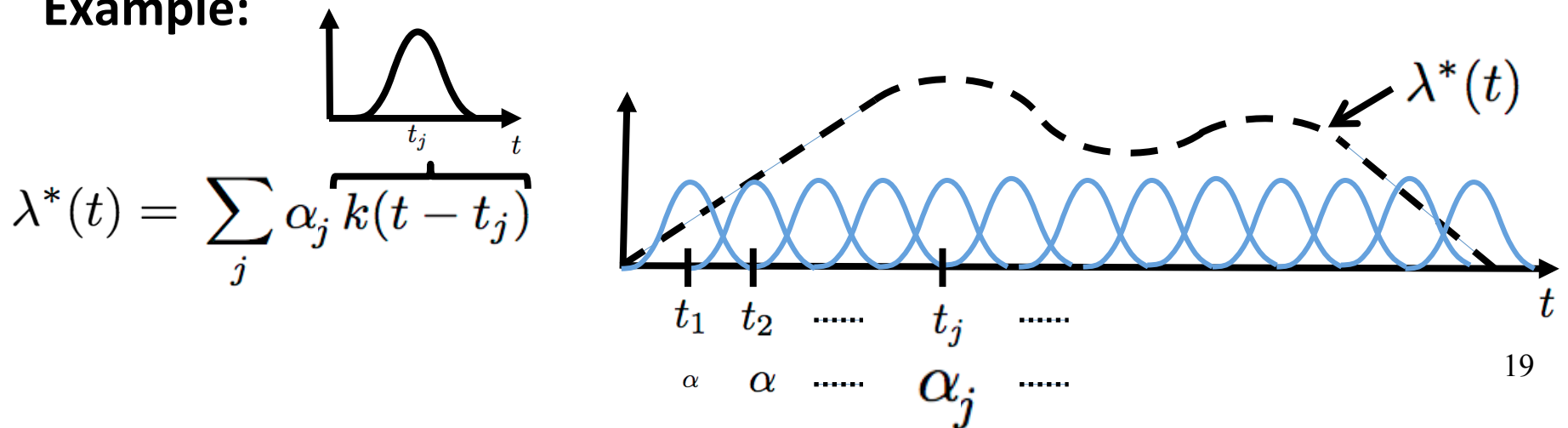
# Inhomogeneous Poisson process



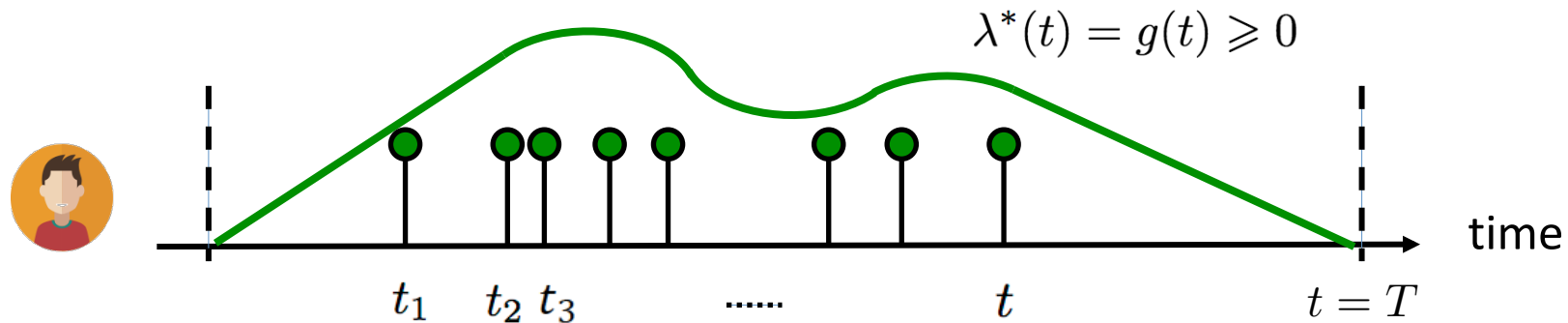
## Intensity of an inhomogeneous Poisson process

$$\lambda^*(t) = g(t) \geq 0 \quad (\text{Independent of history})$$

Example:



# Fitting & sampling from inhomogeneous Poisson

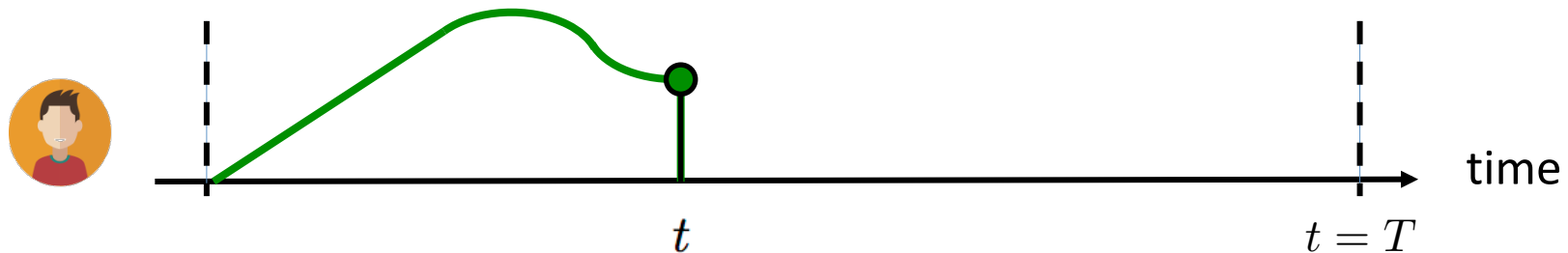


**Fitting by maximum likelihood:**  $\underset{g(t)}{\text{maximize}} \sum_{i=1}^n \log g(t_i) - \int_0^T g(\tau) d\tau$

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

1. Sample  $t$  from Poisson process with intensity  $\mu$  using inverse sampling
  2. Generate  $u_2 \sim \text{Uniform}(0, 1)$
  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

