Time series in discrete events

Their differences and features of modelling

Maximum likelihood method

Skoltech

Based on ICML Tutorial, July 2018 by I. Valera & M.G. Rodriguez

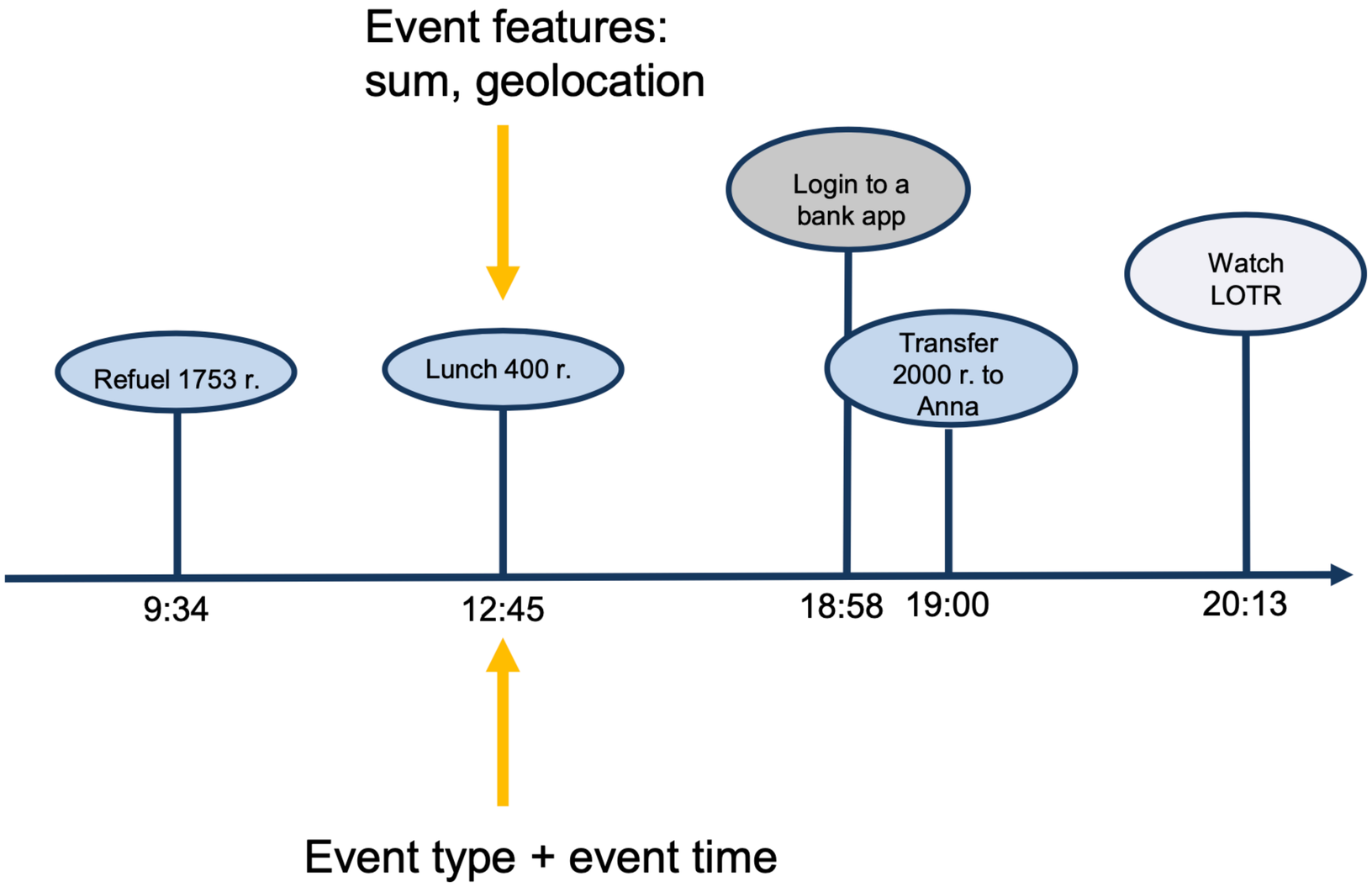
Slides/references: http://learning.mpi-sws.org/tpp-icml18

Classical TS vs TS in discrete events

Skoltech

Example: financial transactions are event sequences

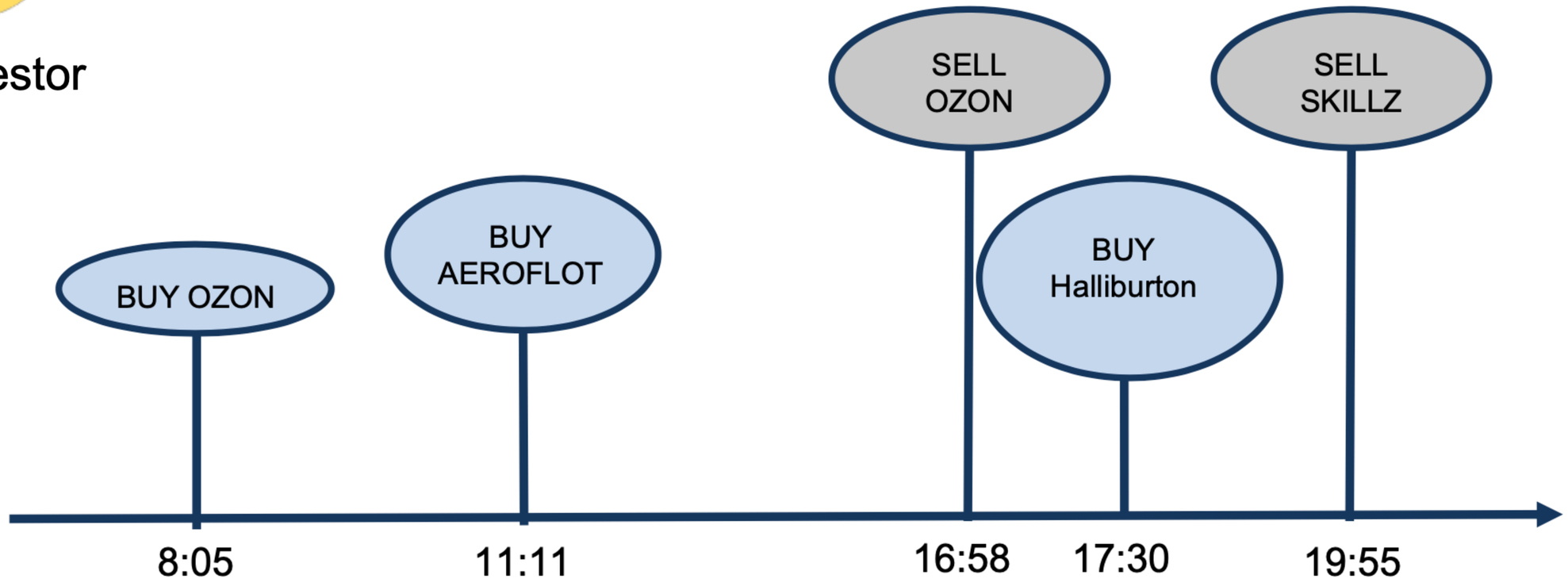
Alex, 33, man
works in retail



Skoltech

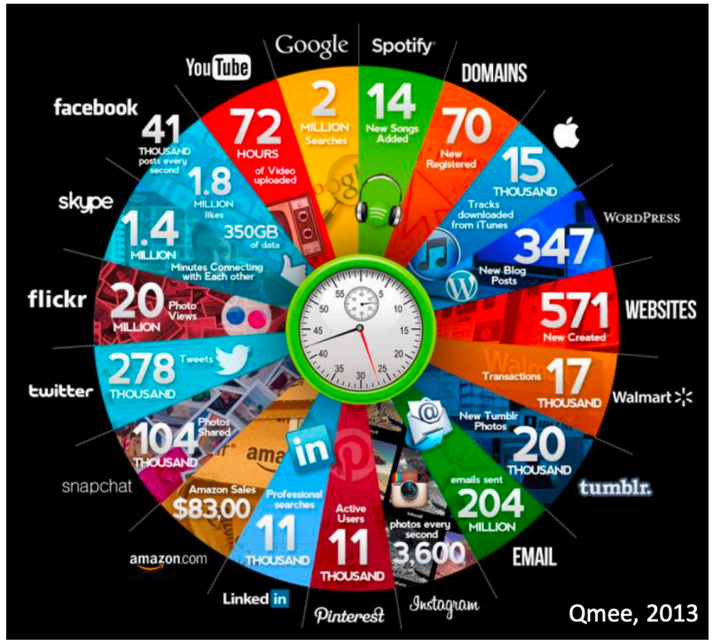
Example: operations in markets are event sequences

