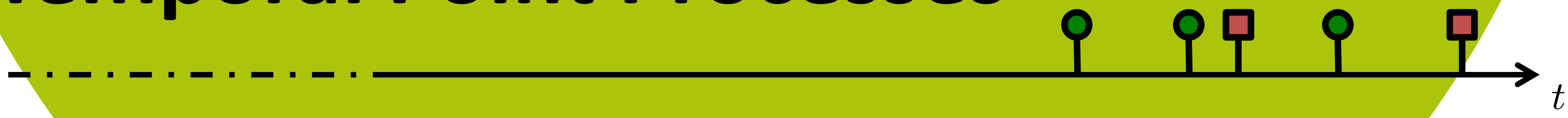
Temporal Point Processes



Alexey Zaytsev

Assistant professor, Skoltech

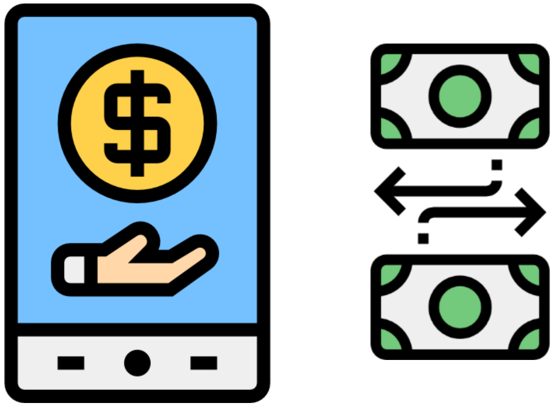
13th of December

Based on ICML Tutorial, July 2018 by I. Valera & M.G. Rodriguez
Slides/references: http://learning.mpi-sws.org/tpp-icml18