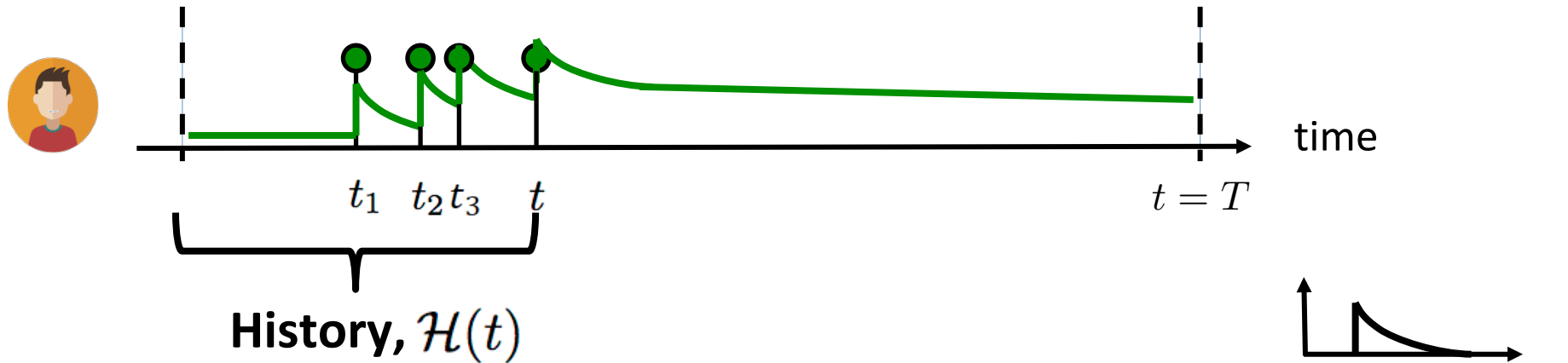
$$\lambda^*(t) = g^*(t)(1 - N(t)) \geq 0$$

Observations:

1. Limited number of occurrences

*Try sampling  
and fitting!*

# Self-exciting (or Hawkes) process



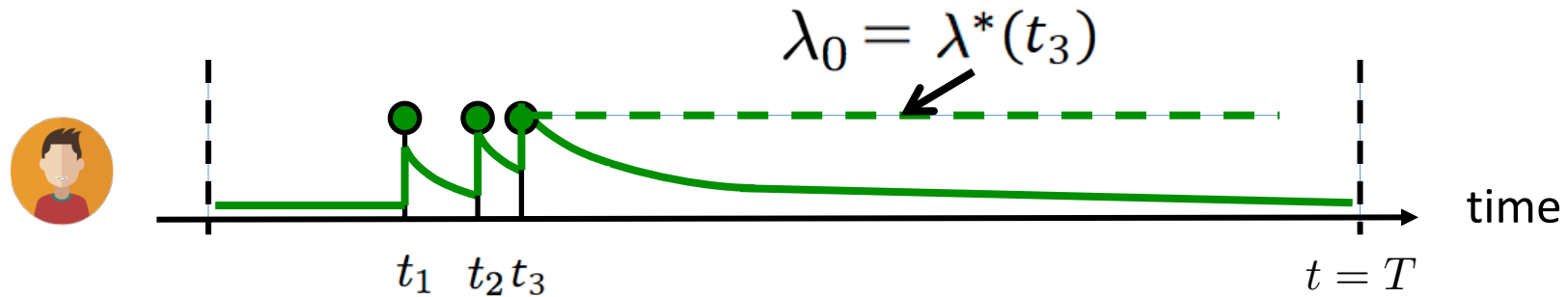
Intensity of self-exciting  
(or Hawkes) process:

$$\begin{aligned}\lambda^*(t) &= \mu + \alpha \sum_{t_i \in \mathcal{H}(t)} \kappa_\omega(t - t_i) \\ &= \mu + \alpha \kappa_\omega(t) \star dN(t)\end{aligned}$$

Observations:

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

# Fitting & Sampling a Hawkes process



Fitting by maximum likelihood:

$$\text{maximize}_{\mu, \alpha} \left. \sum_{i=1}^n \log \lambda^*(t_i) - \int_0^T \lambda^*(\tau) d\tau \right\} \begin{array}{l} \text{The max. likelihood} \\ \text{is jointly convex} \\ \text{in } \mu \text{ and } \alpha \end{array}$$

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

Key idea: the maximum of the intensity  $\lambda_0$  changes over time

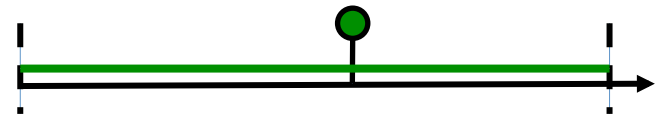
Simulating a  
Hawkes Process!

