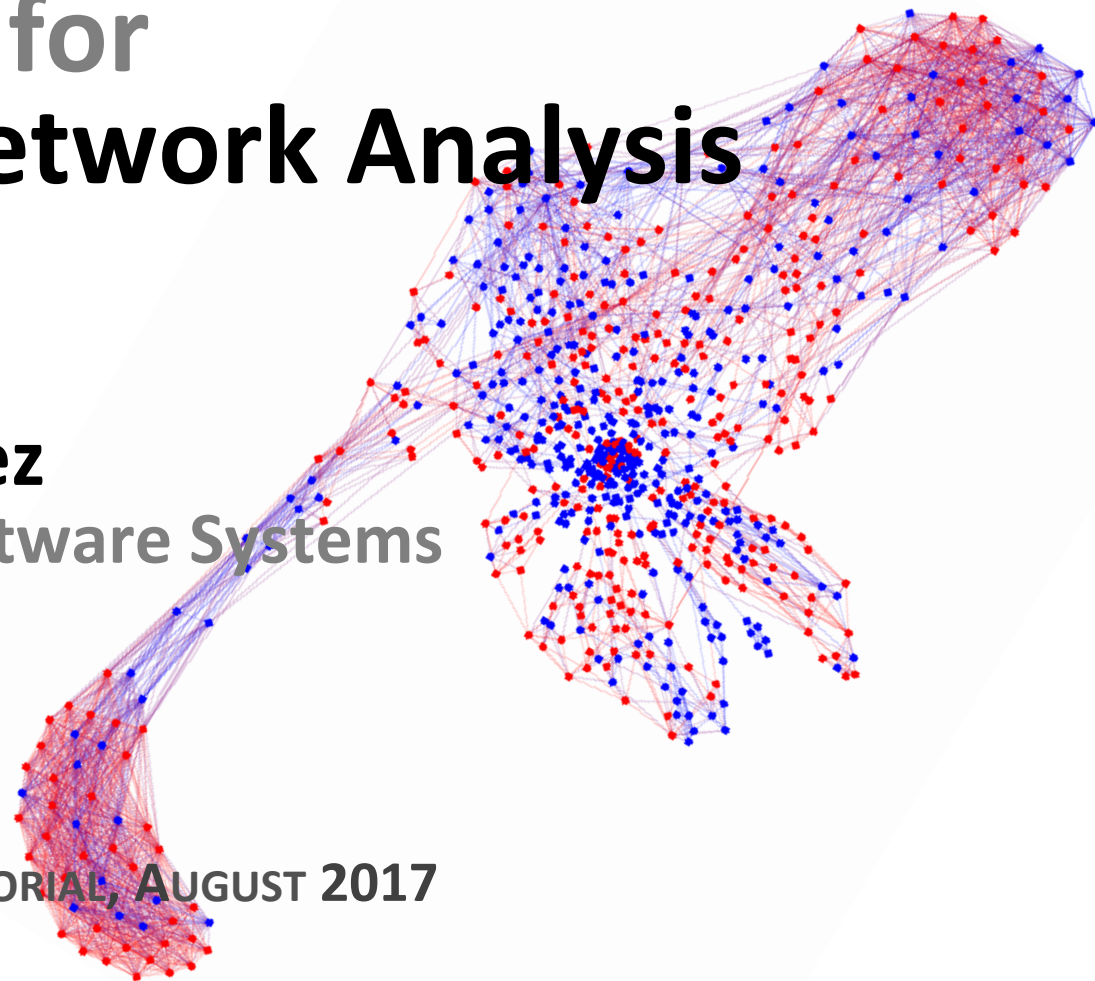


Machine learning for Dynamic Social Network Analysis

Manuel Gomez Rodriguez
Max Planck Institute for Software Systems



IJCAI TUTORIAL, AUGUST 2017

Many discrete *events* in continuous time



Variety of processes behind these events

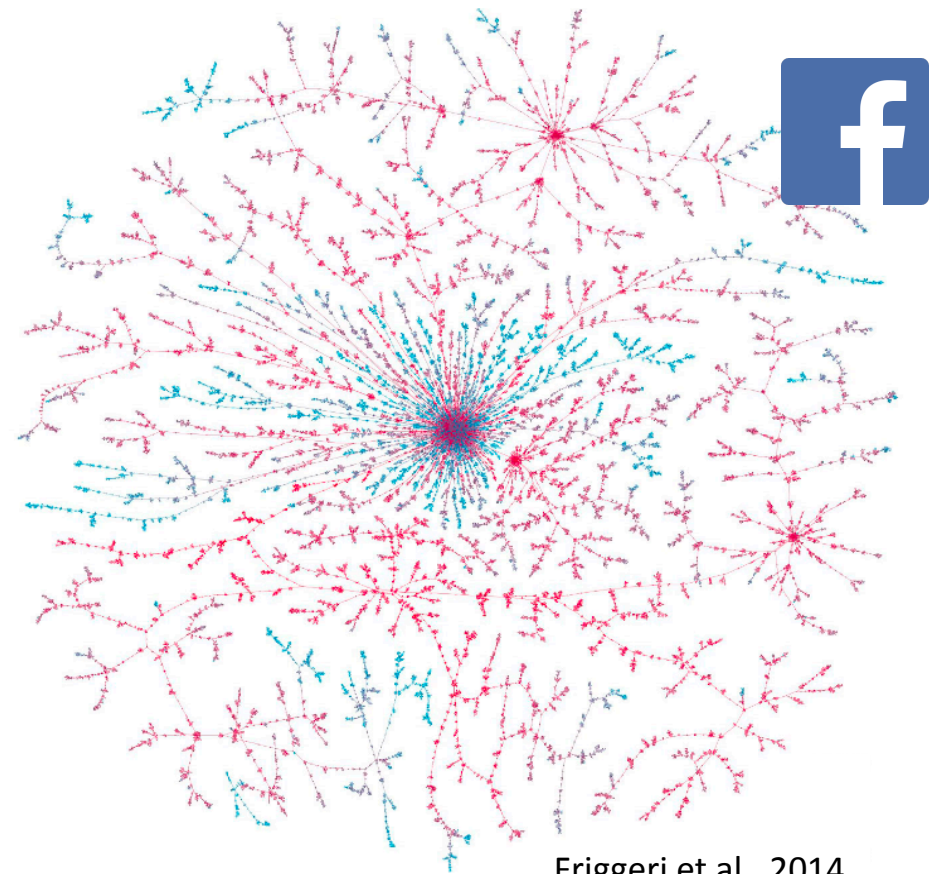
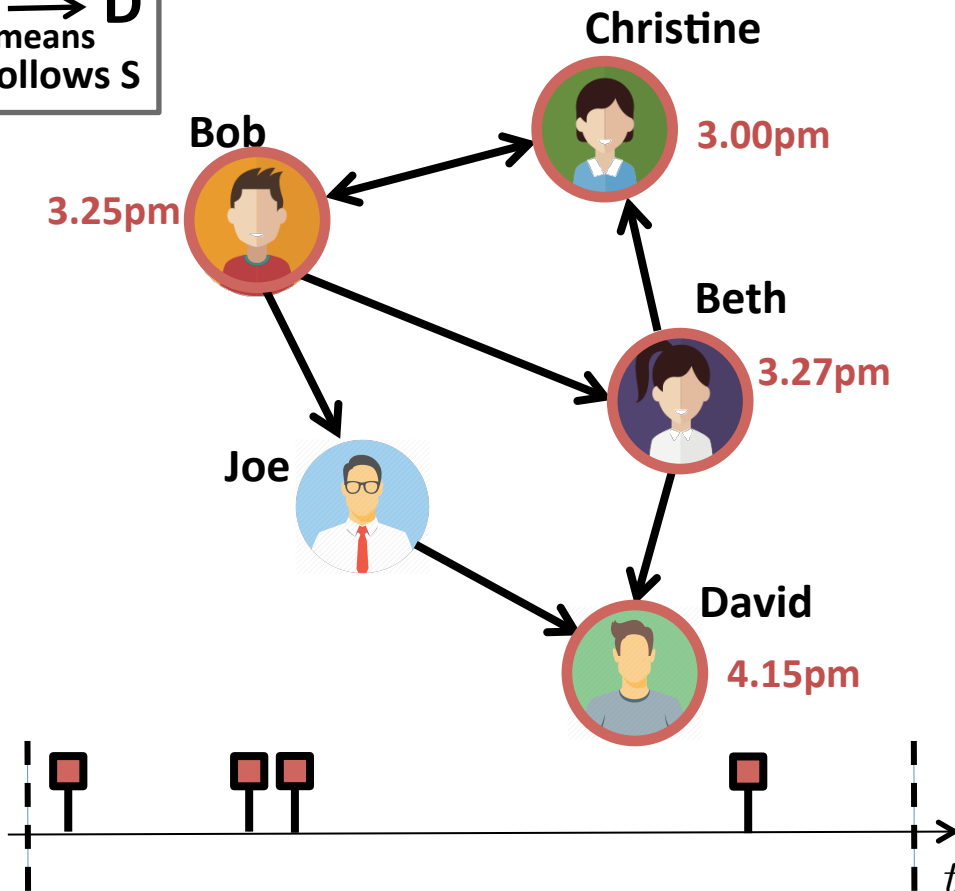
Events are (noisy) observations of a variety of complex dynamic processes...



...in a wide range of temporal scales. ³

Example I: Idea adoption/viral marketing

$S \rightarrow D$
means
D follows S



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: Information creation & curation




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 United States House of Representatives.

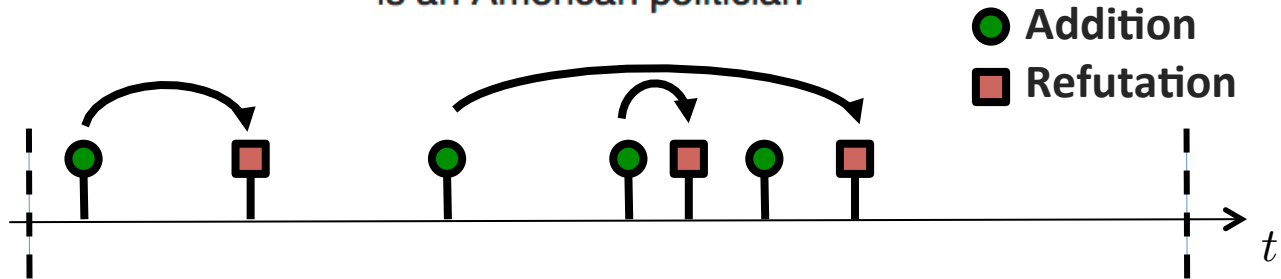
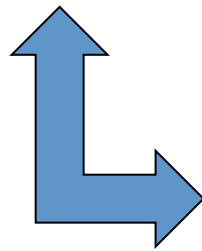
Barack Obama: Revision history

03:41, 28 November 2016	Ranze (talk contribs)	.. (301,105 bytes) (+18) .. (E)
03:32, 28 November 2016	Xin Deui (talk contribs)	.. (301,087 bytes) (-68) .. (E)
00:57, 28 November 2016	SporkBot (talk contribs)	m .. (301,155 bytes) (-37) .. (E)
07:03, 27 November 2016	Saiph121 (talk contribs)	.. (301,192 bytes) (+25) .. (E)

03:21, 20 September 2016  is a **Kenyan** politician

X ↓ possible vandalism by MLM2016

is an American politician



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.