Temporal Point Processes (TPPs): applied problems

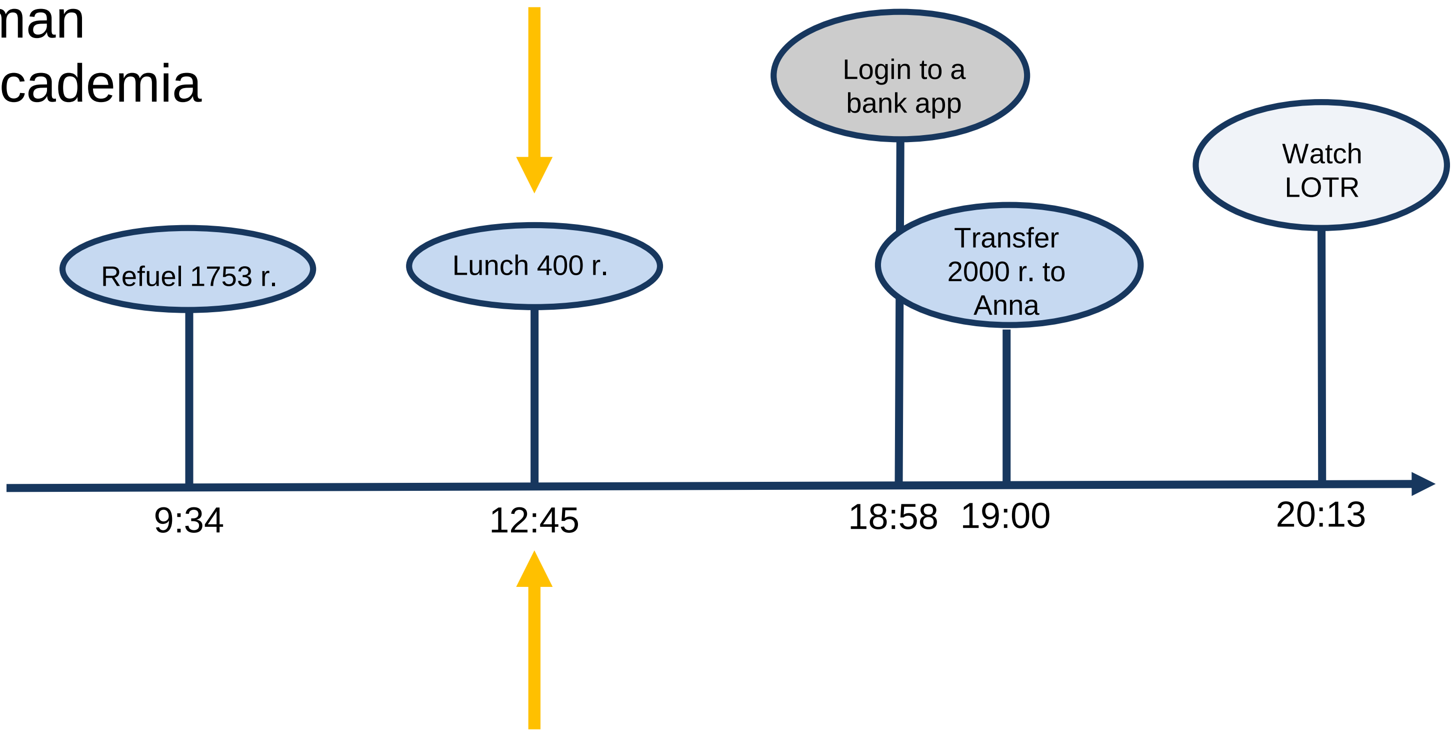
Skoltech

Example: financial transactions are event sequences



Alex, 33, man
works in academia