# Summary

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

Poisson processes:

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



Inho

Term

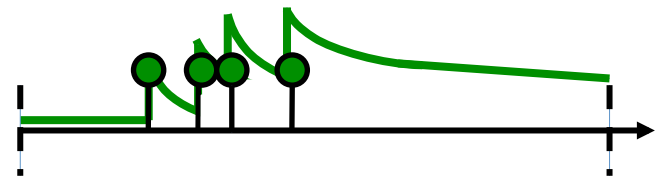
$$\lambda^*(t) = g(t)(1 - N(t))$$

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



Self-exciting point processes:

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



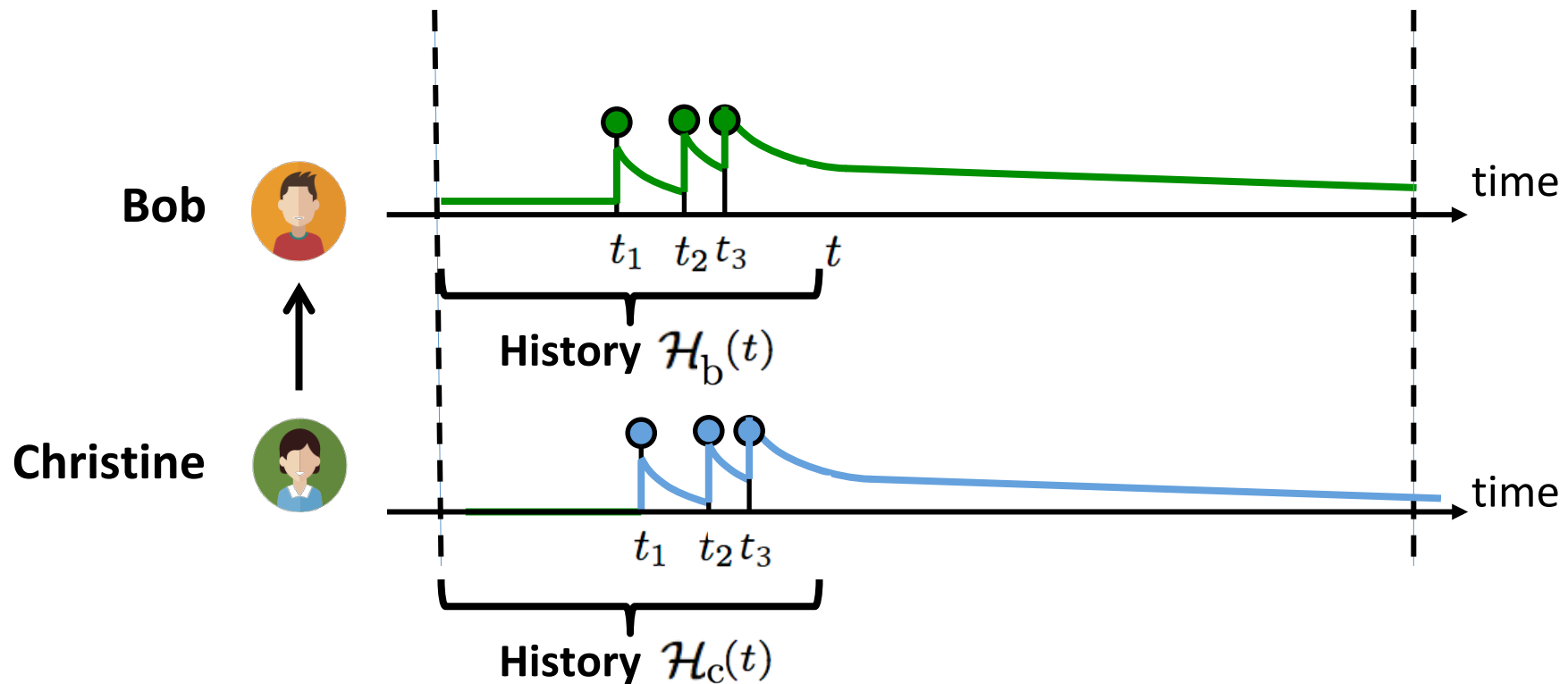


# Representation:

## Temporal Point Processes

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

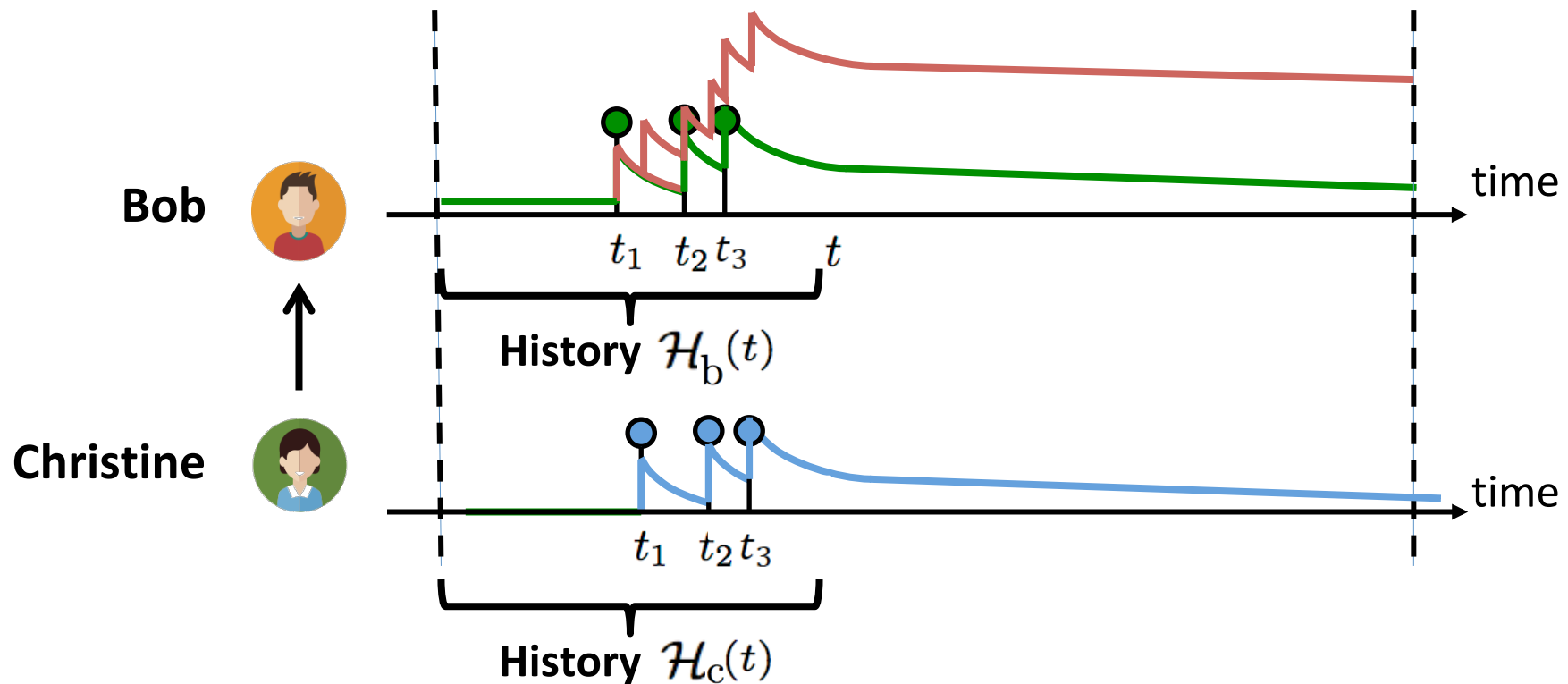
# Total intensity



**Clustered occurrence affected by neighbors**

$$\lambda^*(t) = \mu_b + \mu_c + \alpha \sum_{t_i \in \mathcal{H}_b(t)} \kappa_\omega(t - t_i) + \beta \sum_{t_i \in \mathcal{H}_c(t)} \kappa_\omega(t - t_i)$$