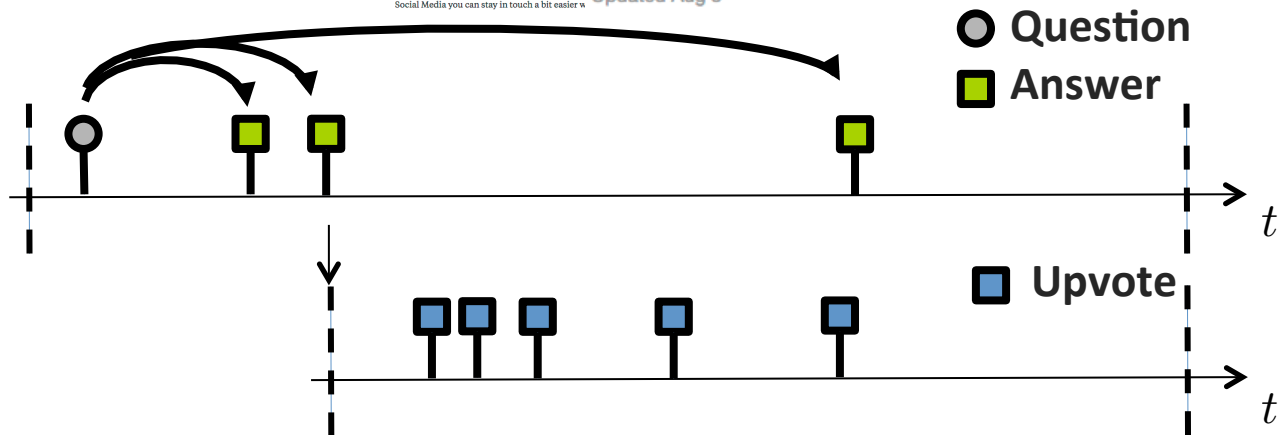
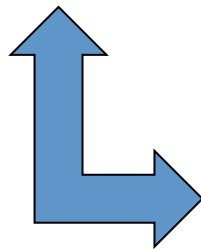
M Sharma, Lived in Australia as Migrant, Student, Worker, Business Owner & Family Man

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
- Society is more and more connected and thanks to Social Media you can stay in touch a bit easier with them

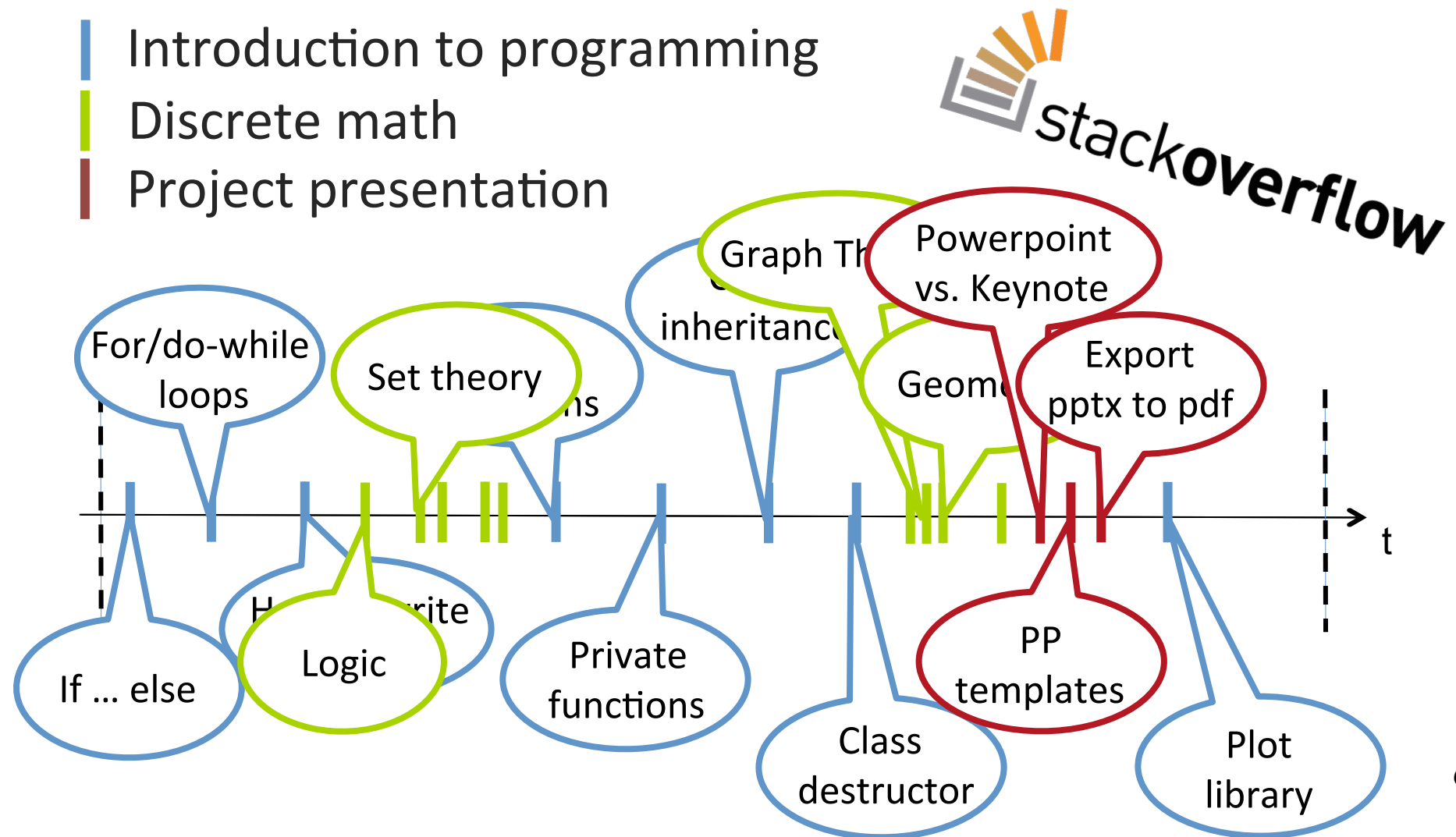
Updated Aug 3



Example III: Learning trajectories



1st year computer science student



Detailed *event traces*

DETAILED TRACES OF ACTIVITY



Warren Buffett ✓
@WarrenBuffett



Warren is in the house.

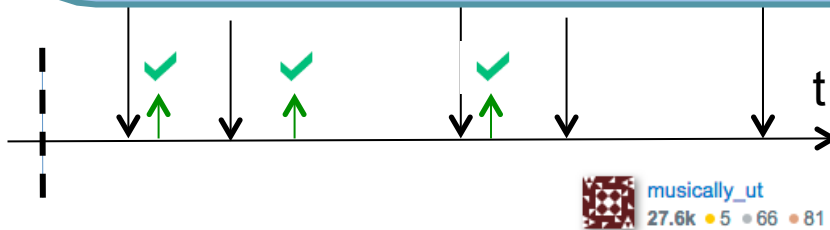


Manuel Gomez Rodriguez updated his cover photo.
April 17 at 1:14pm · 🌐



Pique-Longue, French Pyrenees
Easter 2017

The availability of event traces
boosts a new generation of
data-driven models and algorithms

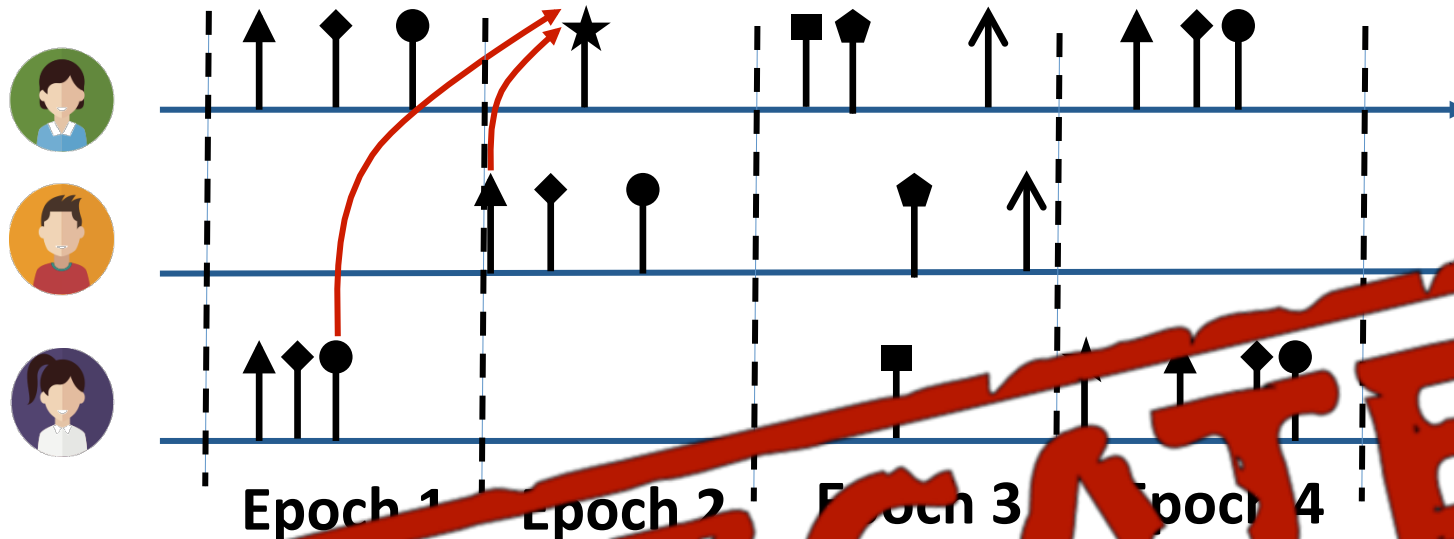


Like Comment Share

Mehrdad Farajtabar, Lili Yavis-Hound and 24 others

Rober Tab Pu 🤩 wow!
Like · Reply · April 17 at 1:32pm

Previously: discrete-time models & algorithms



Discrete-time models artificially introduce epochs:

1. How long is each epoch? Data is very heterogeneous.

2. How to aggregate events within an epoch?

3. What if no event within an epoch?

4. Time is treated as index or conditioning variable, not easy to deal with time-related queries.

Outline of the Seminar

REPRESENTATION: TEMPORAL POINT PROCESSES

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

Next

APPLICATIONS: MODELS

1. Information propagation
2. Information reliability
3. Knowledge acquisition

APPLICATIONS: CONTROL

1. Activity shaping
2. When-to-post

Representation:

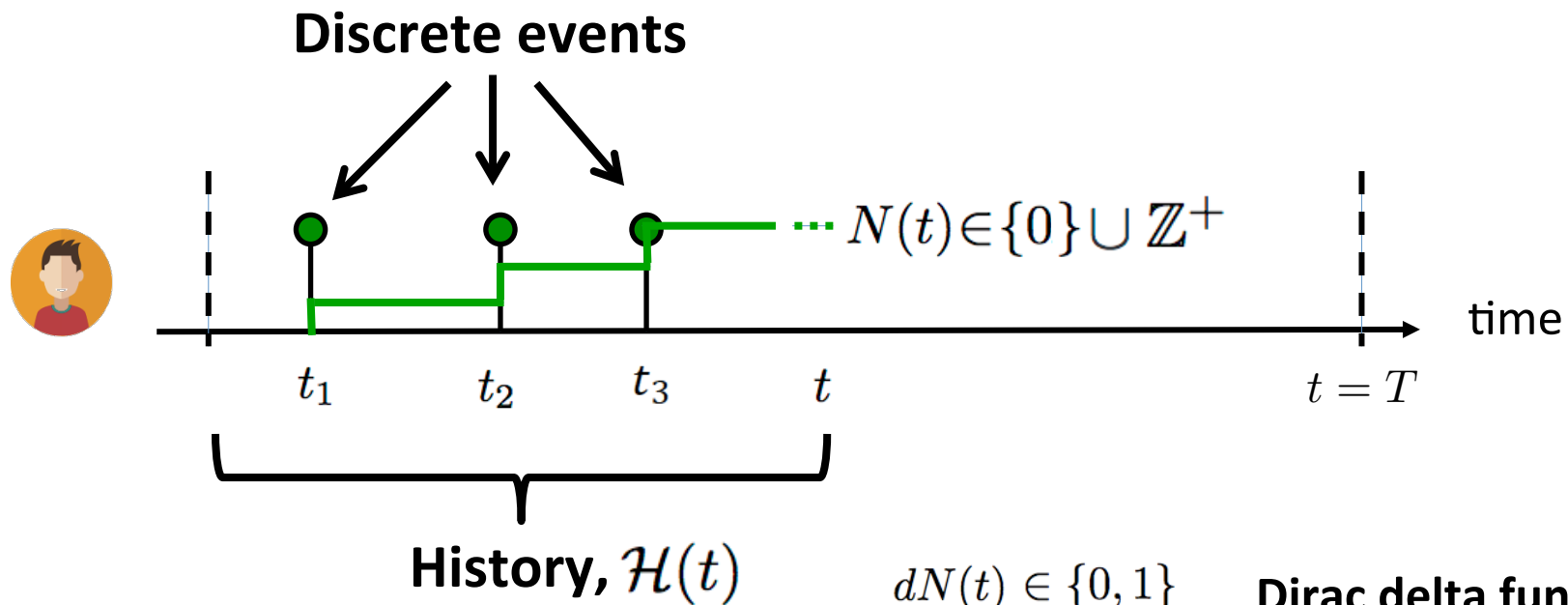
Temporal Point Processes

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

Temporal point processes

Temporal point process:

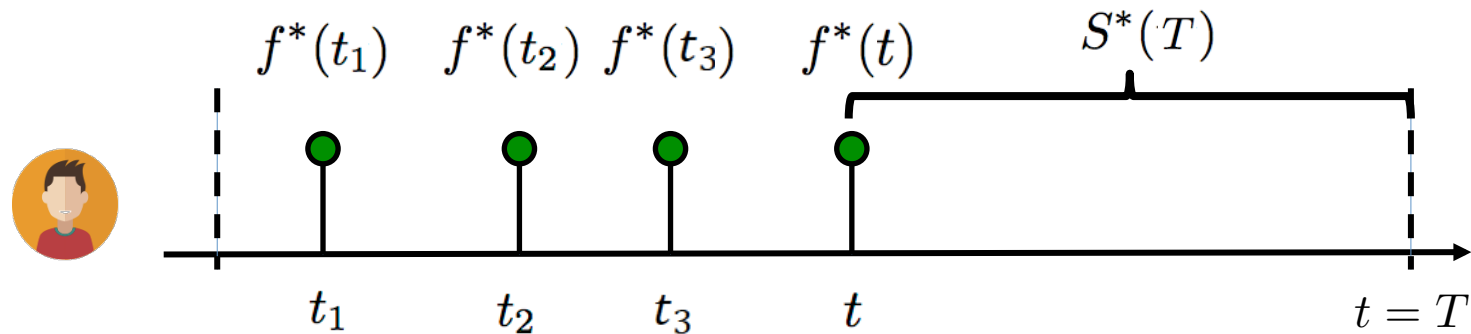
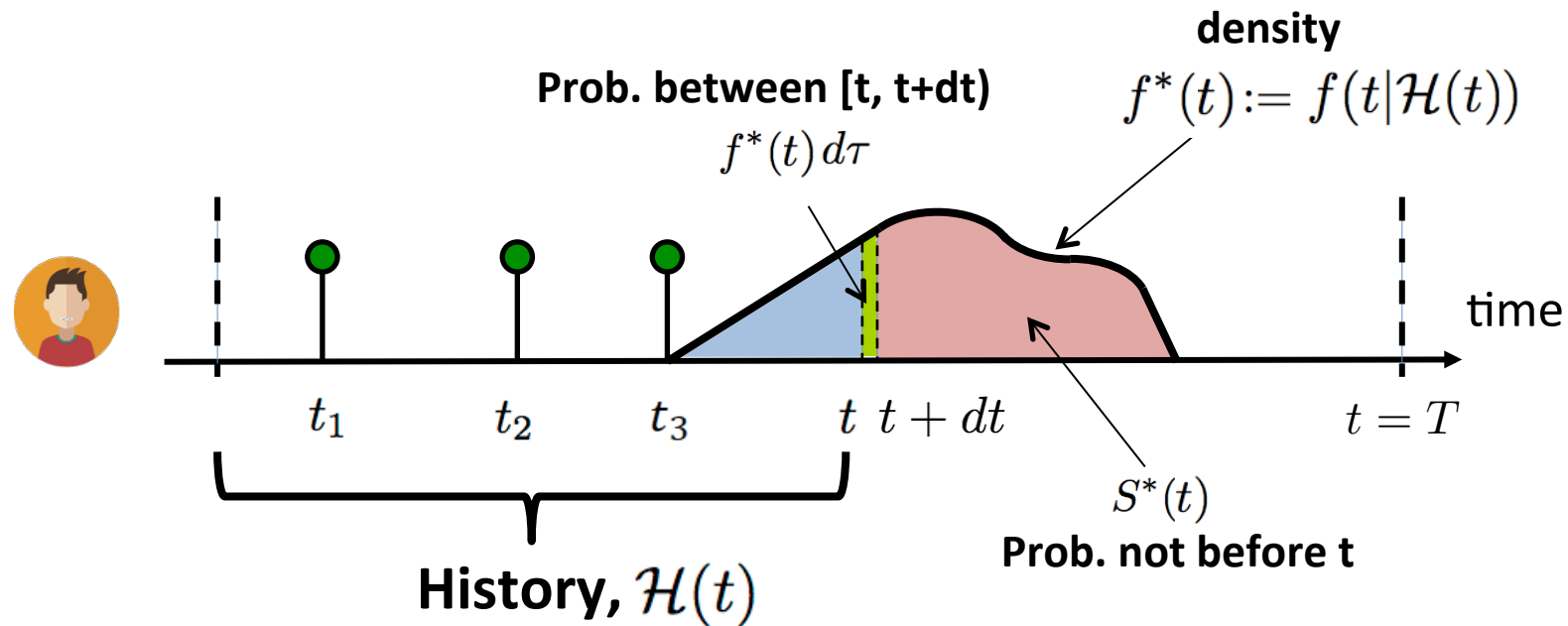
A random process whose realization consists of discrete events localized in time



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

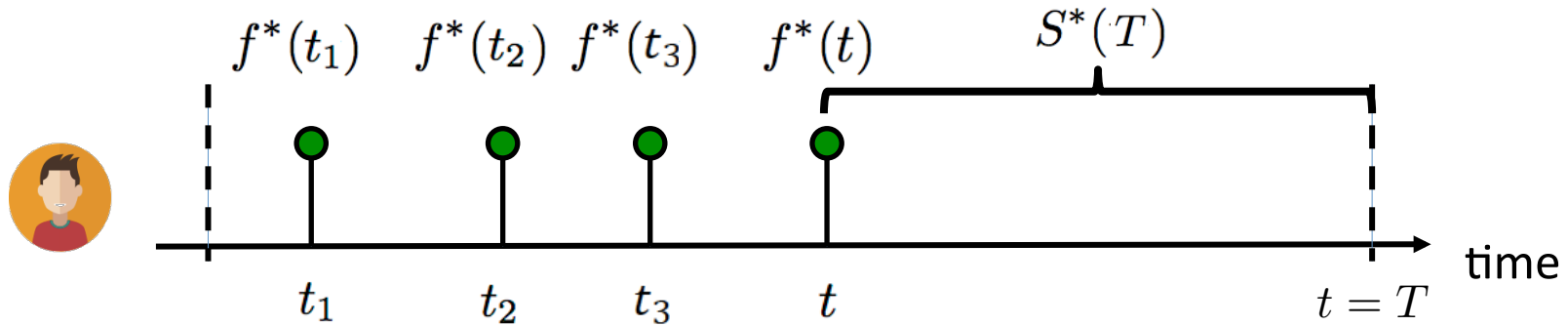
$dN(t) \in \{0, 1\}$ Dirac delta function

Model time as a random variable



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

Problems of density parametrization (I)



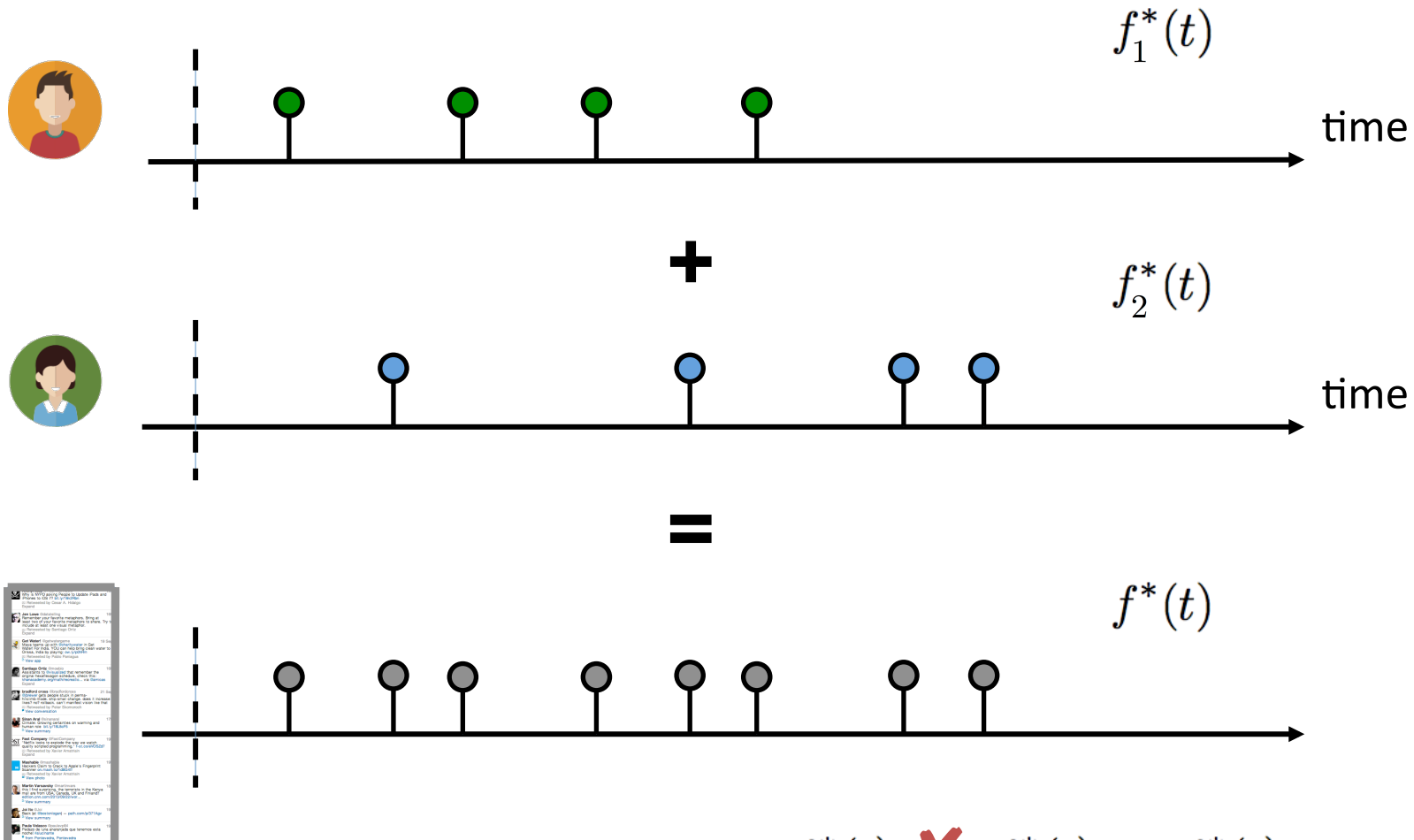
$$\begin{array}{cccccc}
 f^*(t_1) & f^*(t_2) & f^*(t_3) & f^*(t) & S^*(T) & \\
 \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \\
 \frac{\exp\langle w, \psi^*(t_1) \rangle}{Z} & \frac{\exp\langle w, \psi^*(t_2) \rangle}{Z} & \frac{\exp\langle w, \psi^*(t_3) \rangle}{Z} & \frac{\exp\langle w,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

It is **difficult for model design and interpretability**:

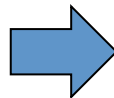
1. Densities need to integrate to 1 (i.e., partition function)
2. Difficult to combine timelines

Problems of density parametrization (II)

Difficult to combine timelines:



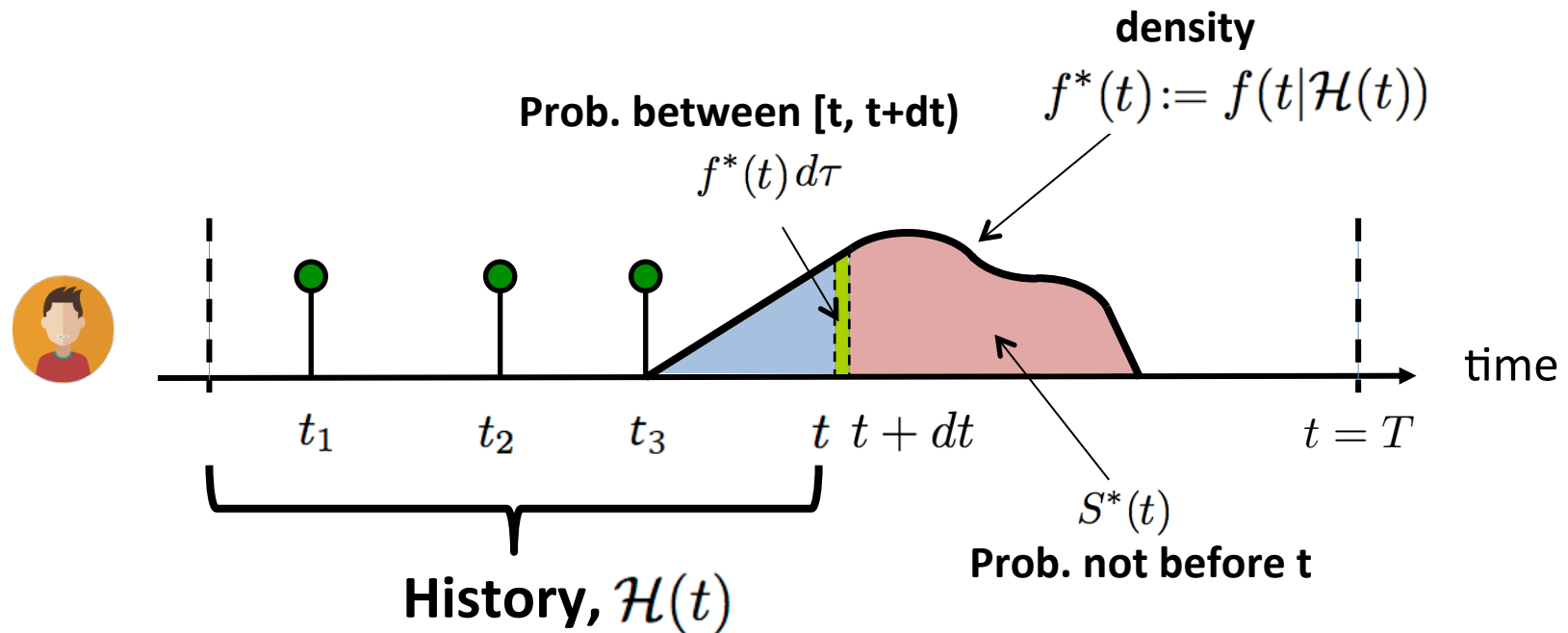
Sum of random processes



$$f^*(t) \neq f_1^*(t) + f_2^*(t)$$

$$f^*(t) \neq f_1^*(t) * f_2^*(t)$$

Intensity function



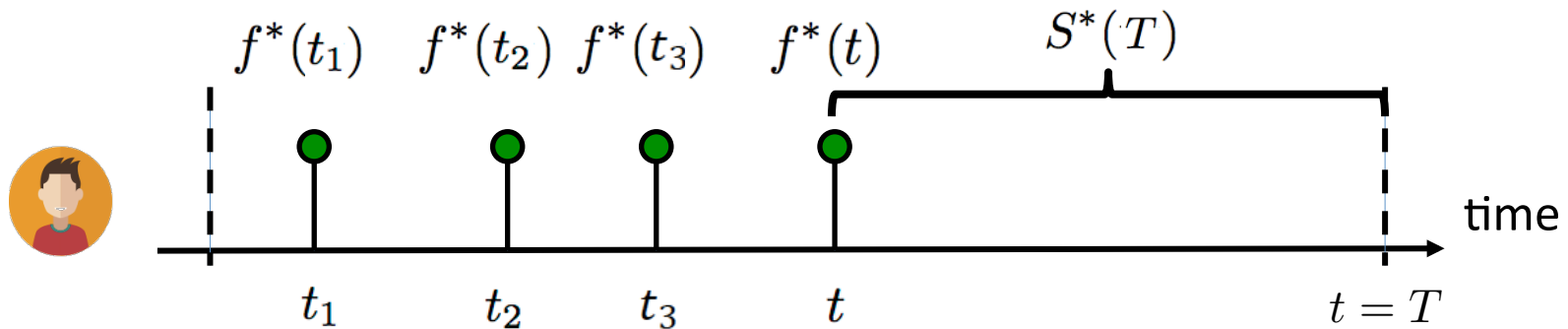
Intensity:

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)



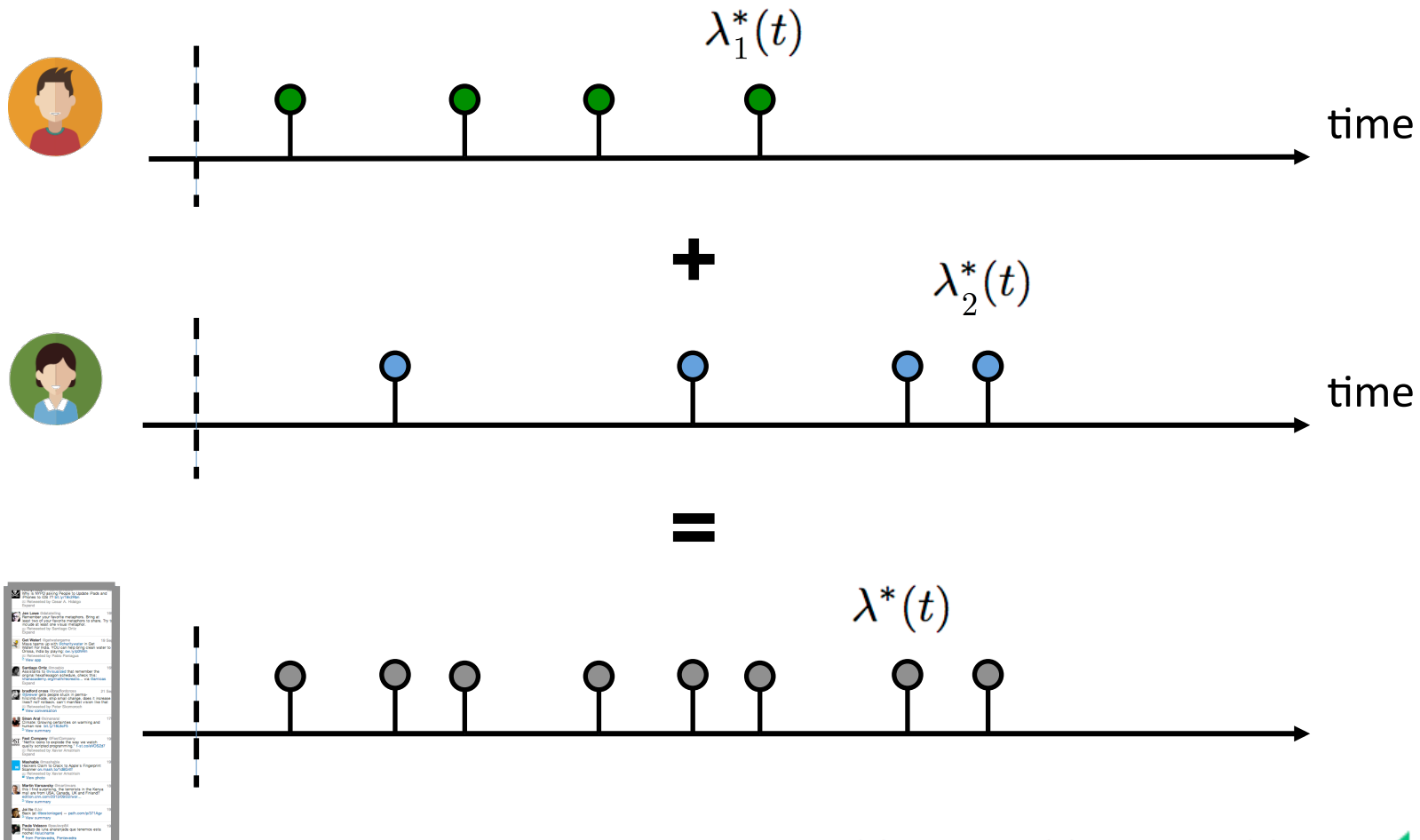
$$\begin{array}{ccccccc}
 \lambda^*(t_1) & \lambda^*(t_2) & \lambda^*(t_3) & \lambda^*(t) & \exp\left(-\int_0^T \lambda^*(\tau) d\tau\right) & & \\
 \nearrow & \nearrow & \uparrow & \nearrow & \nwarrow & & \\
 \langle w, \phi^*(t_1) \rangle & & \langle w, \phi^*(t_3) \rangle & & & & \\
 \langle w, \phi^*(t_2) \rangle & & & \langle w, \phi^*(t) \rangle & & \exp\left(-\int_0^T \langle w, \phi^*(\tau) \rangle d\tau\right) & \\
 & & & & & \nwarrow & \\
 & & & & & &
 \end{array}$$

Suitable for model design and interpretable:

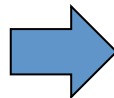
1. Intensities only need to be nonnegative
2. Easy to combine timelines

Advantages of intensity parametrization (II)

Easy to combine timeline:



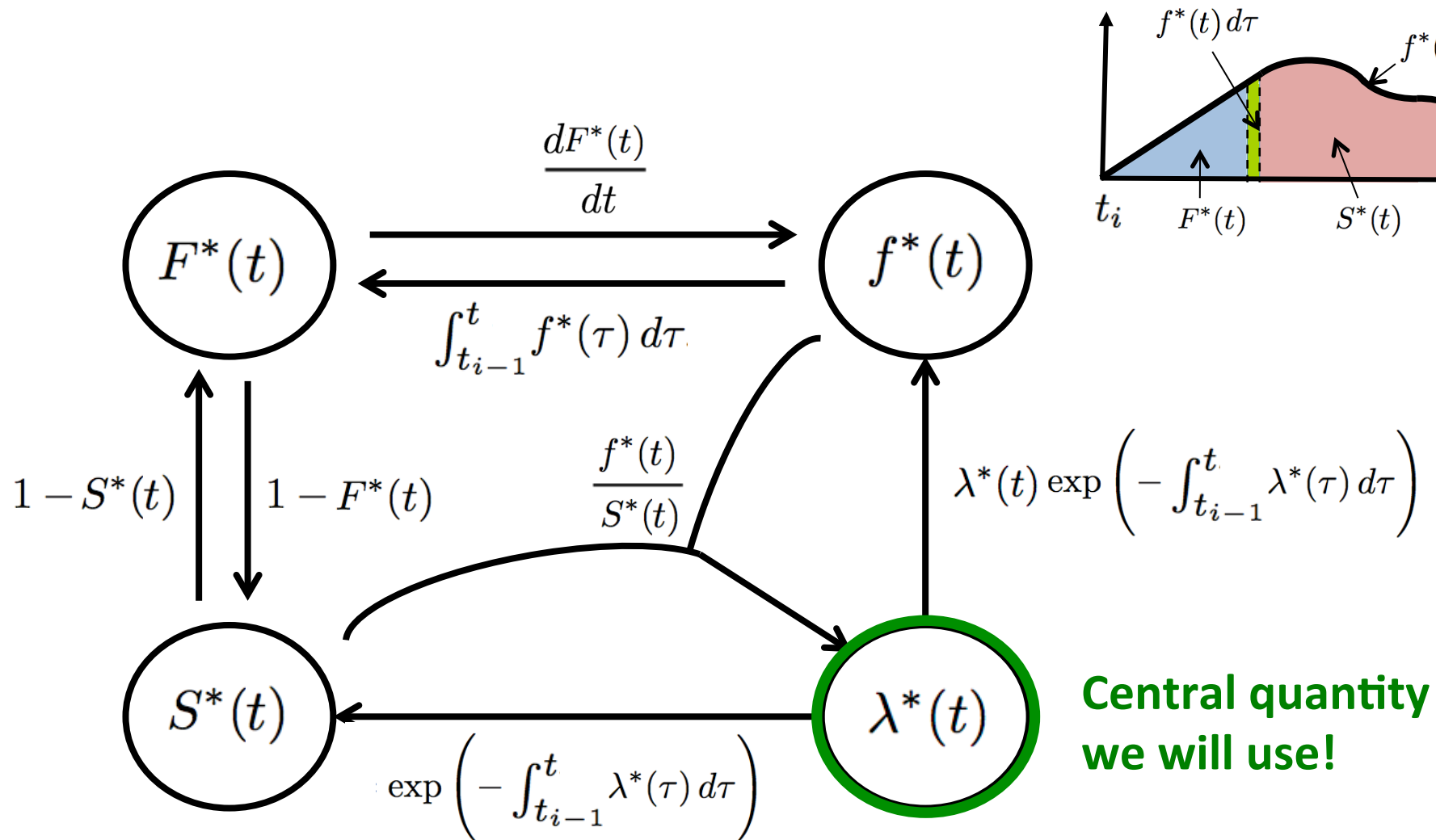
Sum of random processes



$$\lambda^*(t) = \lambda_1^*(t) + \lambda_2^*(t) \quad \checkmark$$

$$\lambda^*(t) \neq \lambda_1^*(t) * \lambda_2^*(t)$$

Relation between f^* , F^* , S^* , λ^*

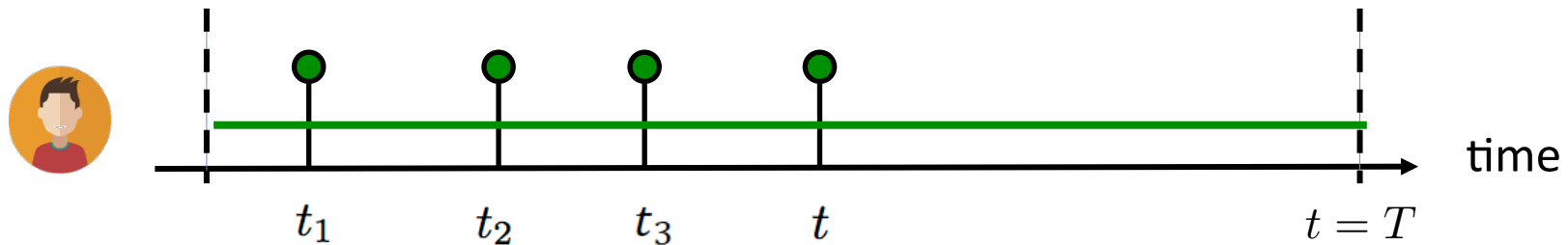


Representation:

Temporal Point Processes

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

Poisson process



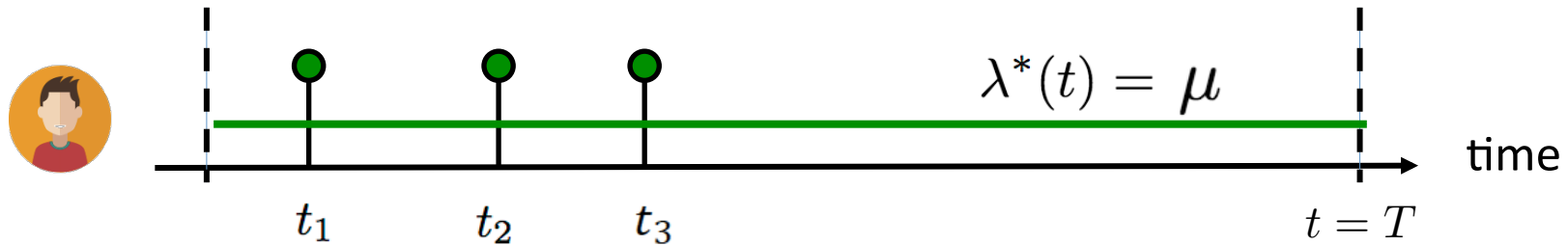
Intensity of a Poisson process

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

Observations:

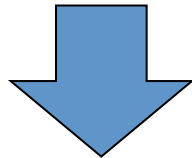
1. Intensity independent of history
2. Uniformly random occurrence
3. Time interval follows exponential distribution

Fitting a Poisson from (historical) timeline



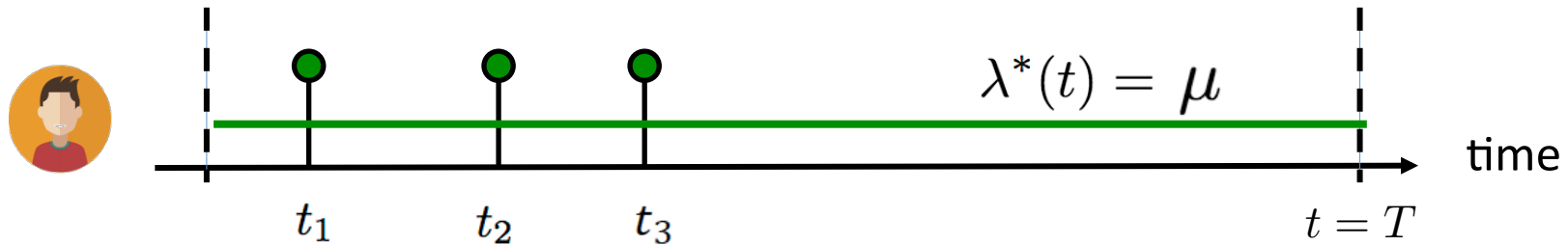
$$\begin{array}{ccccccc} \lambda^*(t_1) & \lambda^*(t_2) & \lambda^*(t_3) & \underbrace{\exp\left(-\int_0^T \lambda^*(\tau) d\tau\right)}_{\exp(-\mu T)} \\ \uparrow & \uparrow & \uparrow & \\ \mu & \mu & \mu & \end{array}$$

Maximum
likelihood



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

Sampling from a Poisson process



We would like to sample: $t \sim \mu \exp(-\mu(t - t_3)) + t_3$

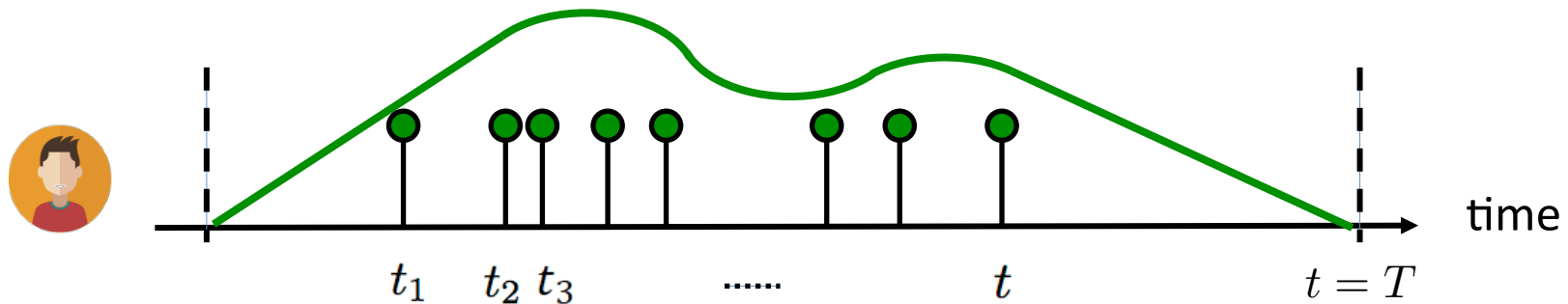
We sample using inversion sampling:

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

Uniform(0, 1)
↓

$$\mathbb{P}(F_t^{-1}(u) \leq t) = \mathbb{P}(u \leq F_t(t)) \stackrel{\substack{\uparrow \\ F_t^{-1}(u) = u}}{=} F_t(t)$$

Inhomogeneous Poisson process



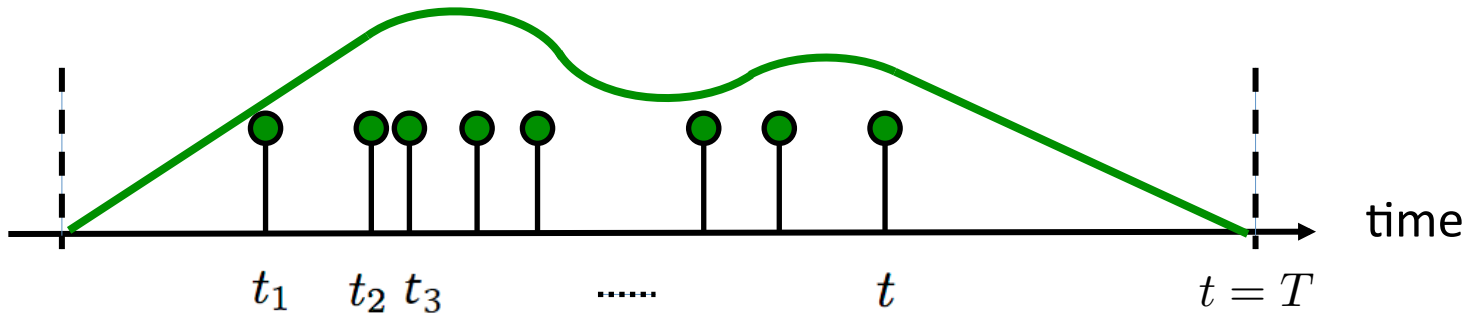
Intensity of an inhomogeneous Poisson process

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

Observations:

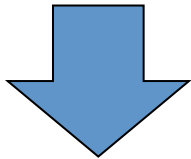
- 1. Intensity independent of history**

Fitting an inhomogeneous Poisson



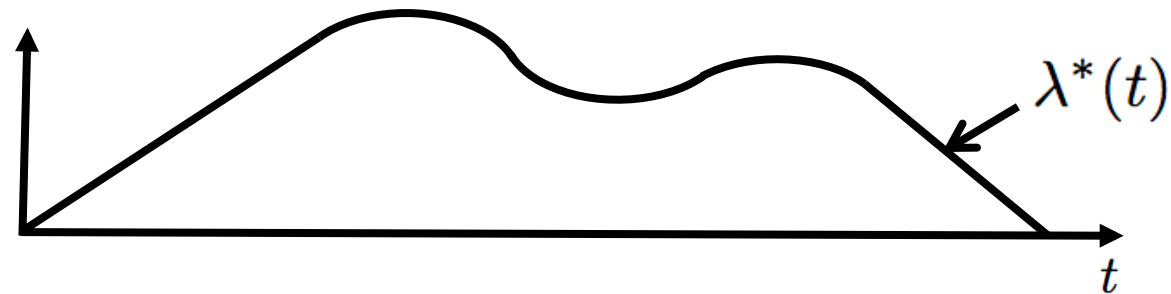
$$\begin{array}{ccccccc}
 \lambda^*(t_1) & \lambda^*(t_2) & \lambda^*(t_3) & \cdots & \lambda^*(t_n) & \underbrace{\exp\left(-\int_0^T \lambda^*(\tau) d\tau\right)} & \\
 \uparrow & \uparrow & \uparrow & & \uparrow & & \\
 g(t_1) & g(t_2) & g(t_3) & & g(t_n) & \underbrace{\exp\left(-\int_0^T g(\tau) d\tau\right)} & \\
 \end{array}$$

Maximum likelihood



$$\left. \begin{array}{l} \text{maximize} \\ g(t) \end{array} \right\} \sum_{i=1}^n \log g(t_i) - \int_0^T g(\tau) d\tau \left. \vphantom{\sum_{i=1}^n} \right\} \begin{array}{l} \text{Design } g(t) \text{ such that} \\ \text{max. likelihood is } \mathbf{convex} \\ \text{(and use CVX)} \end{array}$$