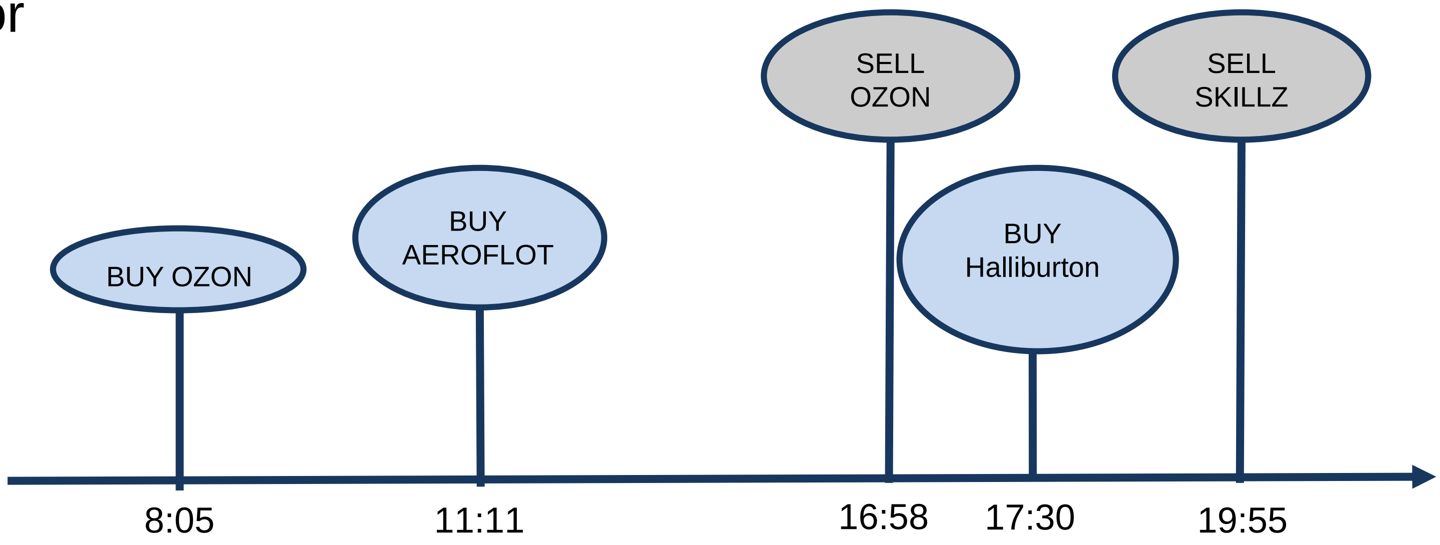
Event features:
sum, geolocation



Event type + event time

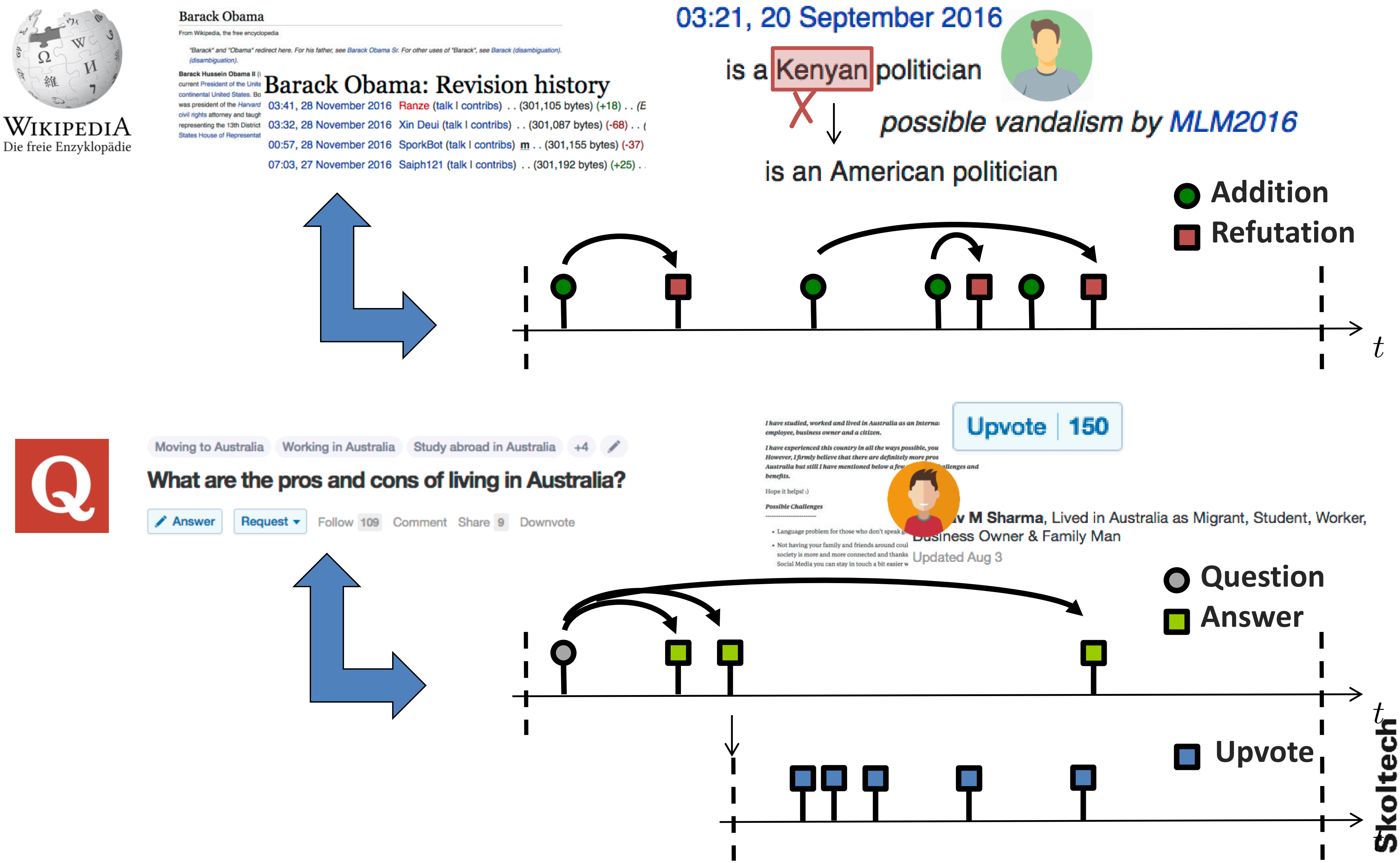
Example: operations in markets are event sequences