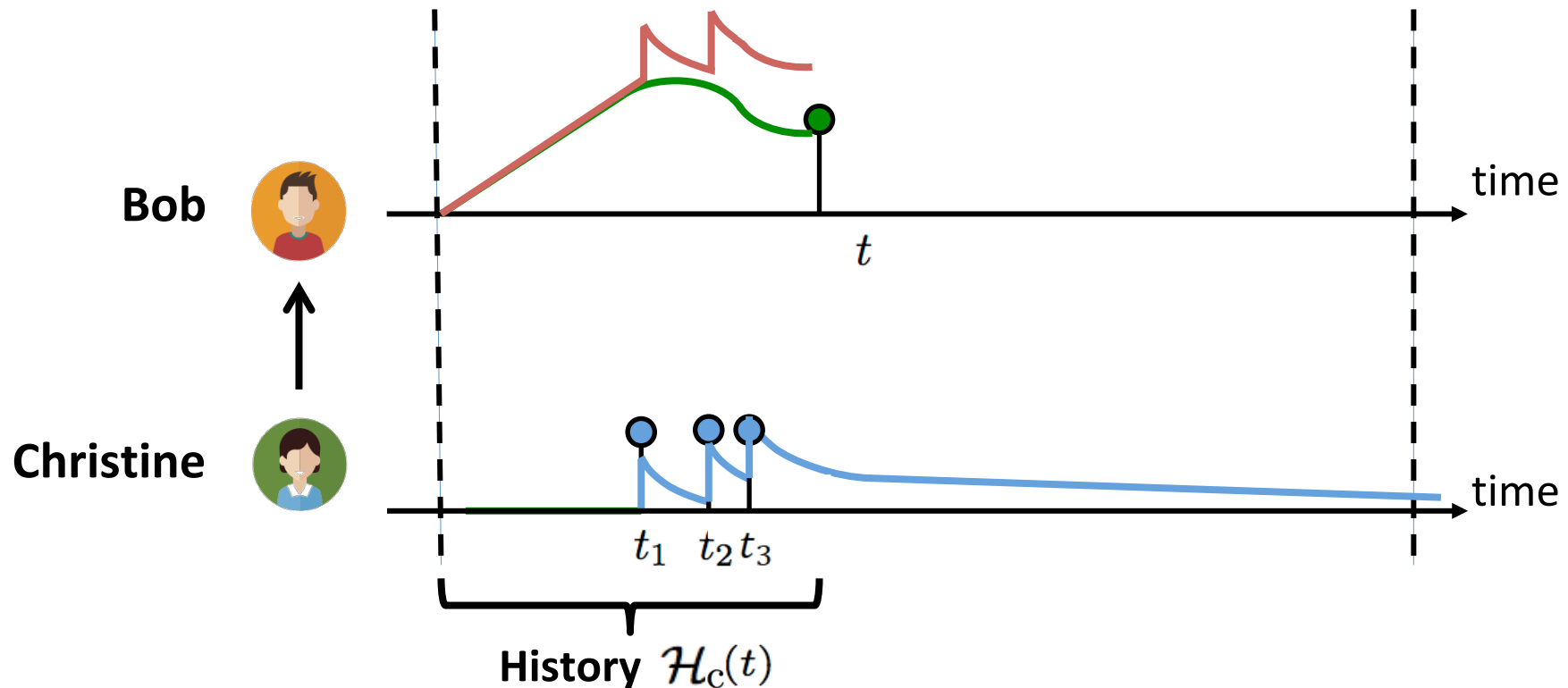
# Mutually exciting process



**Clustered occurrence affected by neighbors**

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

# Mutually exciting terminating process



**Clustered occurrence affected by neighbors**

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