An investor

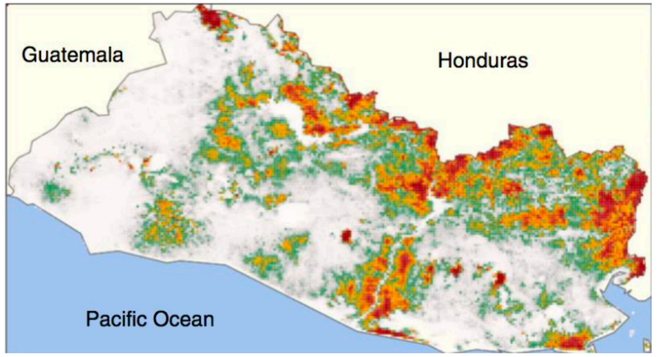


Skoltech

Many discrete *events* in continuous time

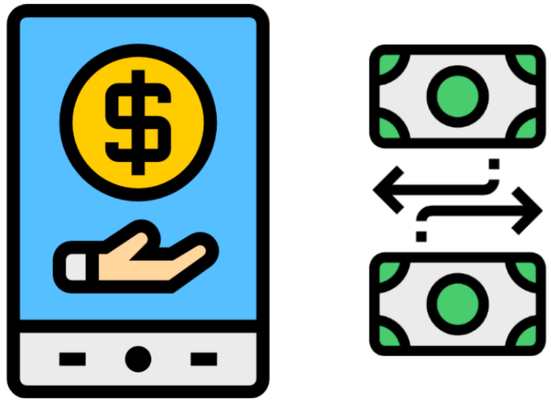


Online actions



Disease dynamics

Financial trading



Financial transactions

Skoltech

Variety of processes behind these events

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

News spread in Twitter

Reviews and sales in Amazon

Stock trading

Flu spreading

Article creation in Wikipedia

Ride-sharing requests

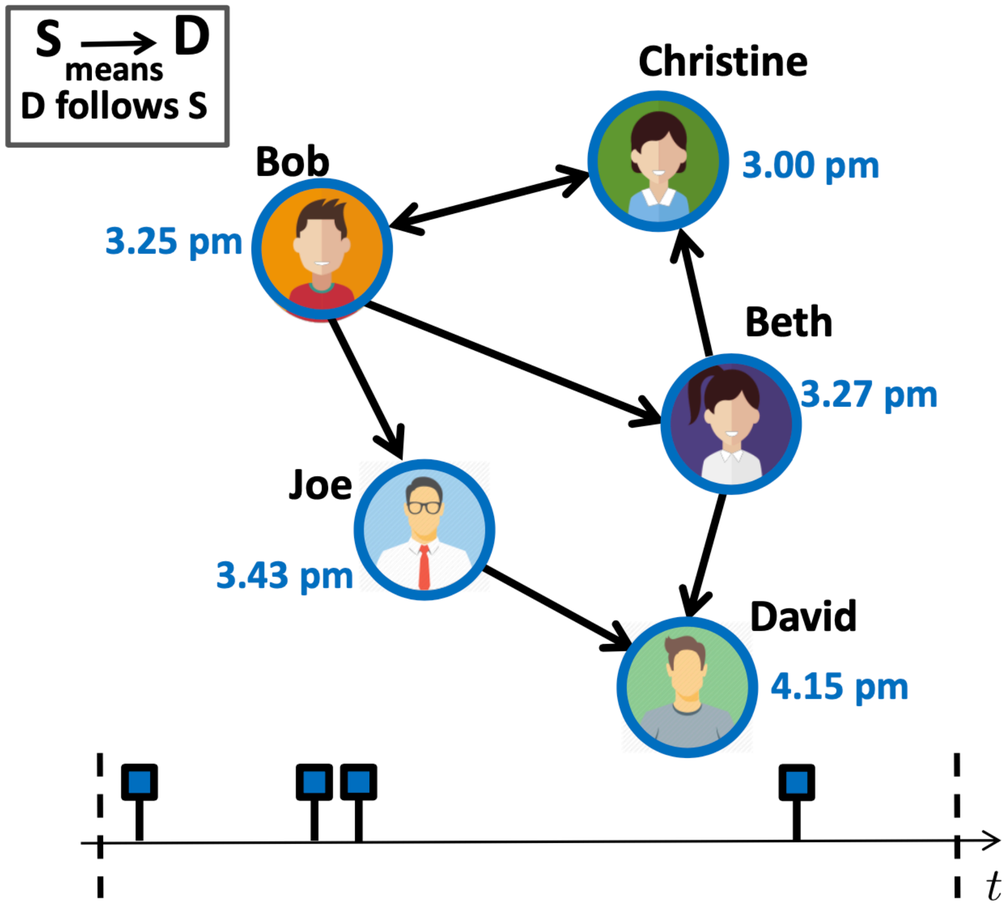
FAST

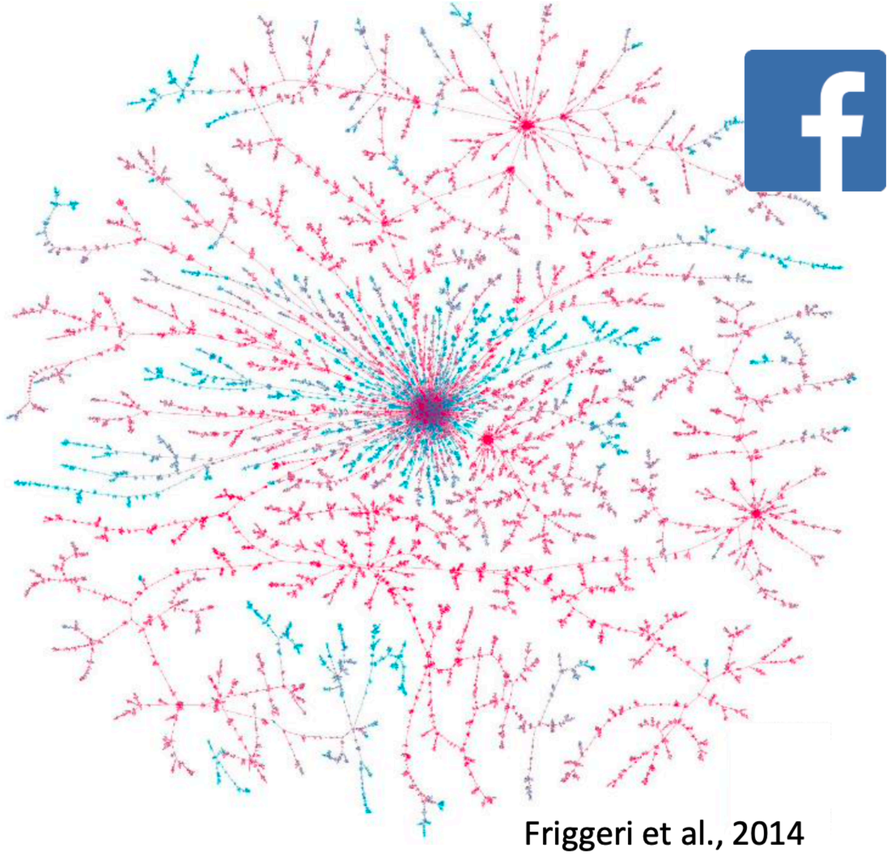
SLOW