An investor



Skoltech

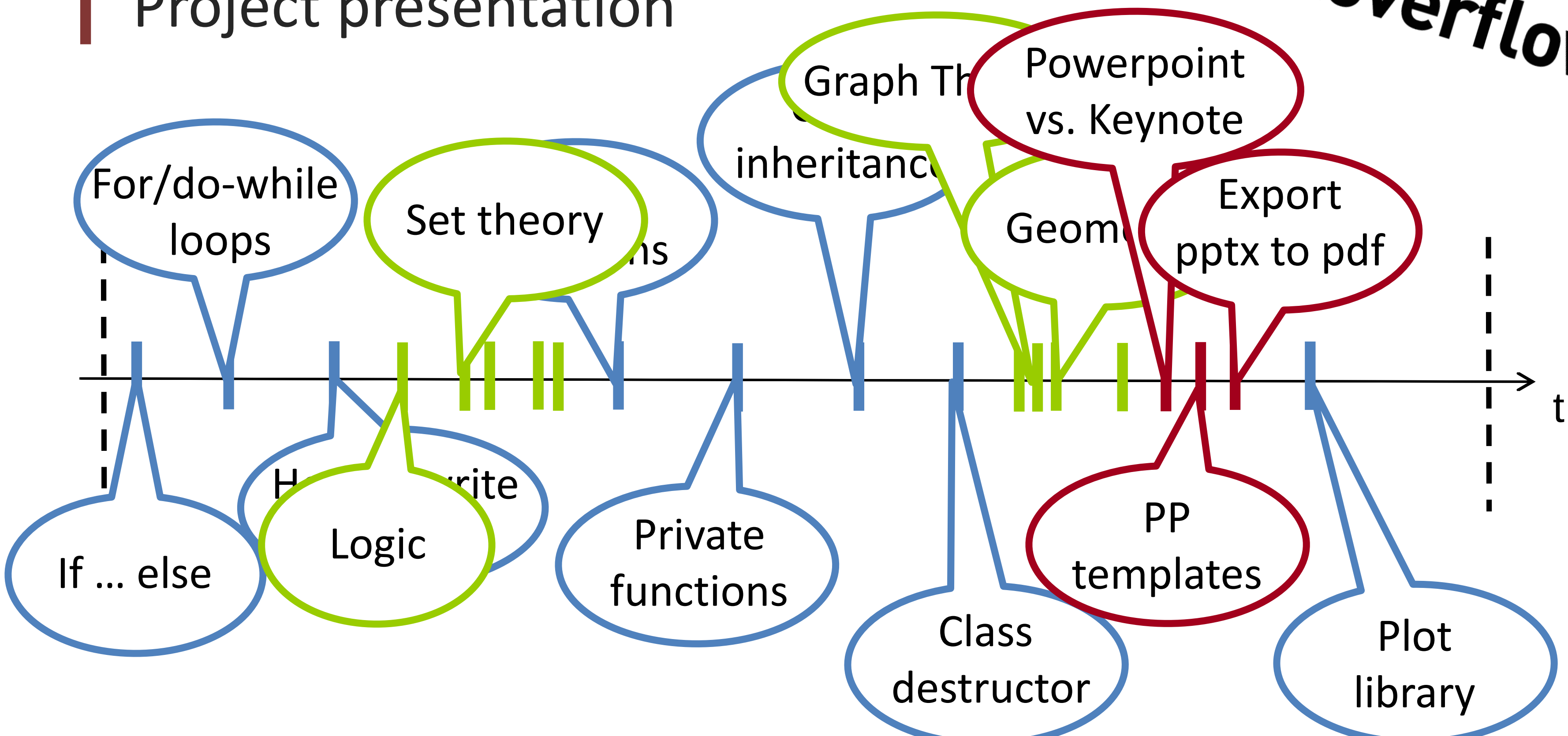
More complex example: response history



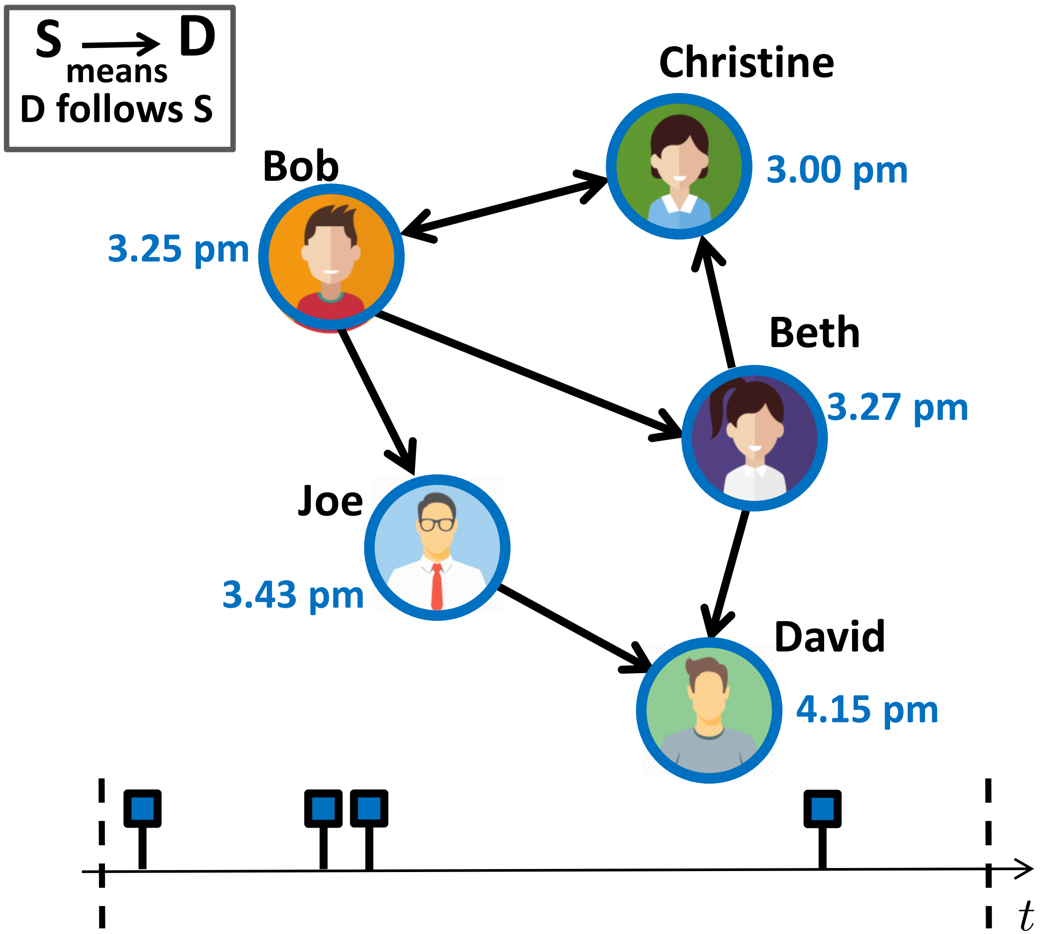
Example: development

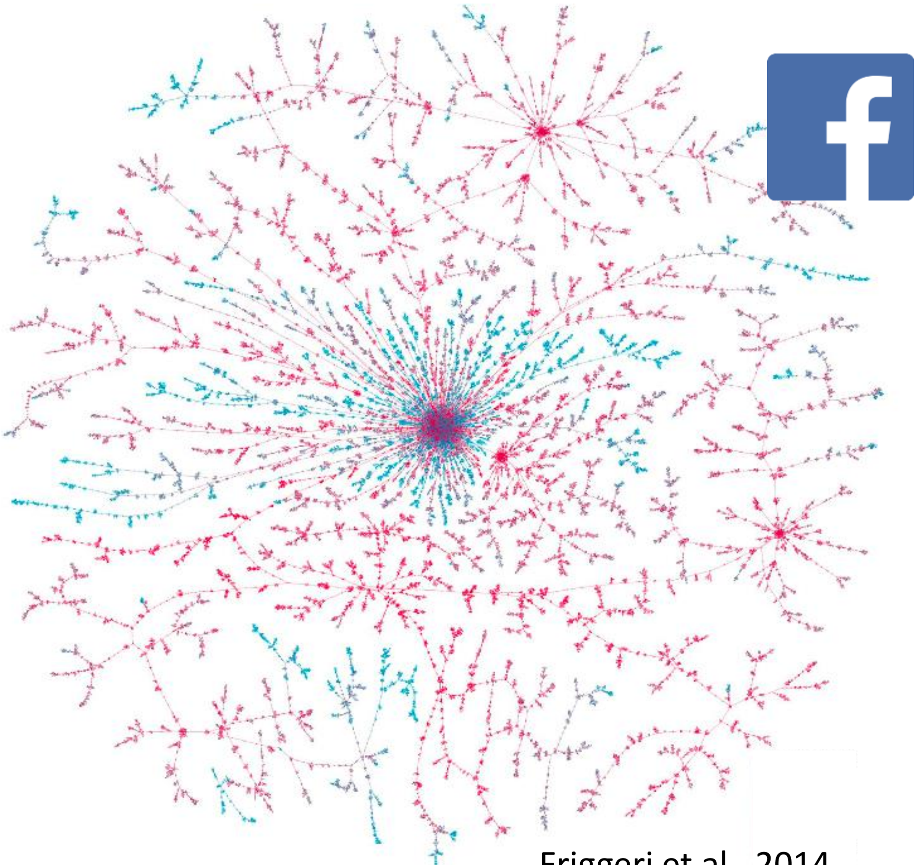
1st year computer science student

- Introduction to programming
- Discrete math
- Project presentation



Example: Information propagation in graphs

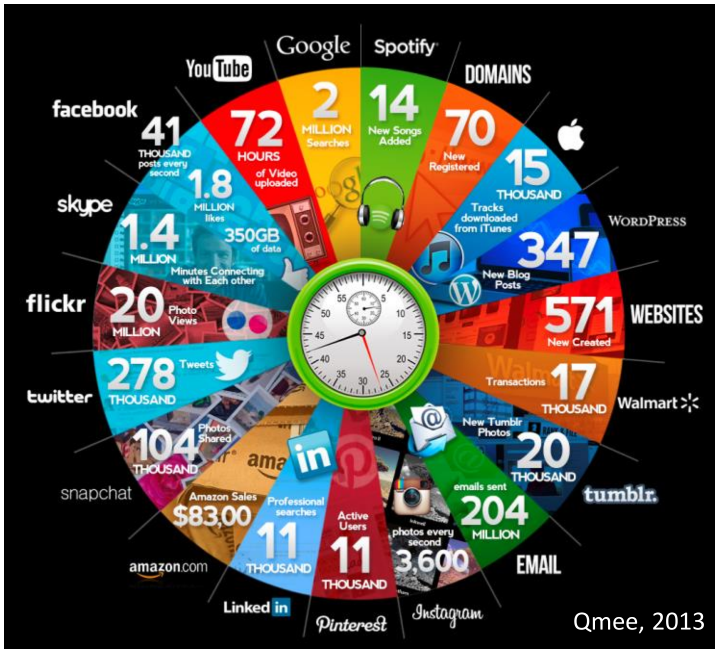




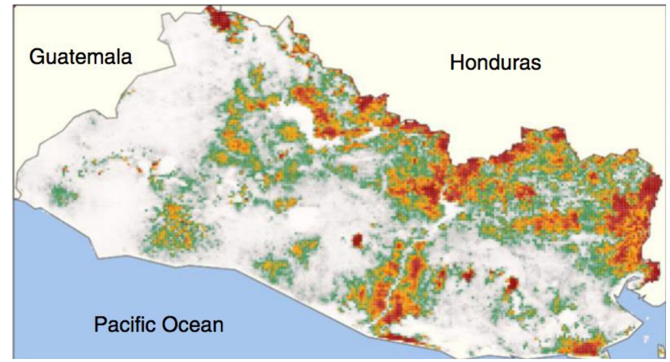
Friggeri et al., 2014