# Representation:

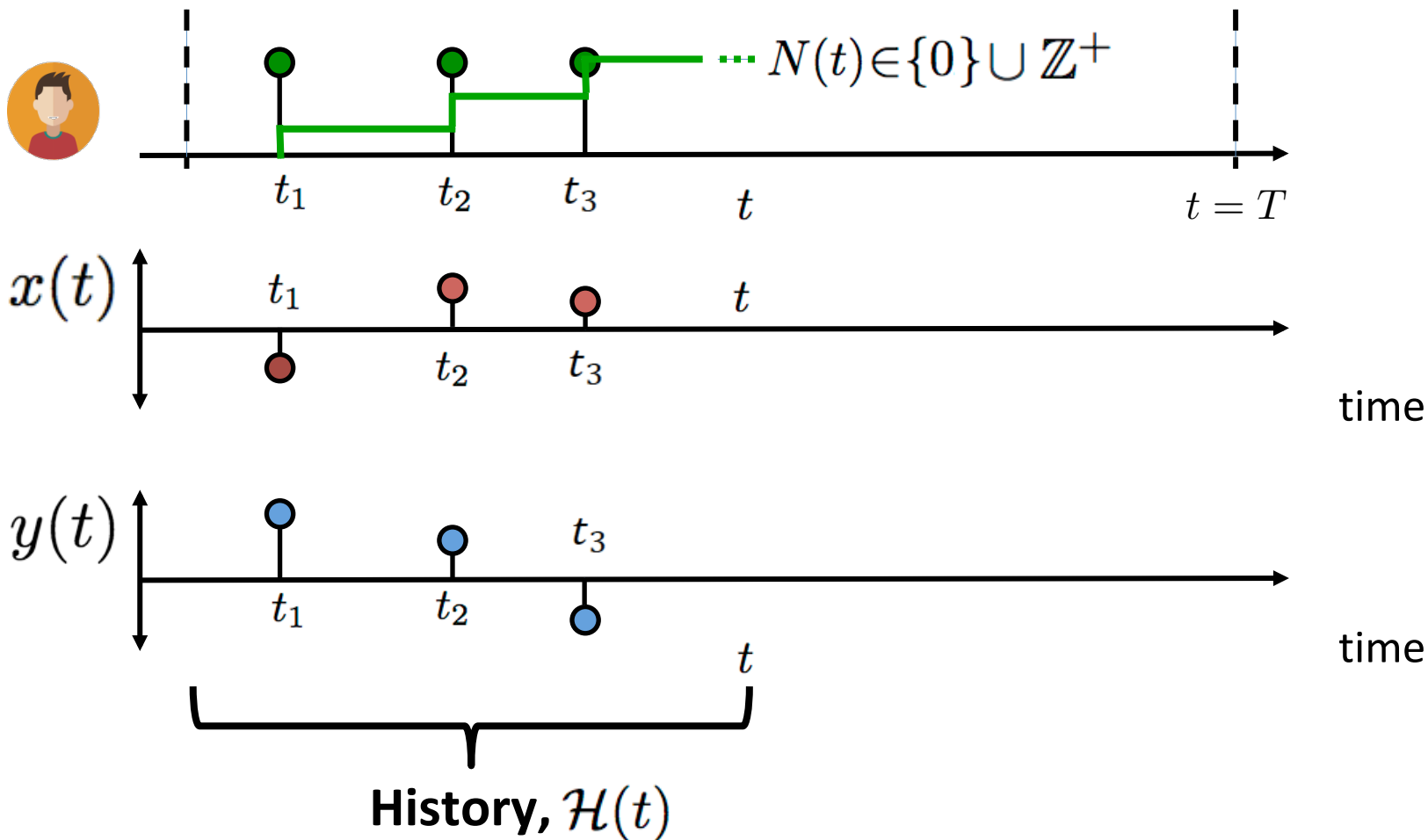
## Temporal Point Processes

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

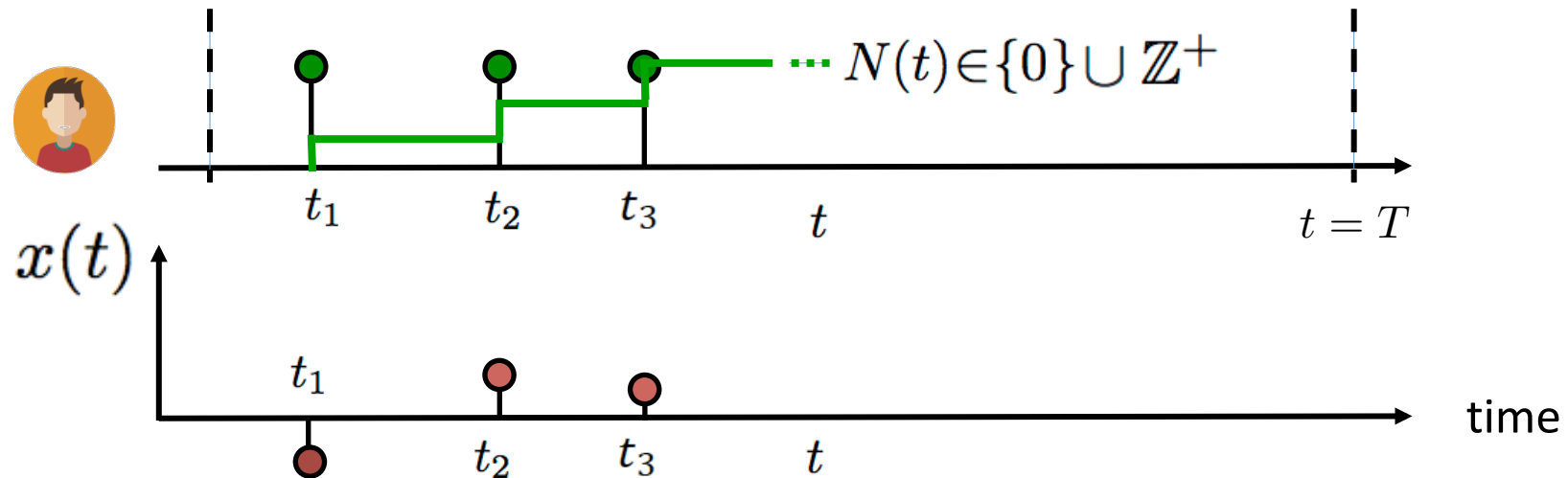
# Marked temporal point processes

## Marked temporal point process:

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



# Independent identically distributed marks



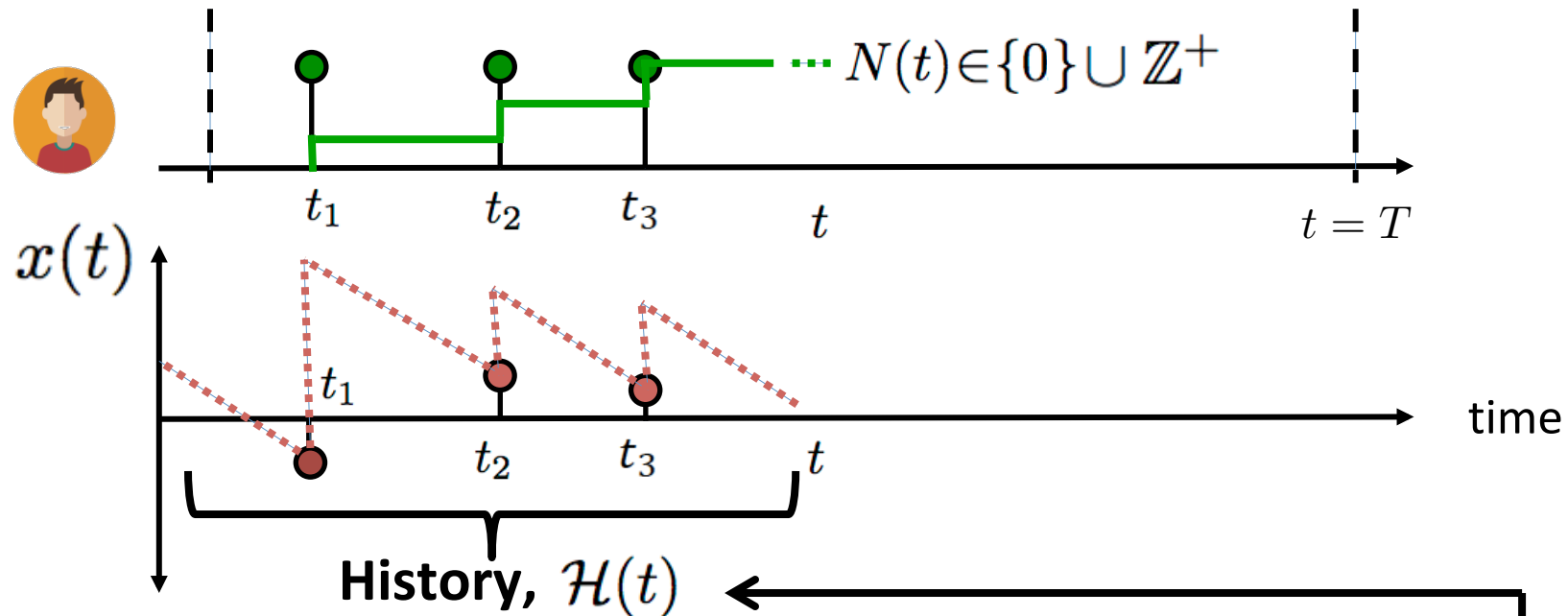