...in a wide range of temporal scales

Skoltech

Example 1: information propagation





Skoltech

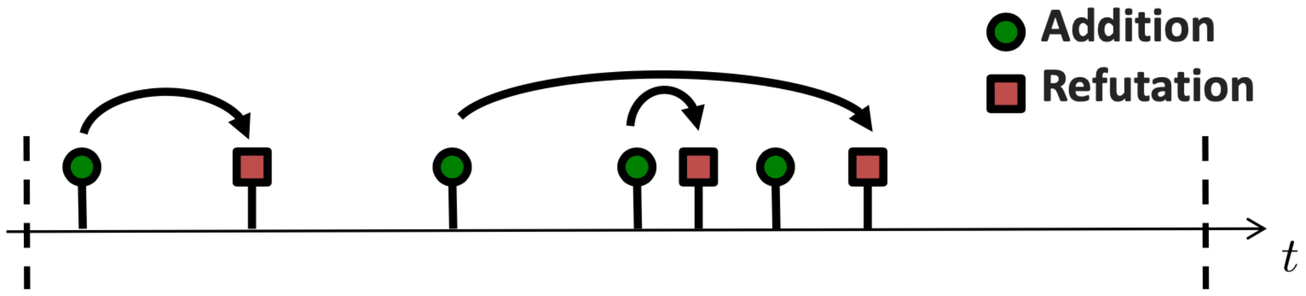
Example 2: response history

03:21, 20 September 2016

is a Kenyan politician

possible vandalism by MLM2016

is an American politician

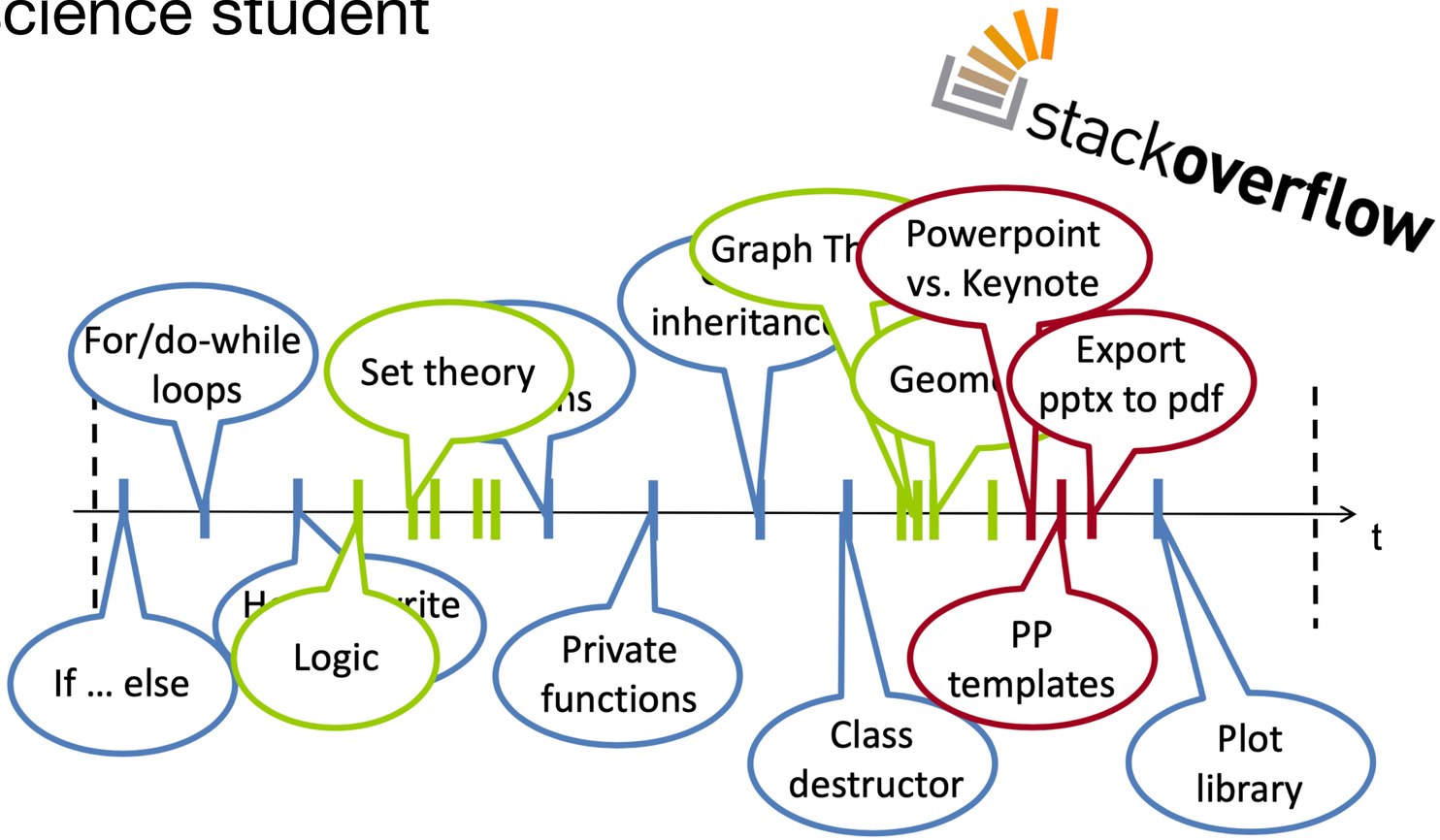


Skoltech

Example 3: development

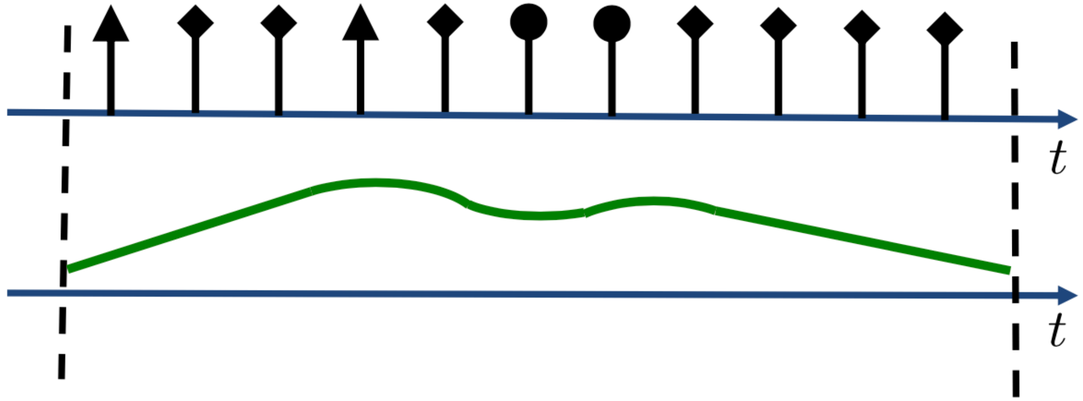
1st year computer science student

- Introduction to programming
- Discrete math
- Project presentation

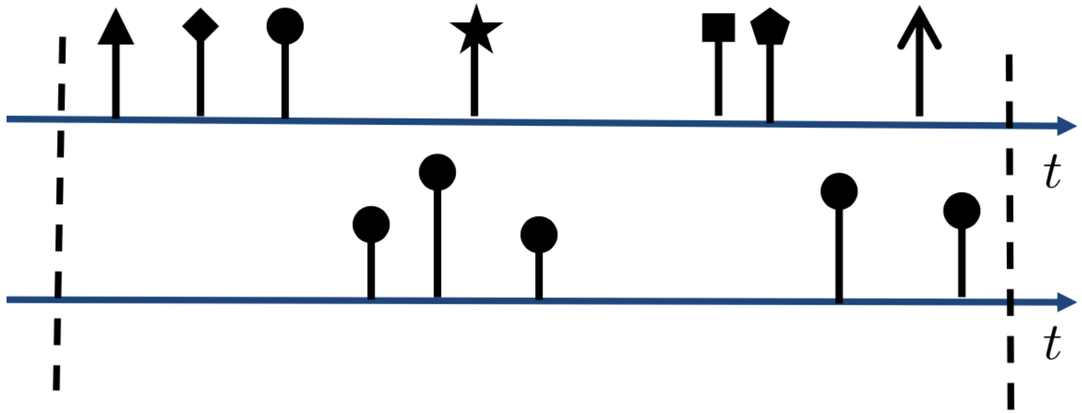


Skoltech

Aren’t these event traces just time series?