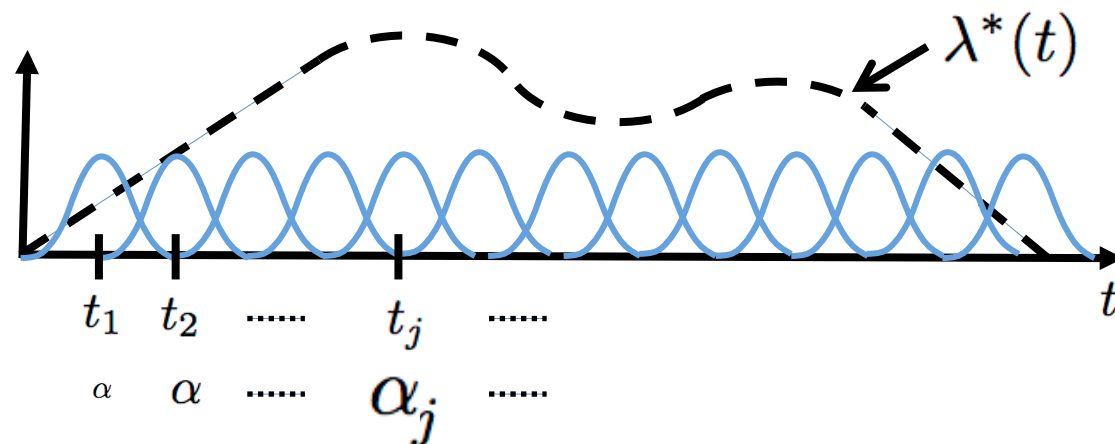
Nonparametric inhomogeneous Poisson process



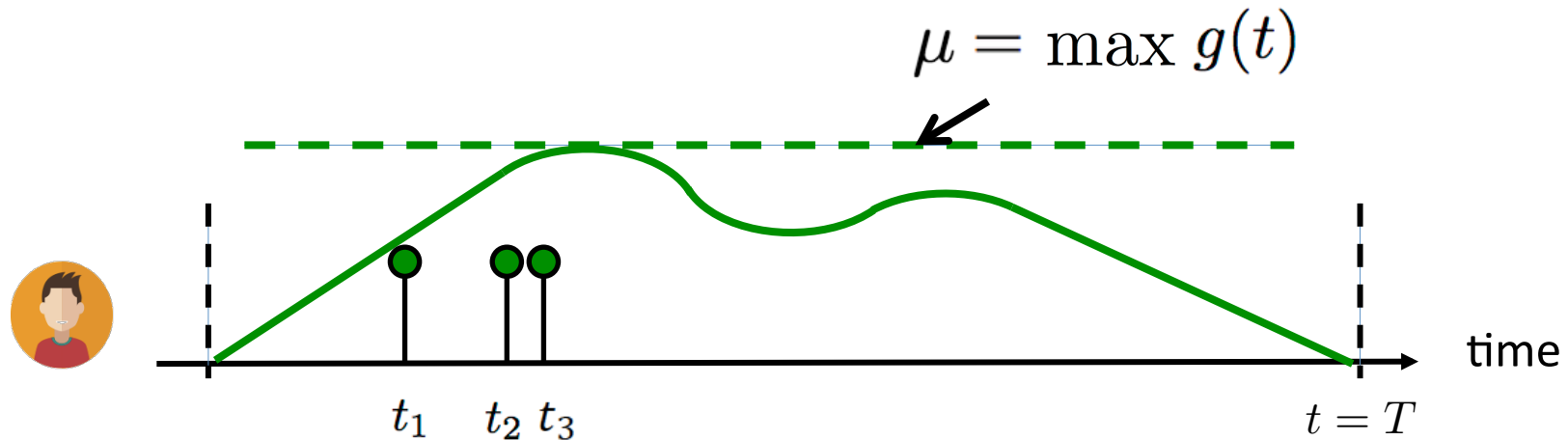
Positive combination of (Gaussian) RFB kernels:

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

The equation is followed by a diagram. A horizontal line represents the time axis. A bracket above it spans a range of time, with an arrow pointing to a graph of a single Gaussian kernel $k(t - t_j)$ centered at t_j . This illustrates that the overall intensity function is a sum of such kernels.



Sampling from an inhomogeneous Poisson



Thinning procedure (similar to rejection sampling):

1. Sample t from Poisson process with intensity μ

$$t \sim -\frac{1}{\mu} \log(1 - u) + t_3$$

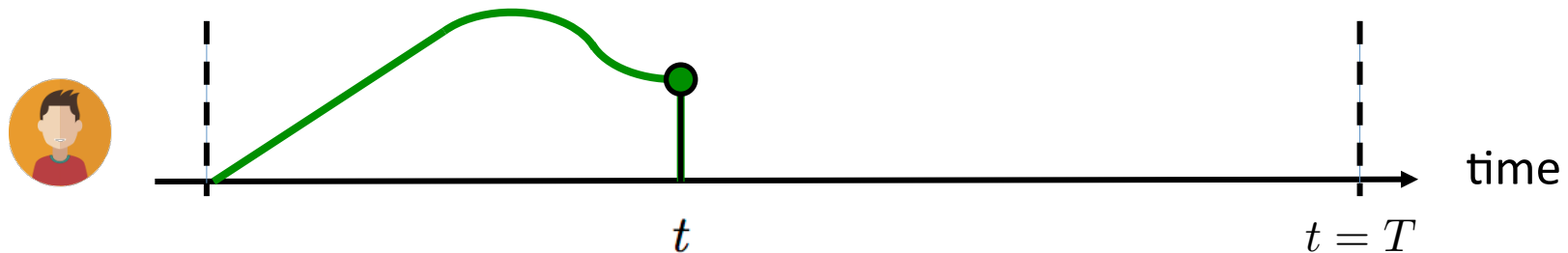
$Uniform(0, 1)$
↓

} Inversion 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

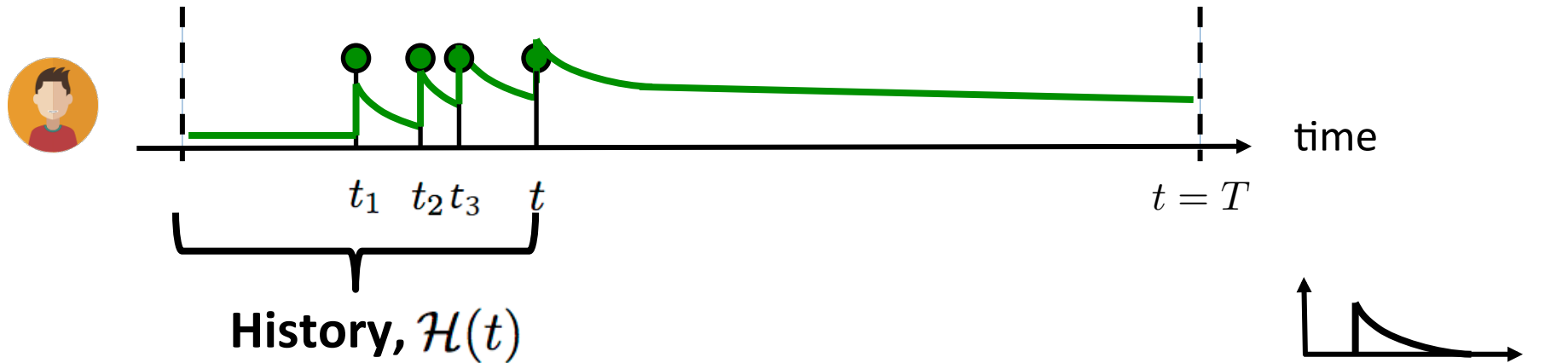
$$\lambda^*(t) = g^*(t)(1 - N(t)) \geq 0$$

Observations:

1. Limited number of occurrences

*Try sampling
and fitting!*

Self-exciting (or Hawkes) process



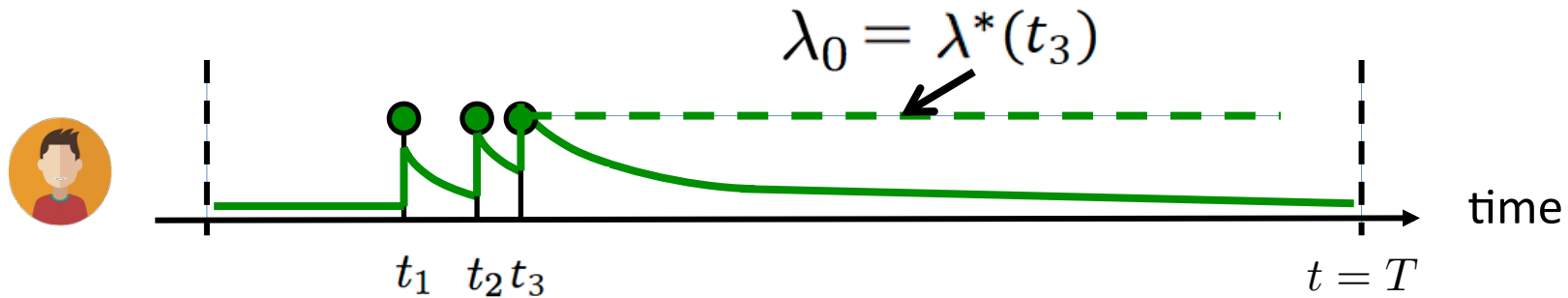
Intensity of self-exciting
(or Hawkes) process:

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

Observations:

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

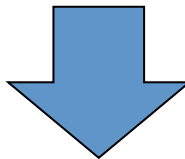
Fitting a Hawkes process from a recorded timeline



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

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

Maximum
likelihood

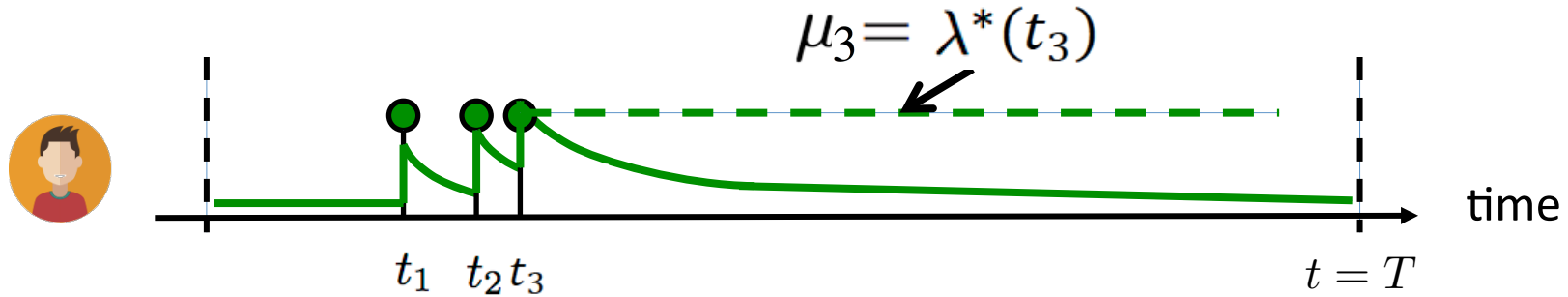


$$\text{maximize}_{\mu, \alpha} \sum_{i=1}^n \log \lambda^*(t_i) - \int_0^T \lambda^*(\tau) d\tau$$

The max. likelihood
is **jointly convex**
in μ and α

(use CVX!)