**Distribution for the marks:**

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

**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:

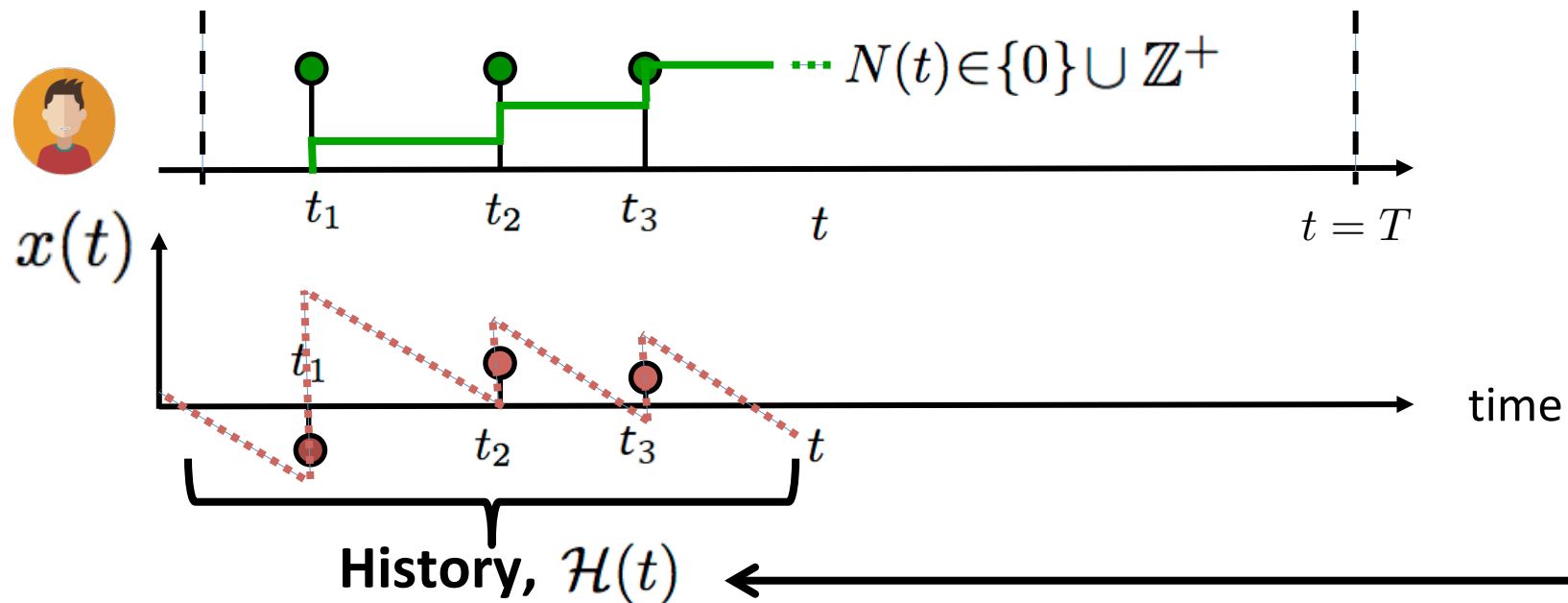
$$x(t + dt) - x(t) = dx(t) = \underbrace{f(x(t), t)dt}_{\text{Drift}} + \underbrace{h(x(t), t)dN(t)}_{\text{Event influence}}$$

Observations:

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



# Dependent marks: distribution + SDE with jumps



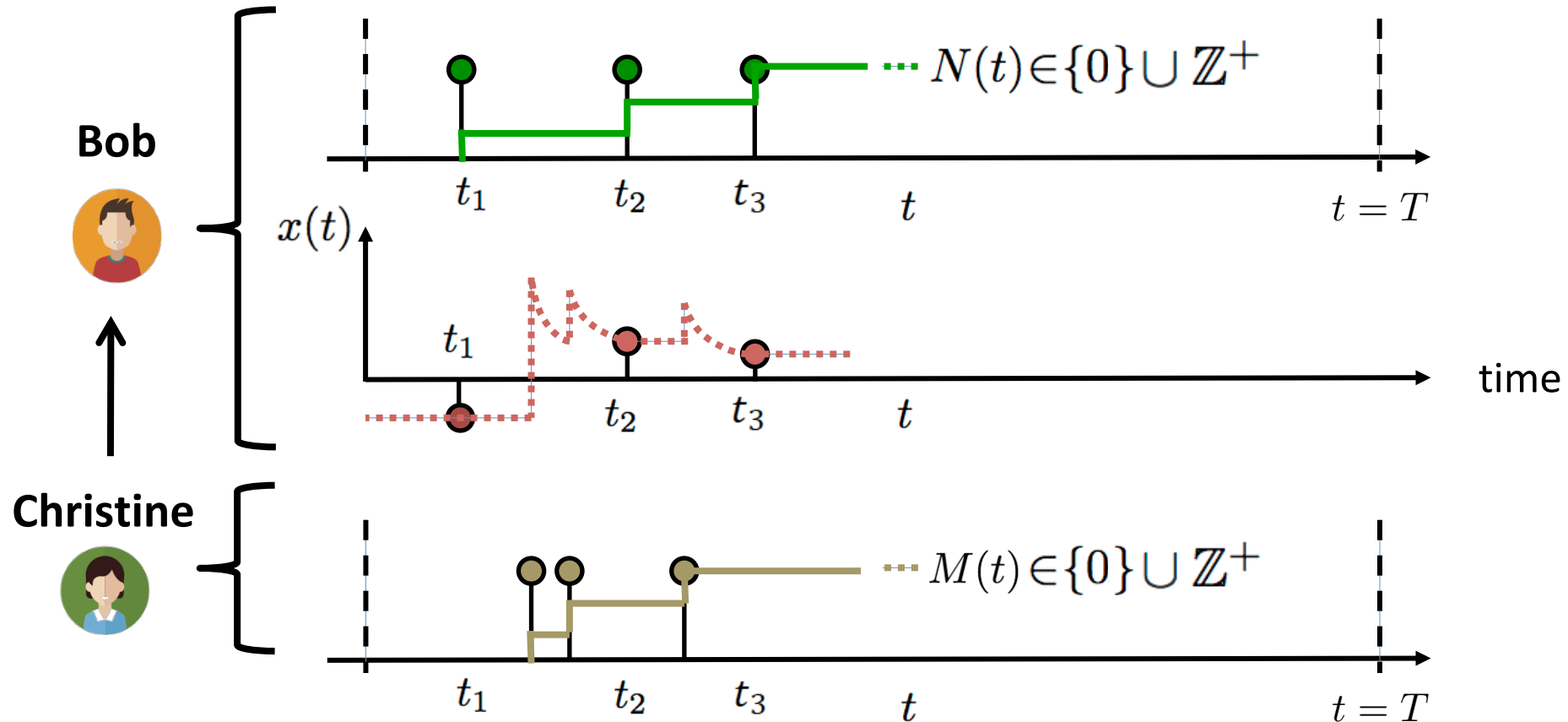
Distribution for the marks:

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

Observations:

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

# Mutually exciting + marks

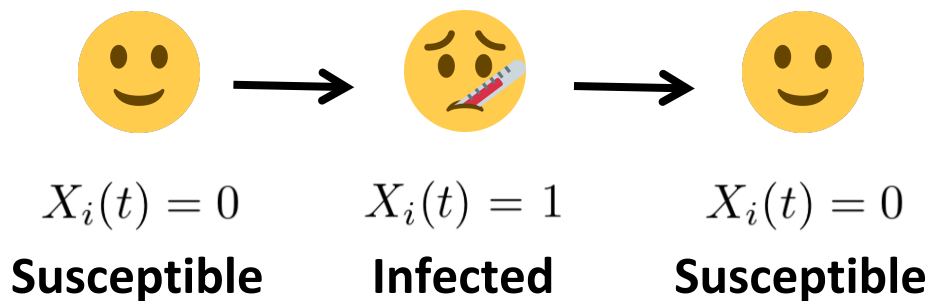


Marks affected by neighbors

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

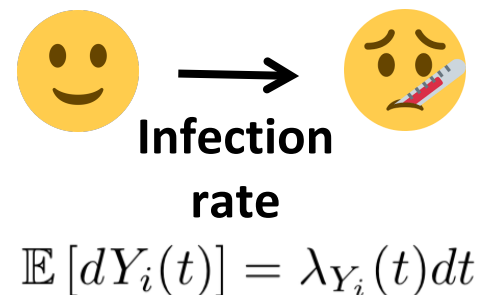
# Marked TPPs as stochastic dynamical systems

## Example: Susceptible-Infected-Susceptible (SIS)



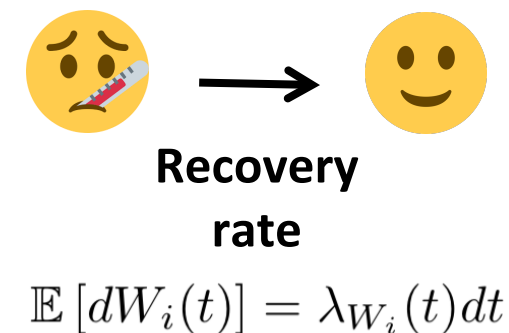
SDE with jumps

$$dX_i(t) = \underbrace{dY_i(t)}_{\text{It gets infected}} - \underbrace{dW_i(t)}_{\text{It recovers}}$$



Node is susceptible

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



SDE with jumps

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

# Outline of the Lecture

## **INTRO TO TEMPORAL POINT PROCESSES (TPPs)**

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

## **APPLICATION: CLUSTERING EVENT SEQUENCES**

1. Problem Statement
2. Introduction to DPMM
3. CRP + HP (a.k.a. HDHP)
4. Generative process