These cascades can deliver valuable insights

Many discrete *events* in continuous time



Clickstreams: Online actions of a user



Disease dynamics

Financial trading

Financial transactions

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

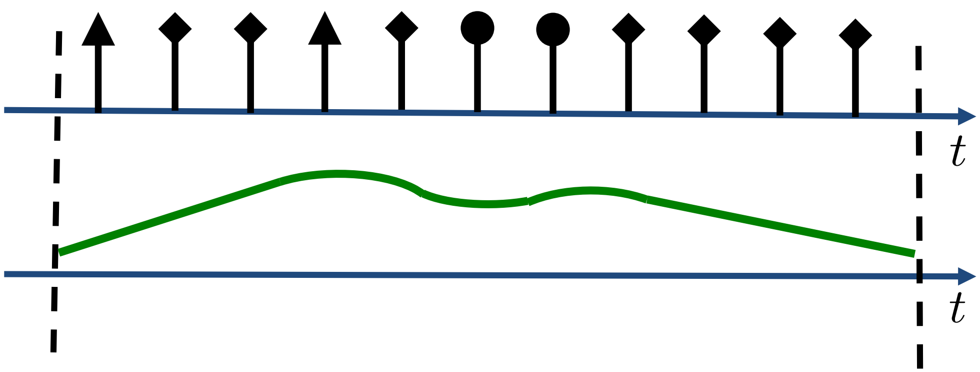
Ride-sharing requests

A user's reputation in Quora

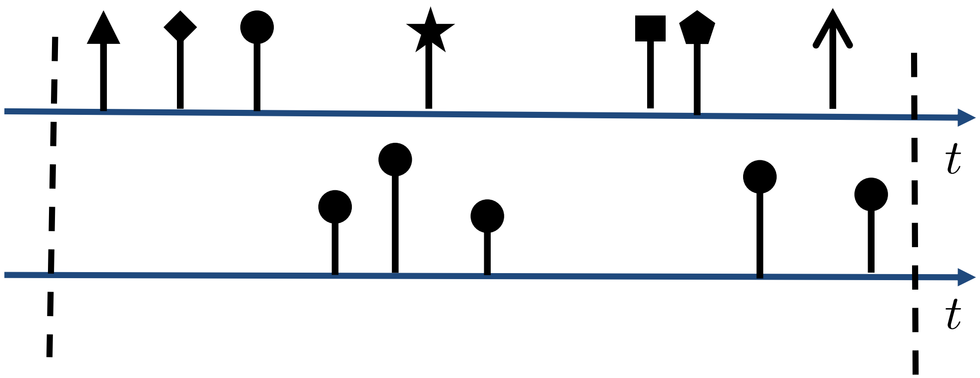
FAST SLOW

…in a wide range of temporal scales.