Sampling from a Hawkes process



Thinning procedure (similar to rejection sampling):

1. Sample t from Poisson process with intensity μ_3

$$t \sim -\frac{1}{\mu_3} \log(1 - u) + t_3$$

Uniform(0, 1)
↓

} Inversion sampling

2. Generate $u_2 \sim \text{Uniform}(0, 1)$

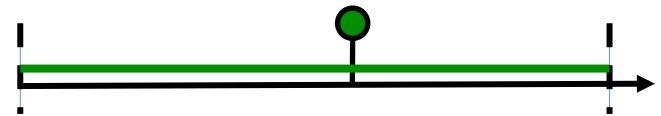
3. Keep the sample if $u_2 \leq g(t) / \mu_3$
- } Keep sample with prob. $g(t) / \mu_3$

Summary

Building blocks to represent different dynamic processes:

Poisson processes:

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



Inho

Term

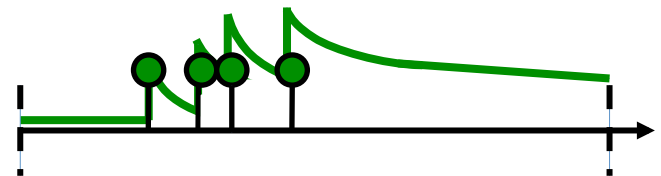
$$\lambda^*(t) = g(t)(1 - N(t))$$

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



Self-exciting point processes:

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

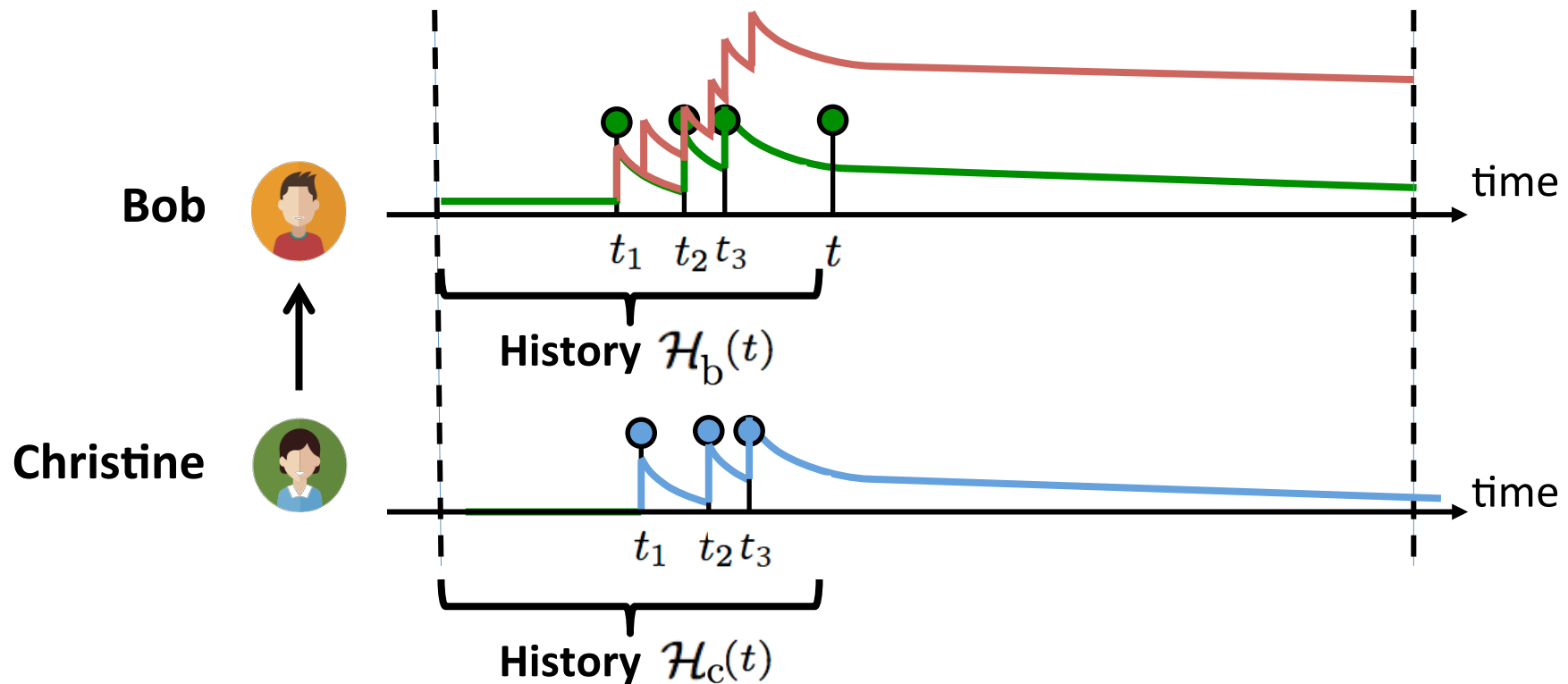


Representation:

Temporal Point Processes

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

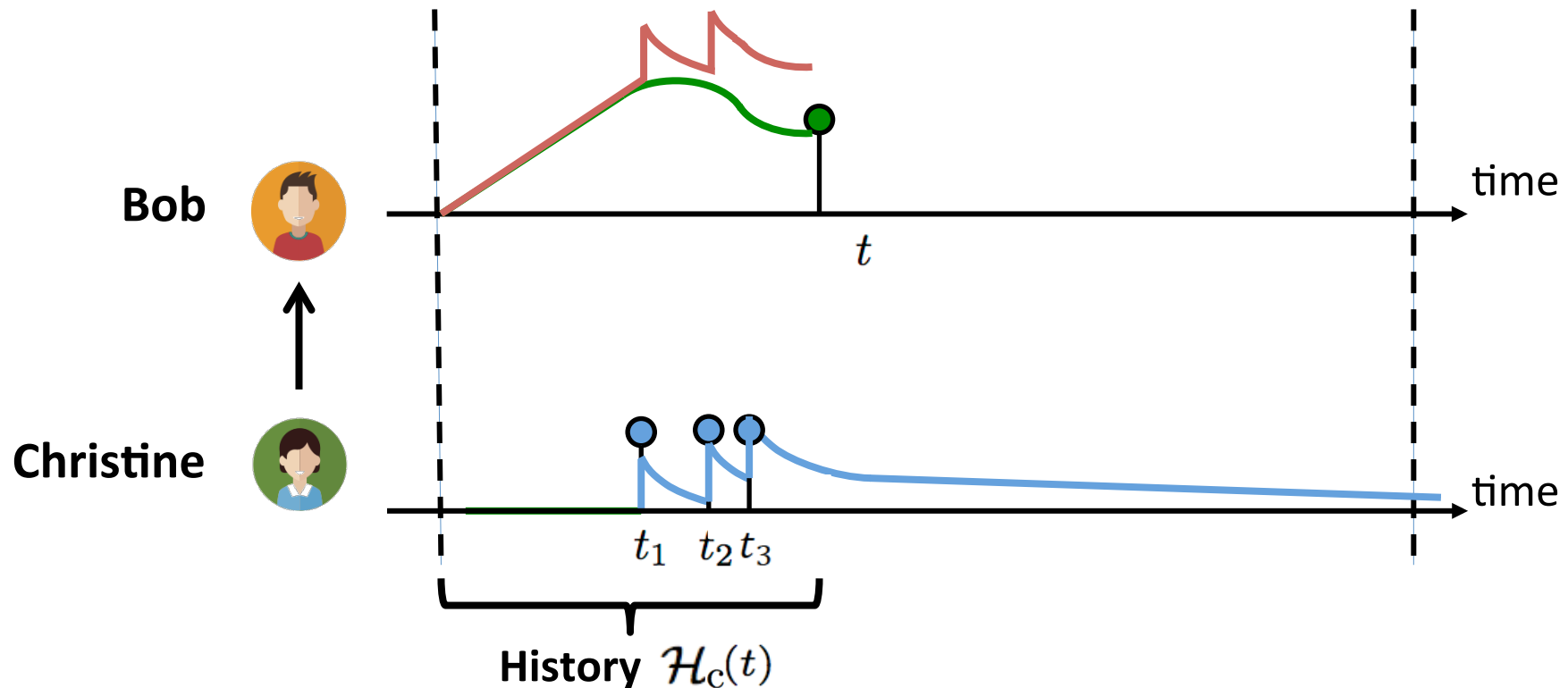
Mutually exciting process



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:

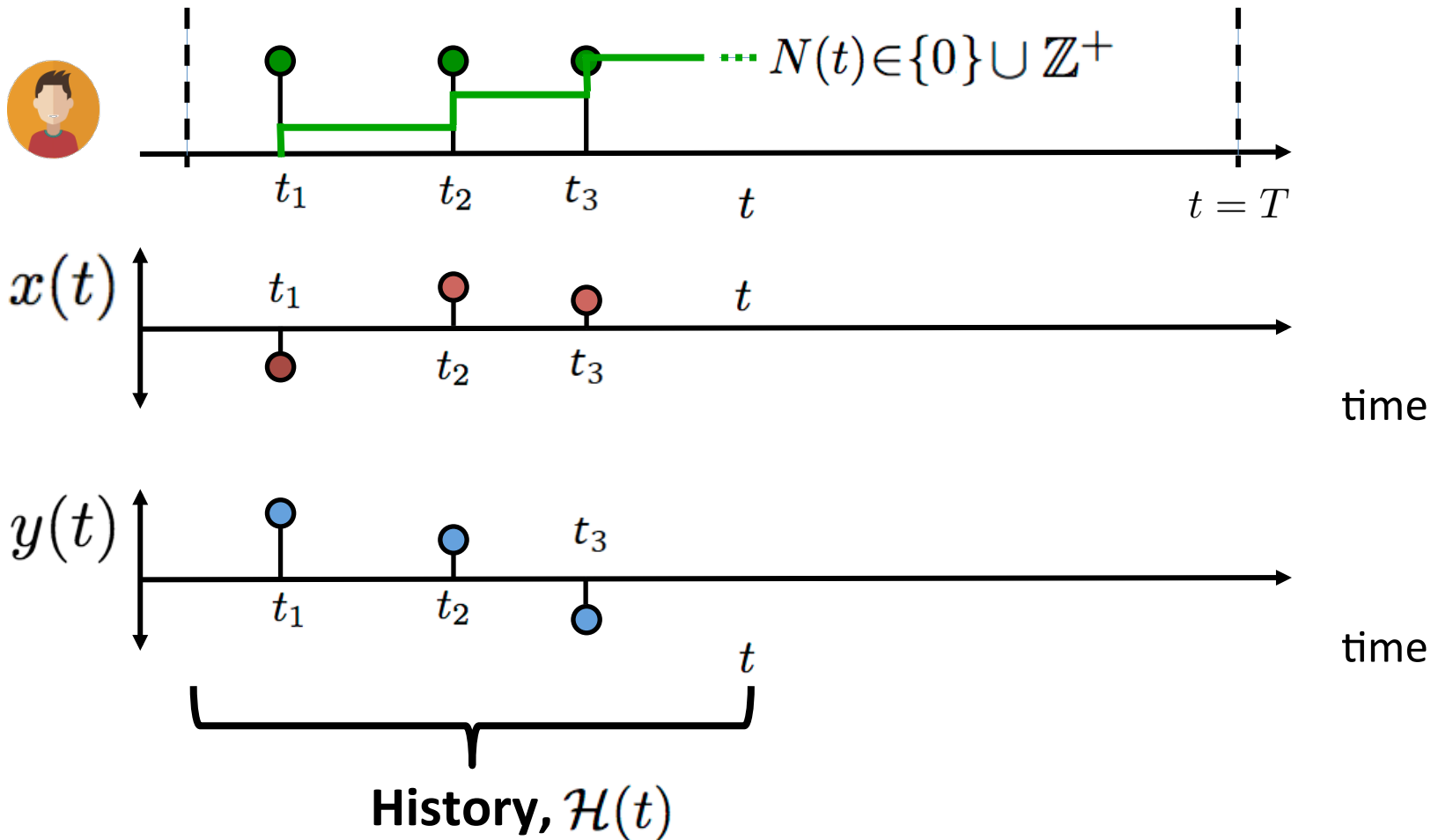
Temporal Point Processes

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

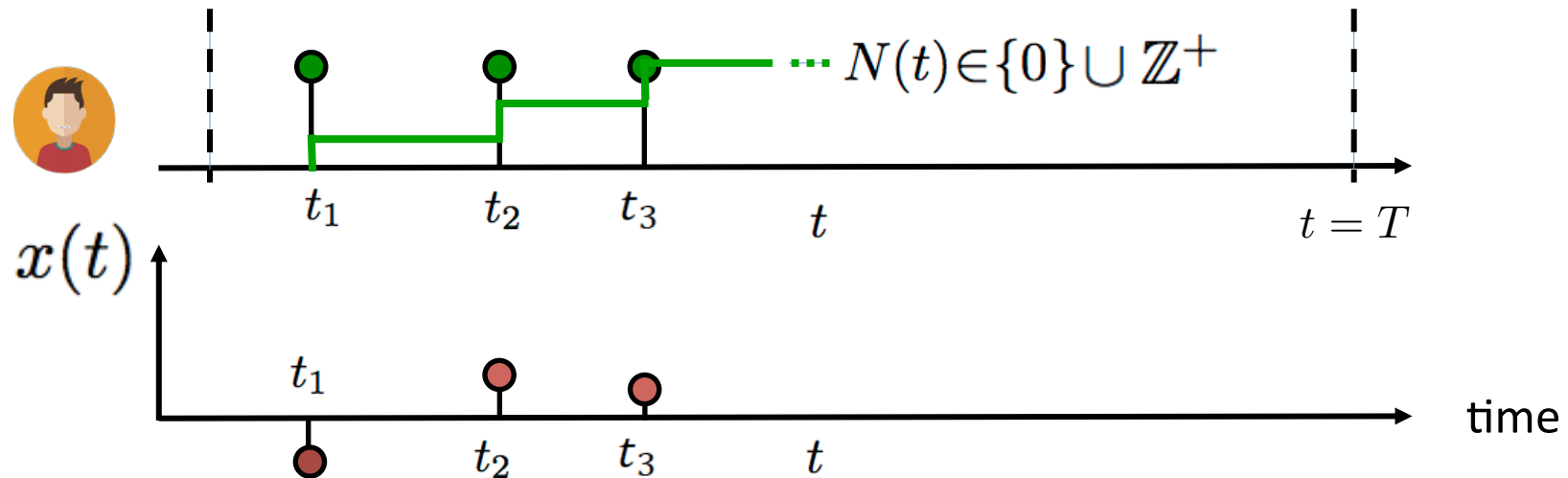
Marked temporal point processes

Marked temporal point process:

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



Independent identically distributed marks



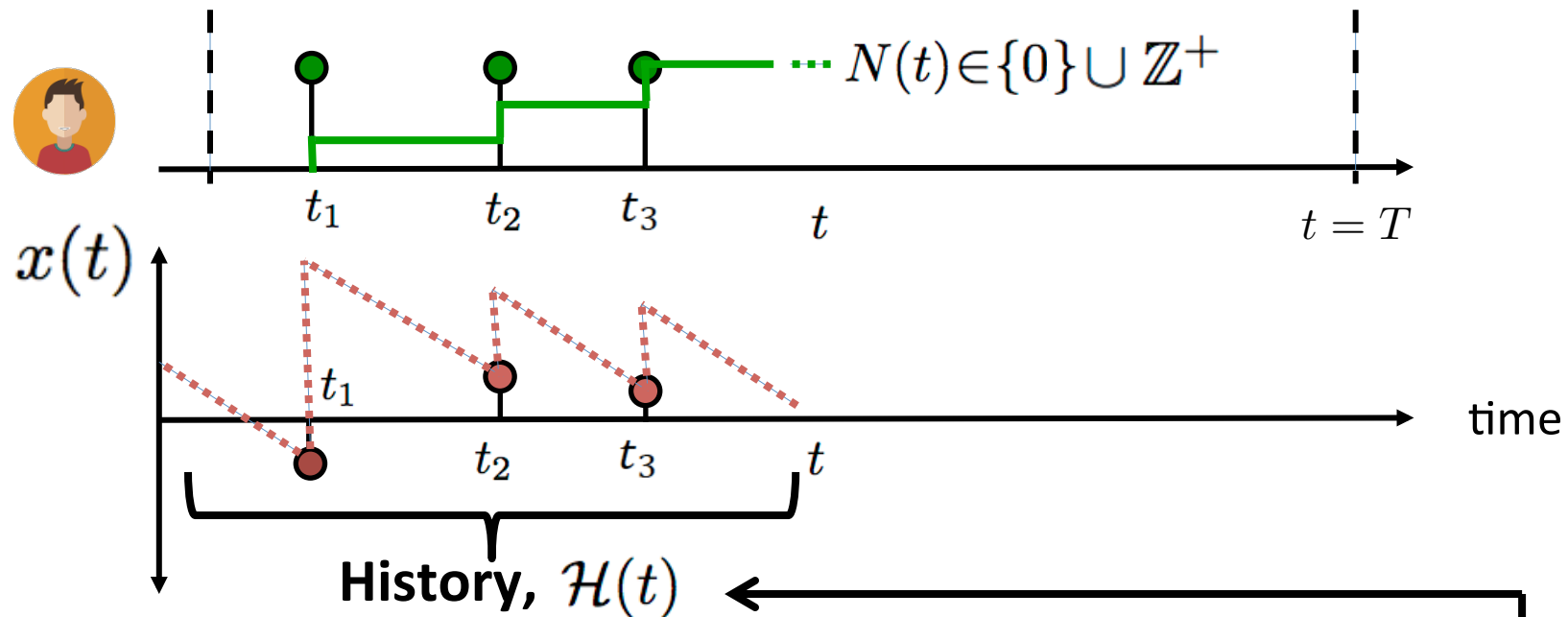
Distribution for the marks:

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

Observations:

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

Dependent marks: SDEs with jumps



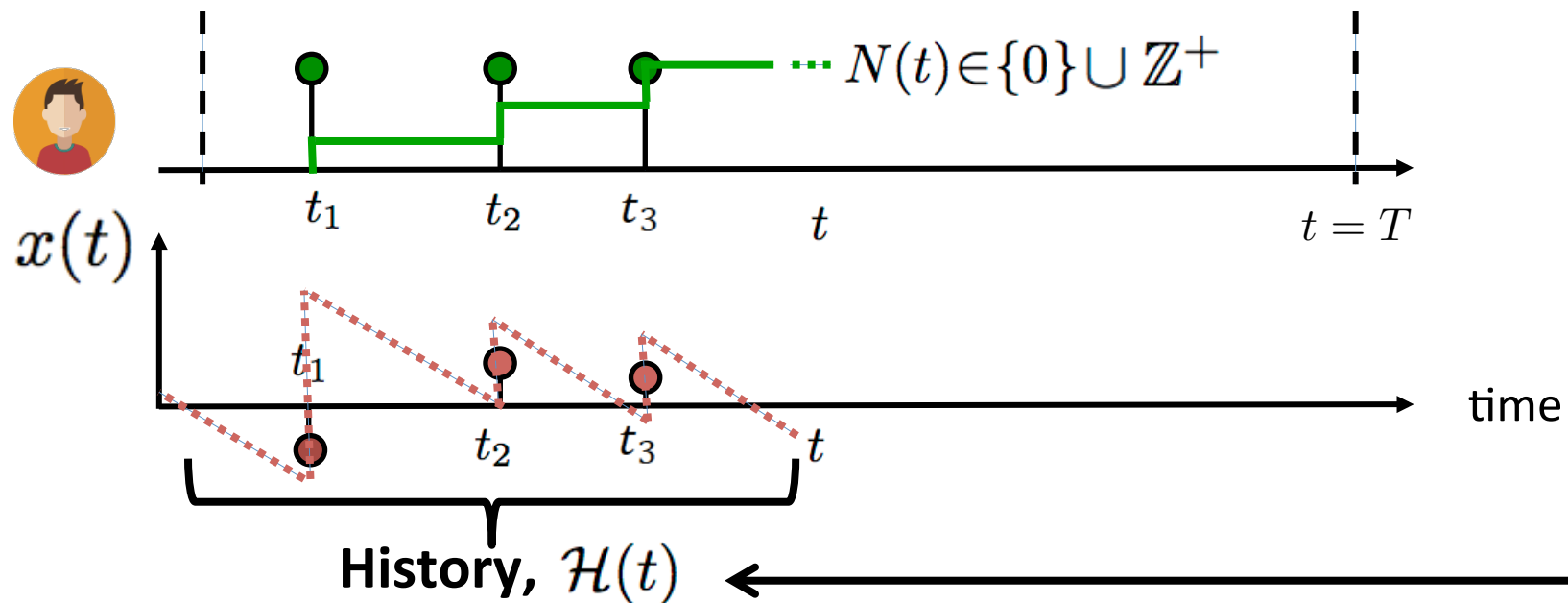
Marks given by stochastic differential equation with jumps:

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

Observations:

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

Dependent marks: distribution + SDE with jumps



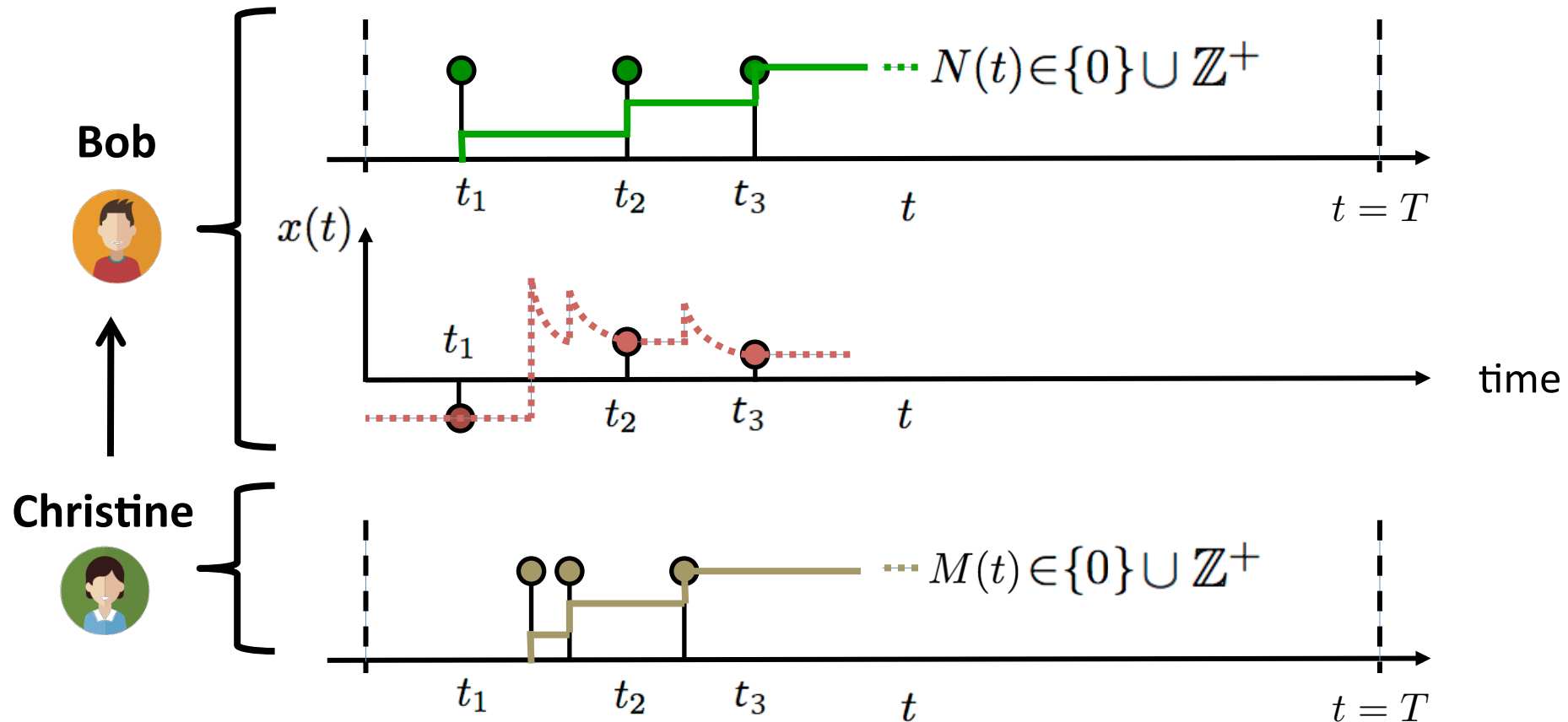
Distribution for the marks:

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

Observations:

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

Mutually exciting + marks



Marks affected by neighbors

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

REPRESENTATION: TEMPORAL POINT PROCESSES

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

APPLICATIONS: MODELS

1. Information propagation
2. Information reliability
3. Knowledge acquisition

APPLICATIONS: CONTROL

1. Activity shaping
2. When-to-post

Next