Discrete and continuous TS

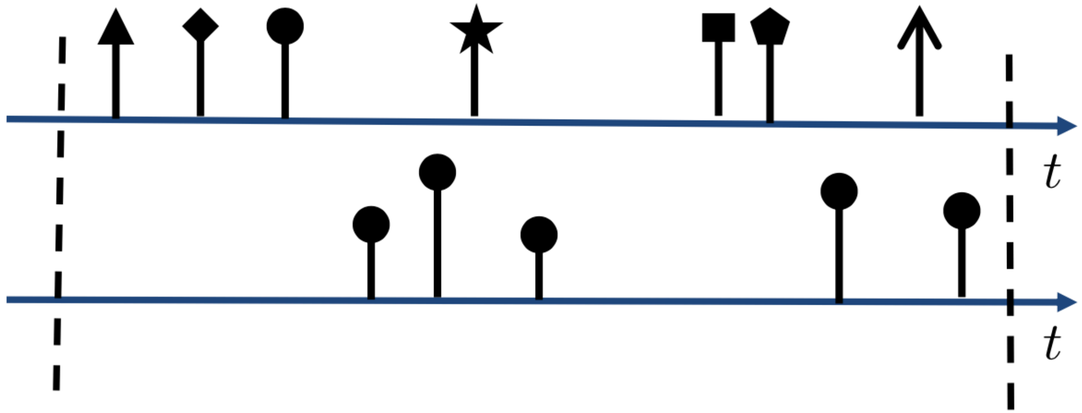


Discrete events in continuous time

Skoltech

Problems for event sequences

- Compact description of data: models
- Interpretation
- Forecasting/Prediction
- Control
- Hypothesis testing
- Simulation



Discrete events in continuous time

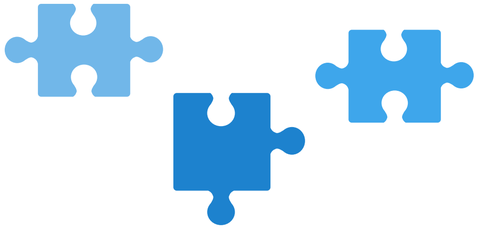
Skoltech

Temporal Point Processes (TPPs)

Skoltech

Motivation (1)

Humans & nature generate …



irregular & asynchronous event sequences

- high-frequency financial transactions
- gene positions in bioinformatics
- earth- quakes geographic

asynchronicity

multi-modality

cross-correlation

Source: https://arxiv.org/pdf/1907.07561.pdf

Motivation (2)

asynchronicity ⇨ multiple events happening in the continuous time domain are sampled with unequal intervals

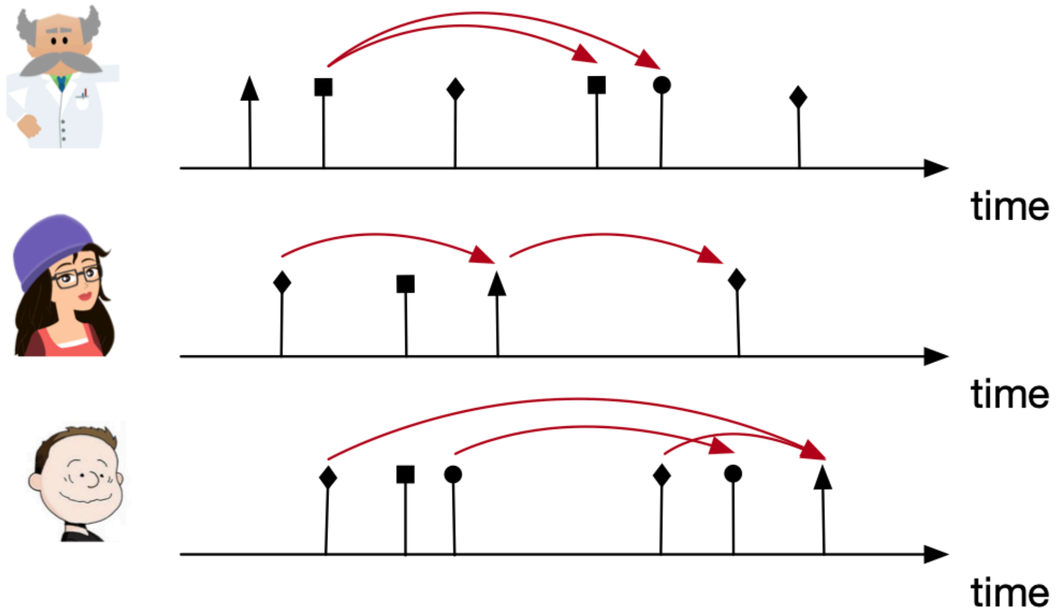
multi-modality ⇨ sequences contain multiple type of events

cross-correlation ⇨ occurrence of one type of event at a certain time can excite/inhibit the happening of future events of the same/another type

Source: https://arxiv.org/pdf/1907.07561.pdf

Motivation (3)

Different types of actions of 3 users. The filled dark symbols in (a) mean 4 action types while the red arrows denote actions influencing another actions. (b) indicates the influence of the column event type j on the row type i on future events.



(a)

	■	▲	◆	●
■	✓			✓
▲			✓	
◆		✓		
●				✓

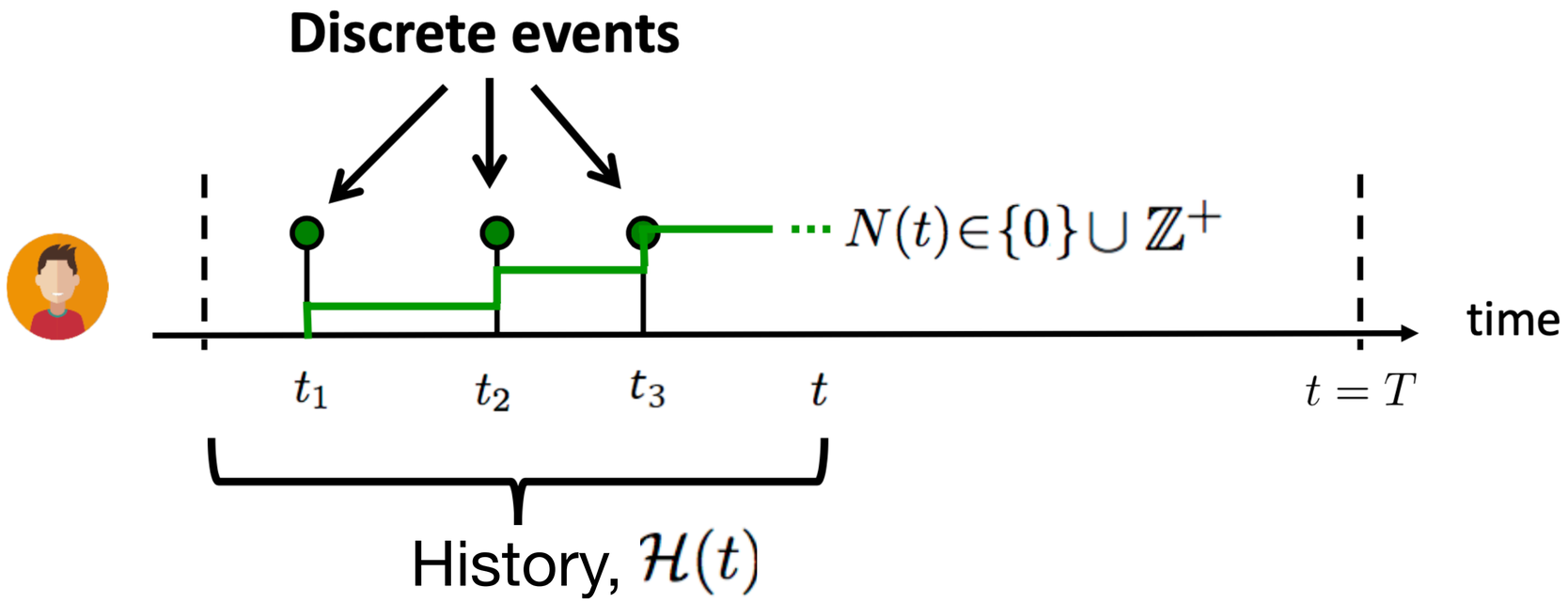
(b)