Aren't these event traces just time series?

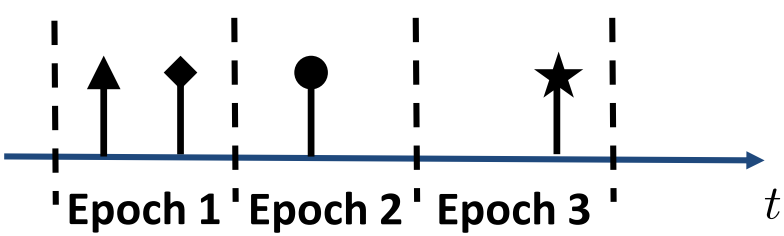


Discrete and continuous times series



Discrete events in continuous time

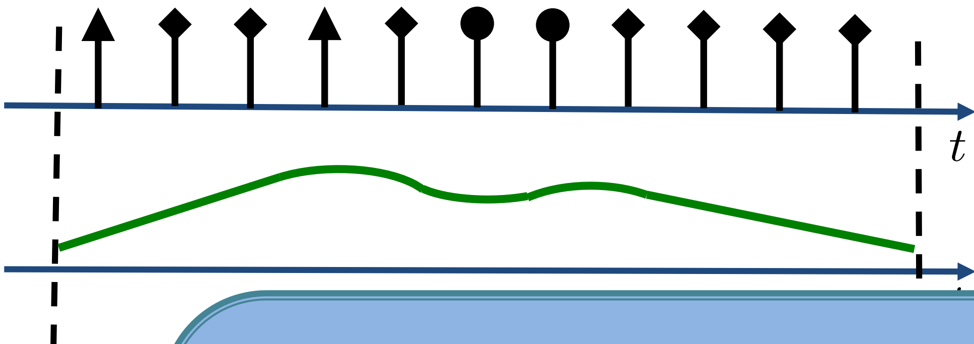
What about aggregating events in *epochs*?

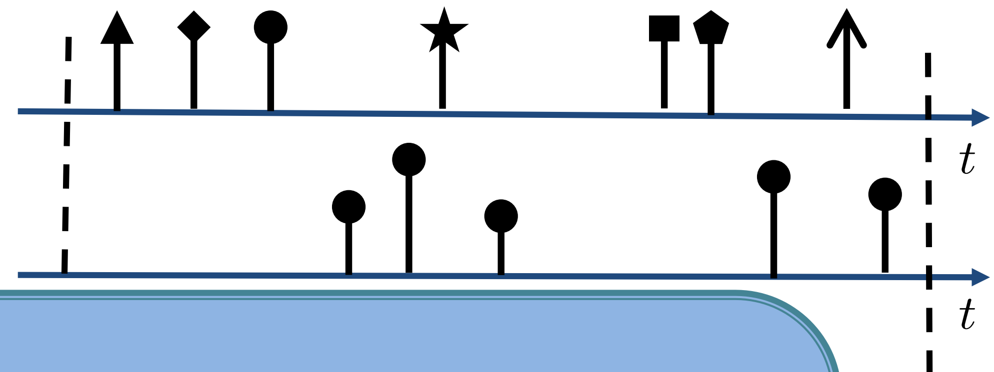


- How long is each epoch?
- How to aggregate events per epoch?
- What if no event in one epoch?
- What about time-related queries?

Skoltech

Aren’t these event traces just time series?