- ***Problem:*** predict *which* type and *when* future events will happen

Source: https://arxiv.org/pdf/1907.07561.pdf

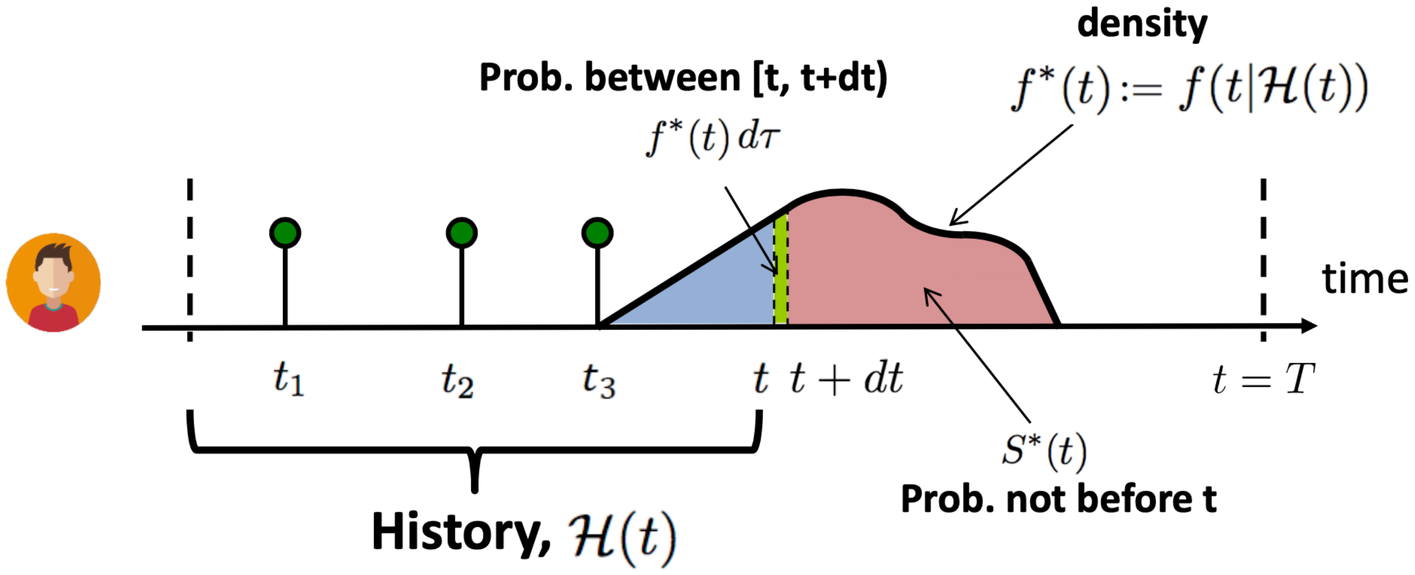
Temporal point processes (TPPs)

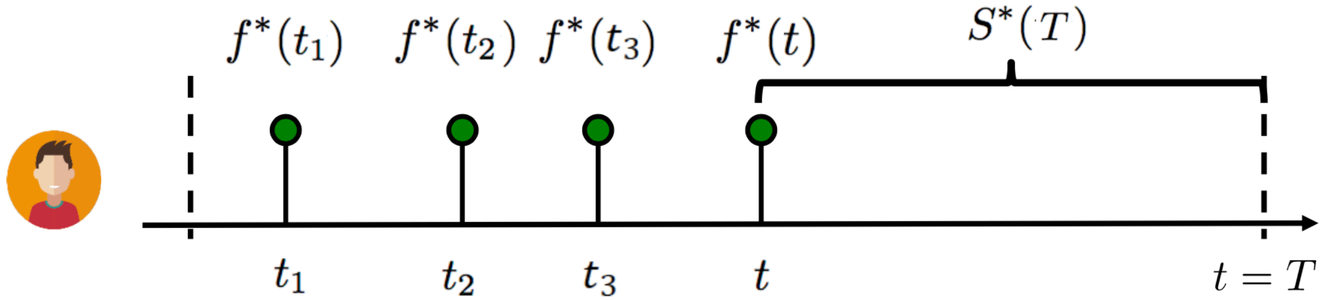
- The occurrence of asynchronous event sequences are often modelled by temporal point processes
- ***Temporal point process*** — a random process whose realization consists of discrete events localised in time $\mathcal{H} = \{t_i\}$



Source: https://courses.mpi-sws.org/hcml-ws18/lectures/TPP.pdf

Model time as a random variable



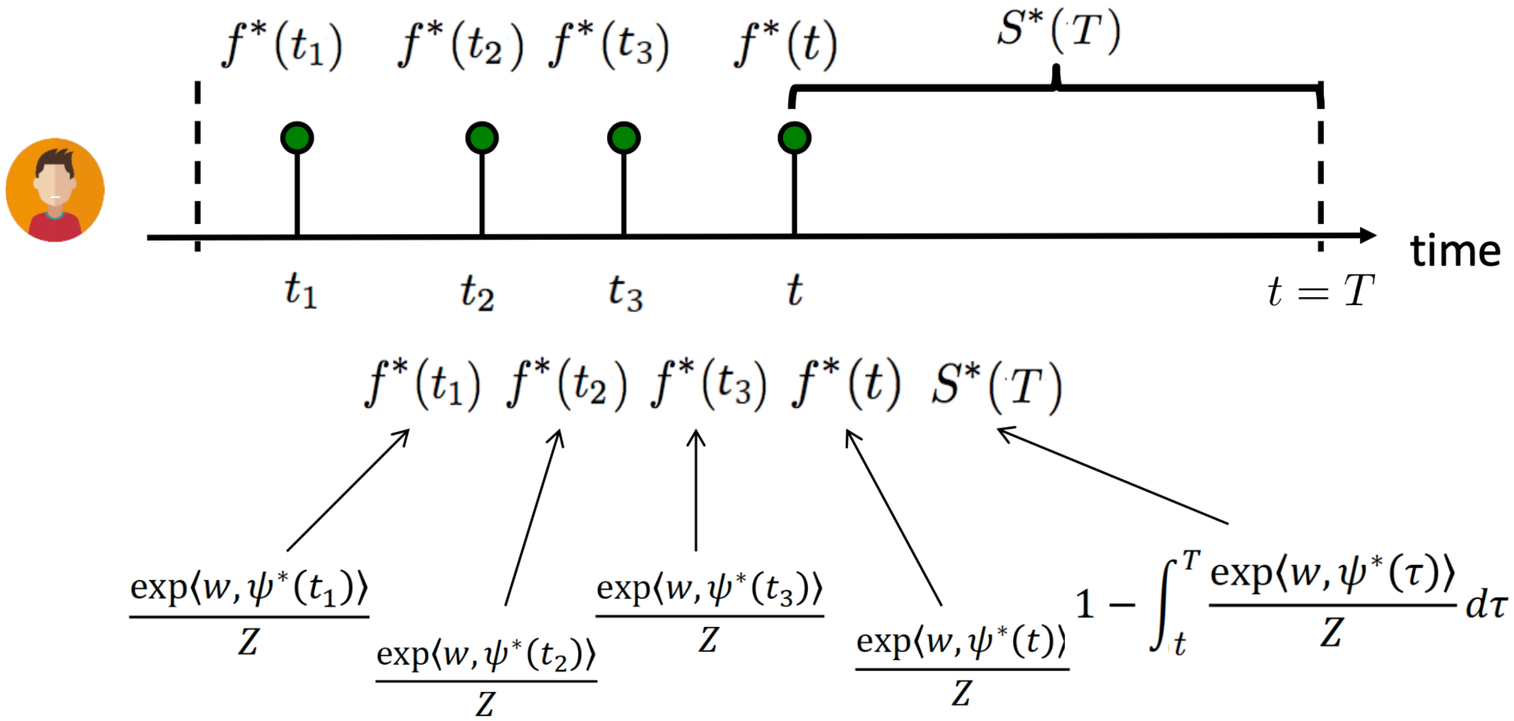


Likelihood of a timeline:

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

Skoltech

Problems of density parametrization

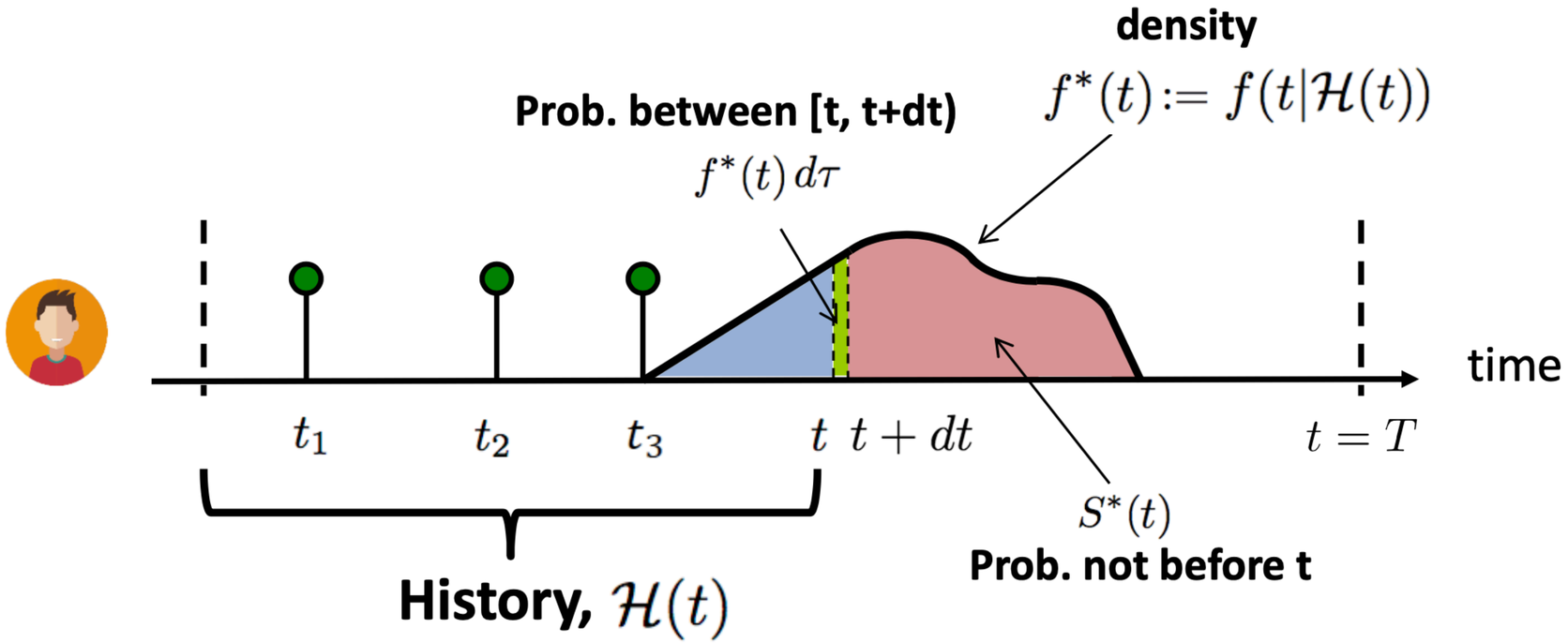


It is difficult for model design and interpretability:

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

Skoltech

Intensity function

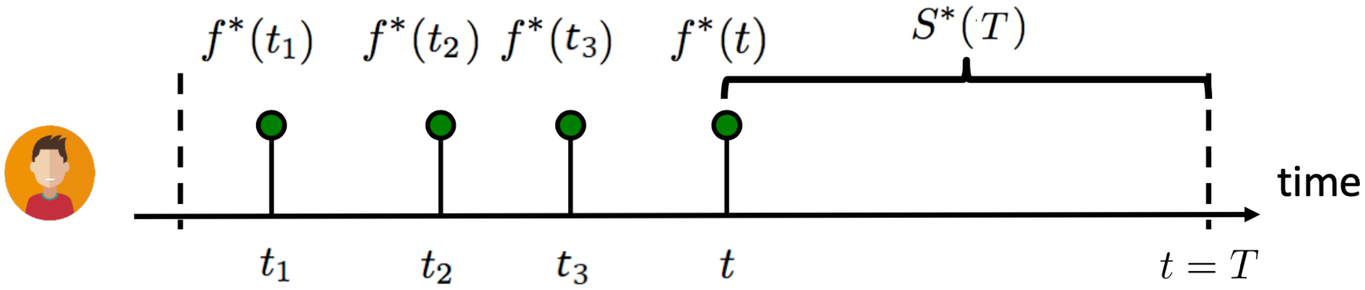


Intensity: probability between $[t, t + dt)$ but not before t

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

Skoltech

Advantages of intensity parametrization



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

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

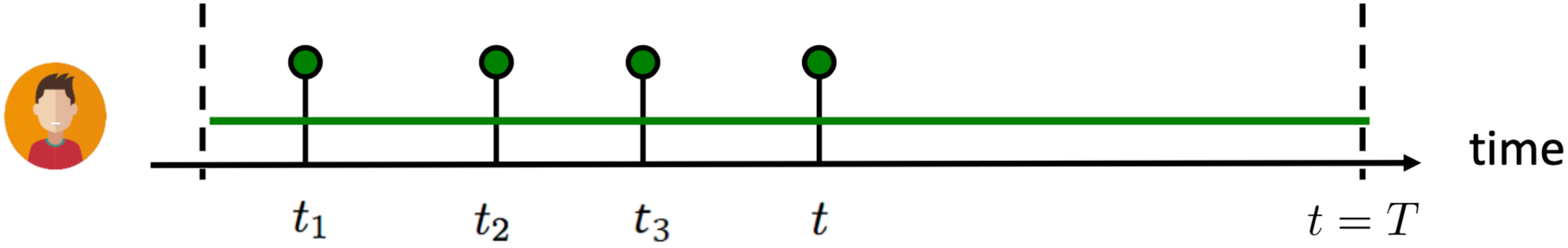
Suitable for model design and interpretable:

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

Skoltech

Poisson process

- Coarse approximation of many real-life processes



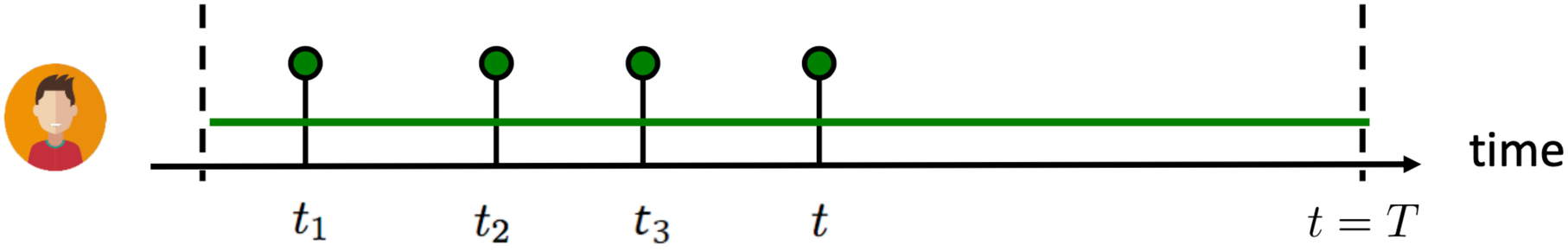
- Intensity of a Poisson process

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

- 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:

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

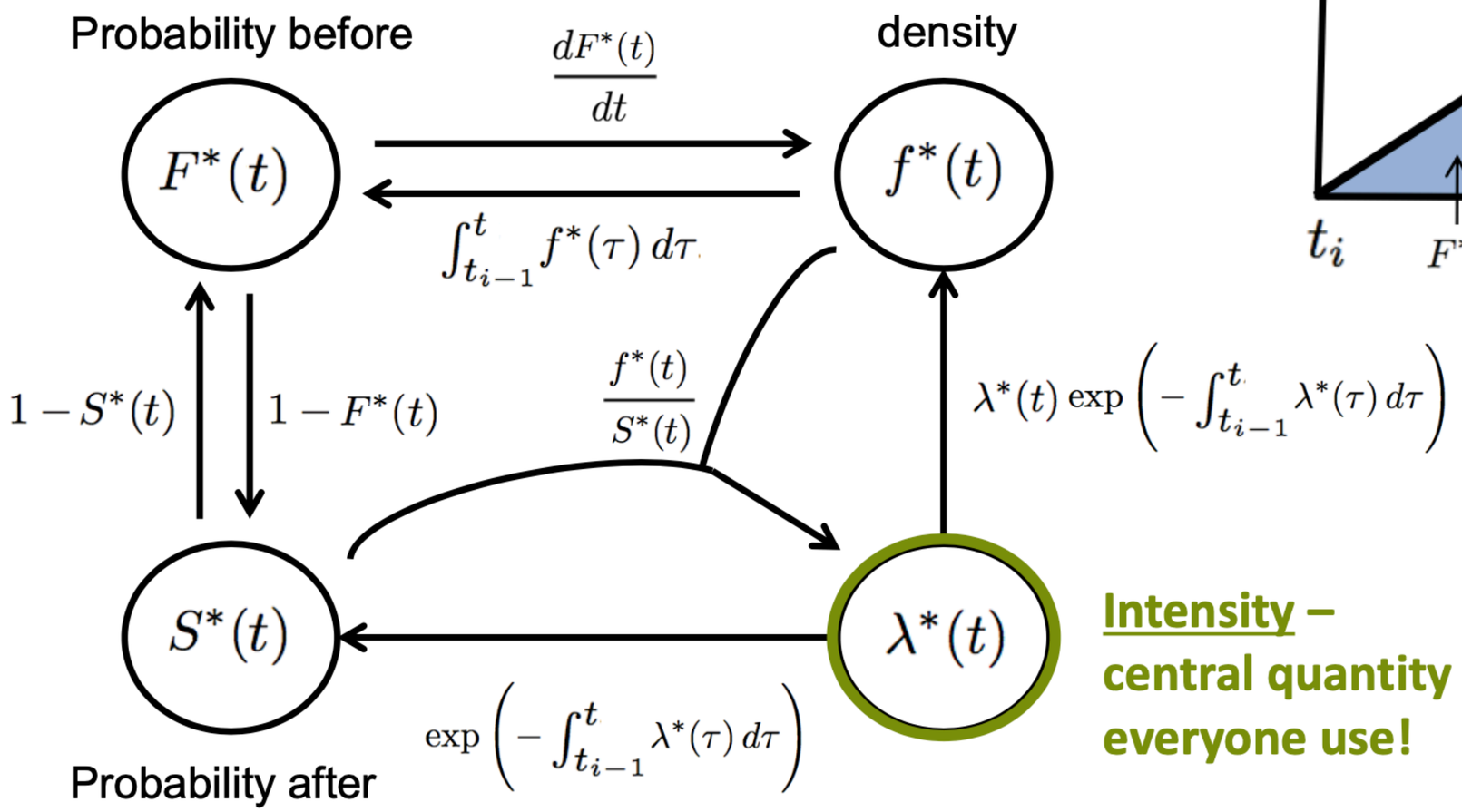
- Sampling using inversion sampling:

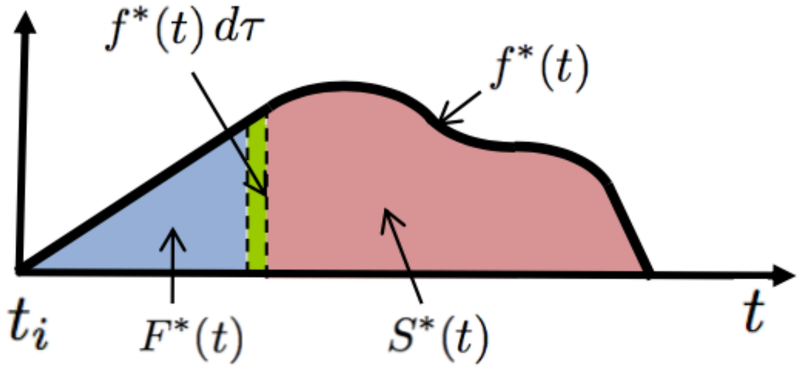
$$t \sim \underbrace{\mu \, \exp\left(-\mu \left(t - t_3\right)\right)}_{f_t^*(t)} \quad \Rightarrow \quad t = \underbrace{-\frac{1}{\mu} \log(1 - \overset{Uniform(0,1)}{u}) + t_3}_{F_t^{-1}(u)}$$

21

Skoltech

Relation between f*, F*, S*, λ*





Intensity – central quantity everyone use!

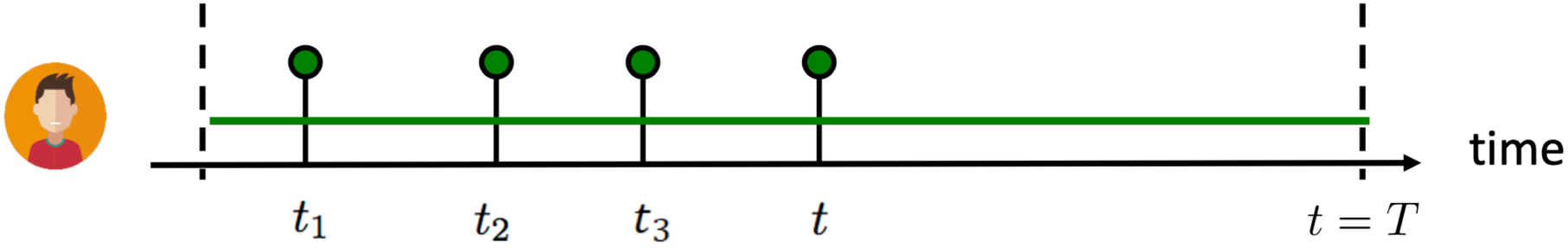
Skoltech

Basic building blocks

Skoltech

Poisson process

- Coarse approximation of many real-life processes



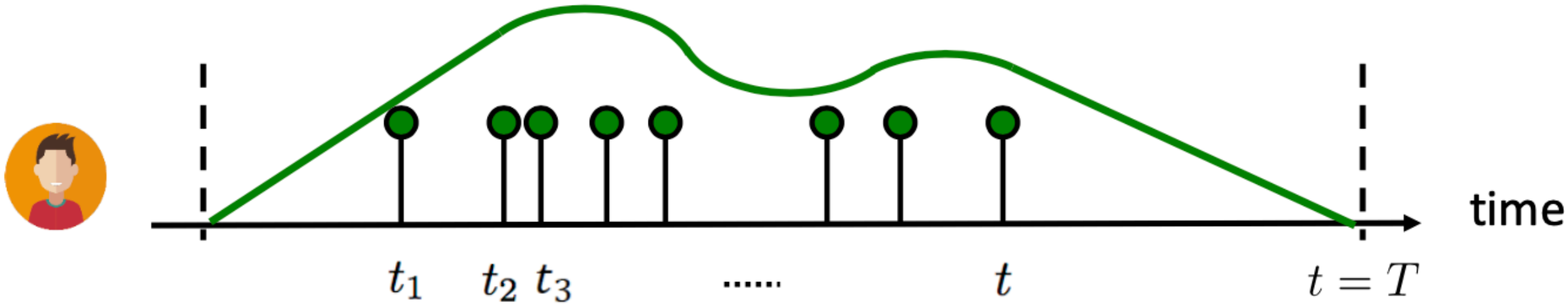
- Intensity of a Poisson process:

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

- Observations:

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

Inhomogeneous Poisson process

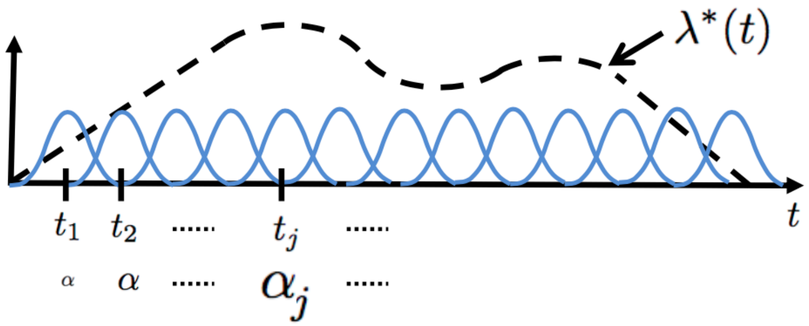


- Intensity of an inhomogeneous Poisson process:

$$\lambda^*(t) = g(t) \geq 0$$ – *independent of history*

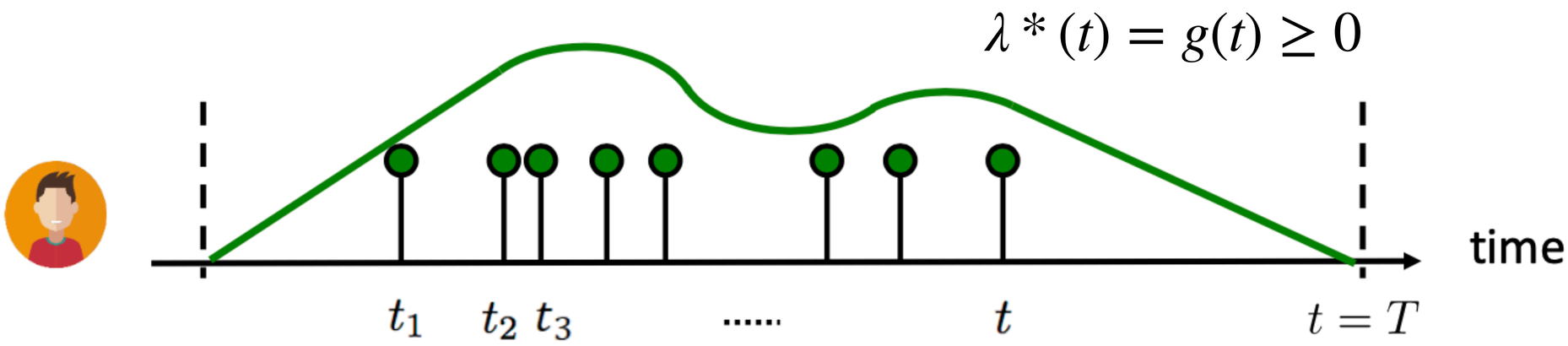
- Example:

$$\lambda^*(t) = \sum_j \alpha_j \overbrace{k(t - t_j)}^{t_j \quad t}$$



Skoltech

Fitting & sampling from inhomogeneous Poisson



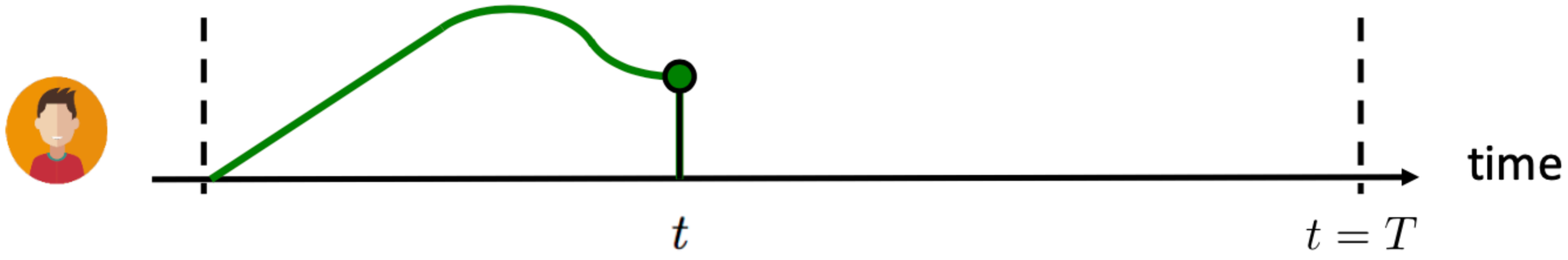
- Fitting by maximum likelihood: $\underset{g(t)}{\text{maximize}} \sum_{i=1}^{n} \log g(t_i) - \int_0^T g(\tau)\, d\tau$
- Sampling using thinning (reject. sampling) + inverse sampling*:

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. $g(t)/\mu$

Skoltech

Terminating (or survival) process



- Intensity of a terminating (or survival) process:

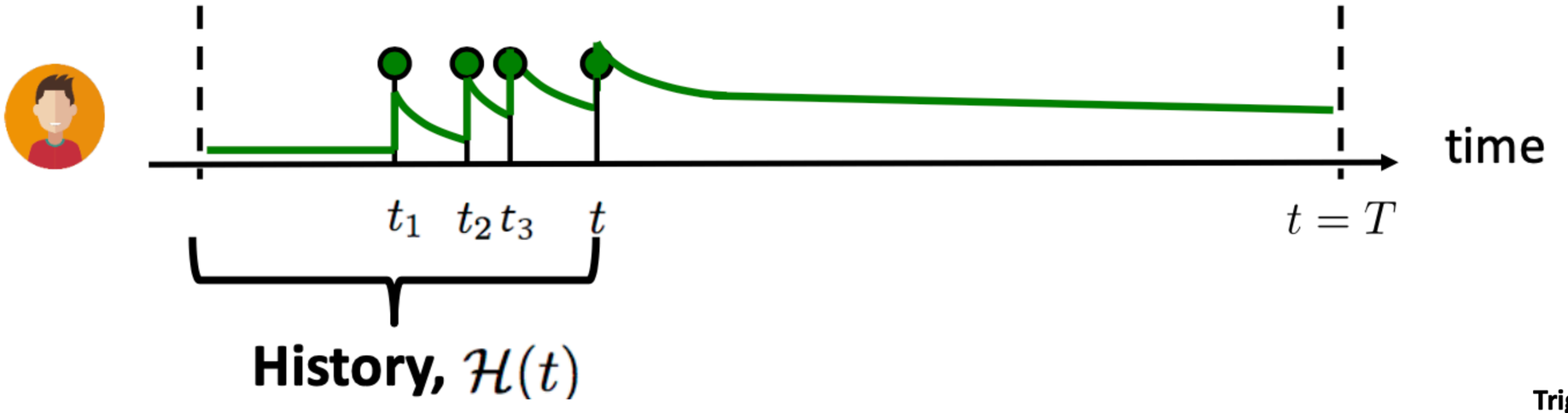
$$\lambda^*(t) = g^*(t)(1 - N(t)) \geq 0$$

- Observations:
  1. Limited number of occurrences

Try sampling and fitting...

Skoltech

Self-exciting Hawkes process



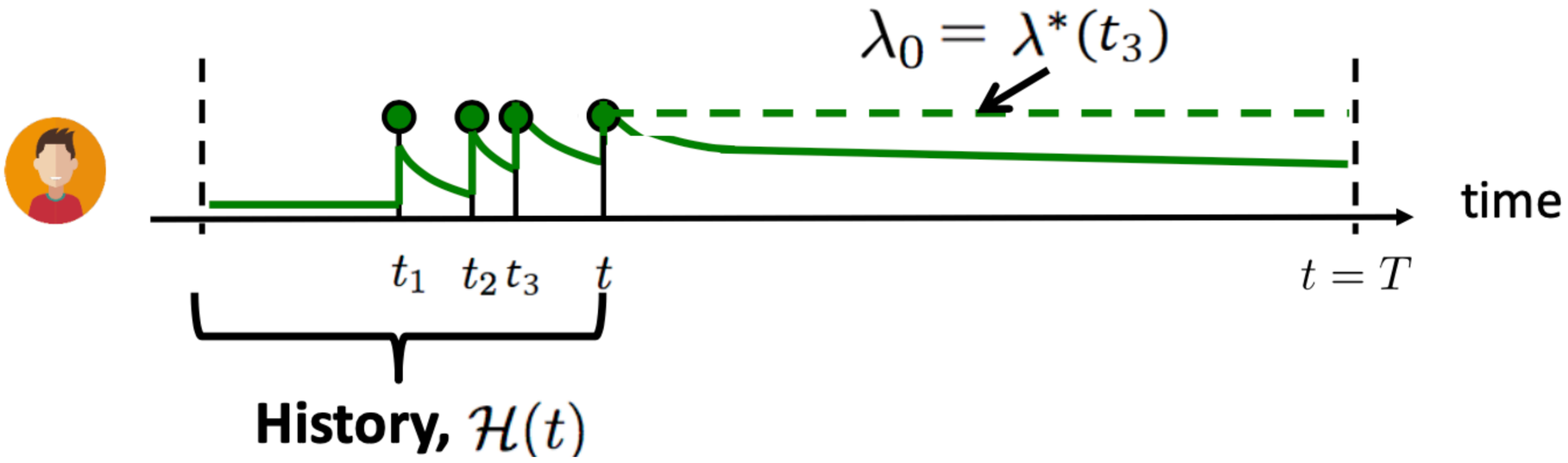
- Intensity of of self-exciting (or Hawkes) process:

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

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

Skoltech

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

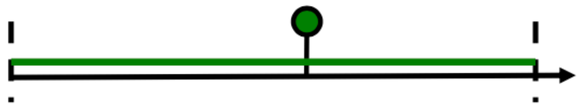
Skoltech

Summary

- Building blocks to represent different dynamic processes:

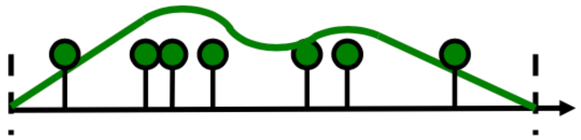
Poisson processes:

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



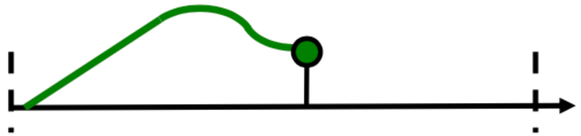
Inhomogeneous Poisson processes:

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



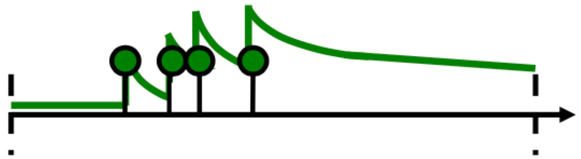
Terminating point processes:

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



Self-exciting point processes:

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



Skoltech