Di

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

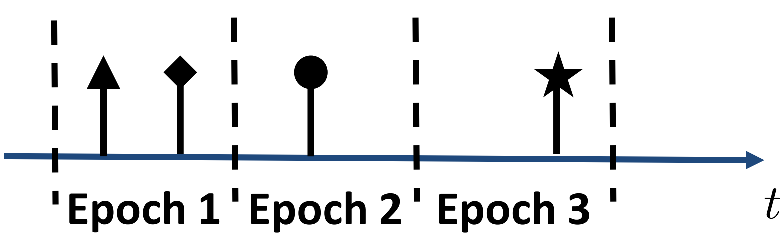
W

events in epochs?

poch?

What if no event in one epoch?

What about time-related queries?



Problems for event sequences we'll solve!

- Compact description of data: models

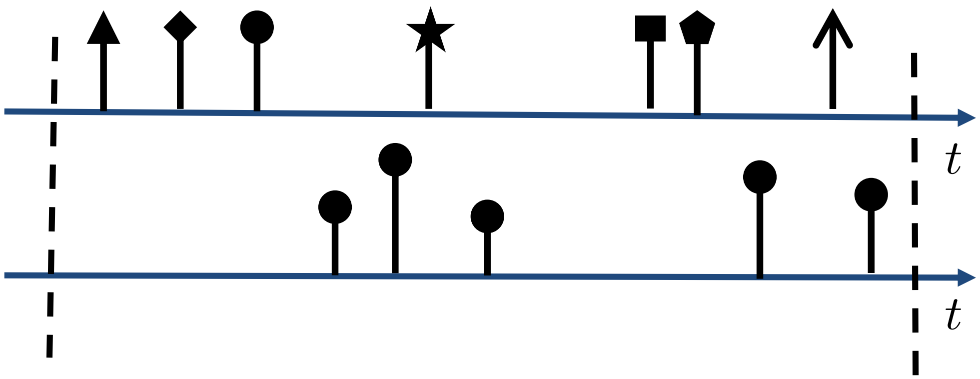
- Forecasting/Prediction: distribution for the next event time; event type

- Interpretation: what causes what?

- Control: what should we do?

- Hypothesis testing

- Simulation



Point process: generates discrete events in continuous time

Skoltech

Outline of the course part

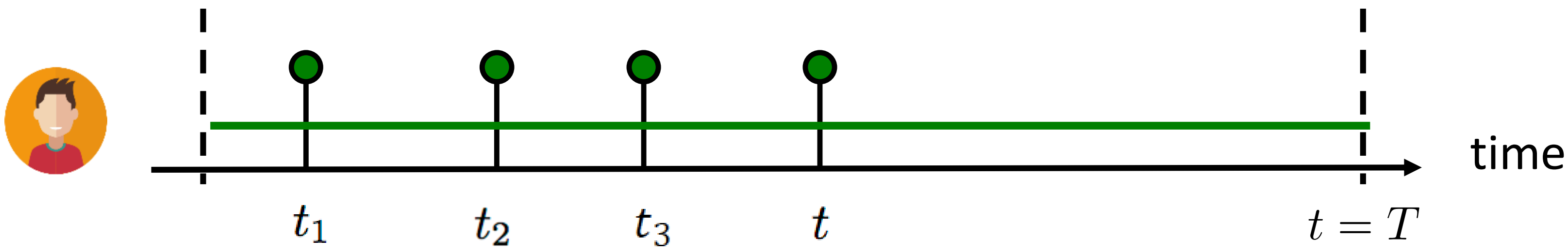
Lecture 1: basic models and concepts

Lecture 2: neural models and their strengths

Poisson process example and estimation for it; basic terms

Poisson process

Coarse approximation of many real-life processes



Intensity of a Poisson process is constant:

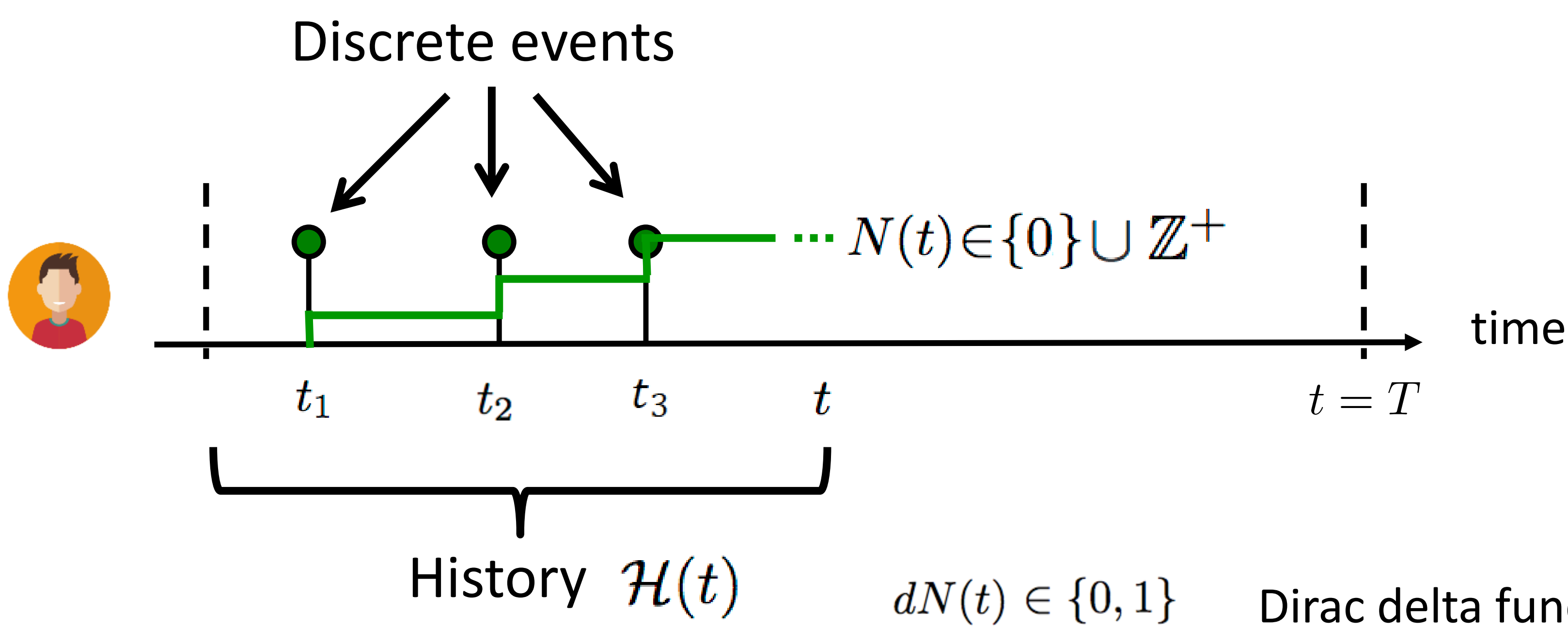
$$\lambda^*(t) = \mu$$

Observations:

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

Defintion: Temporal point processes

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



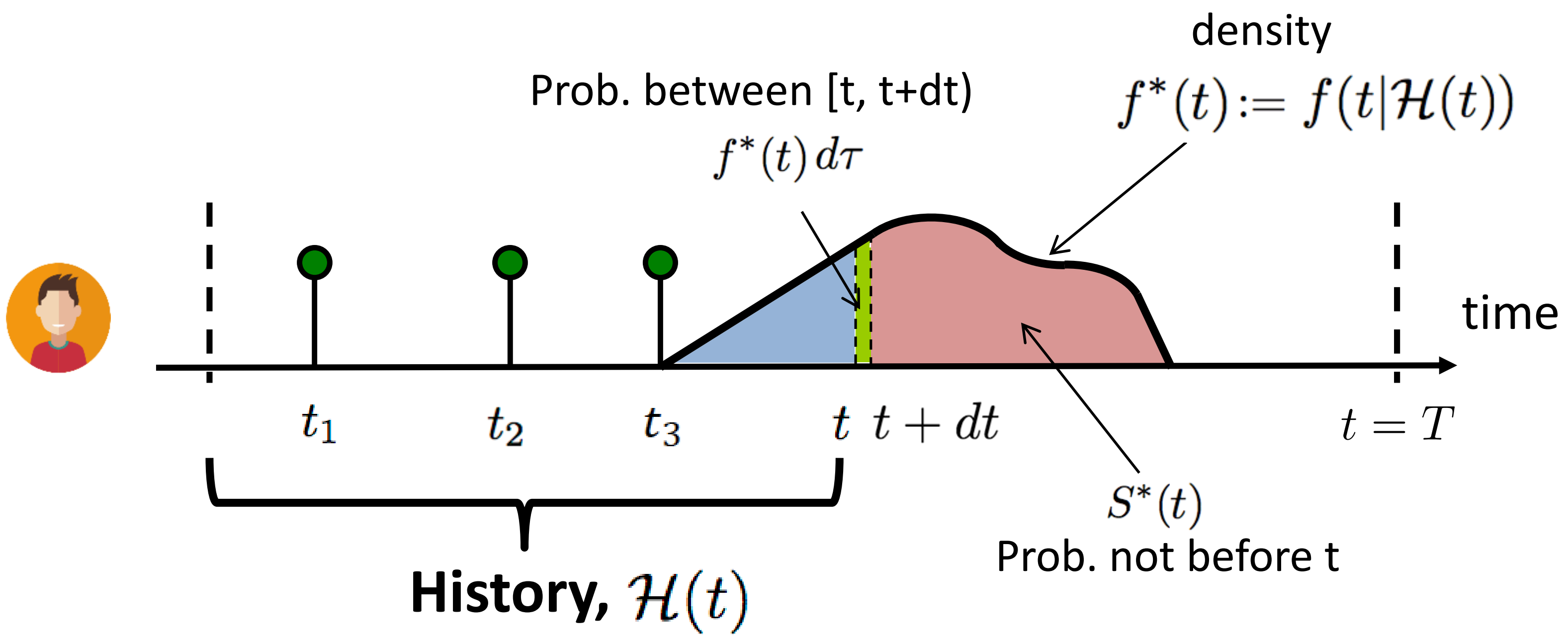
$dN(t) \in \{0, 1\}$

Dirac delta function

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

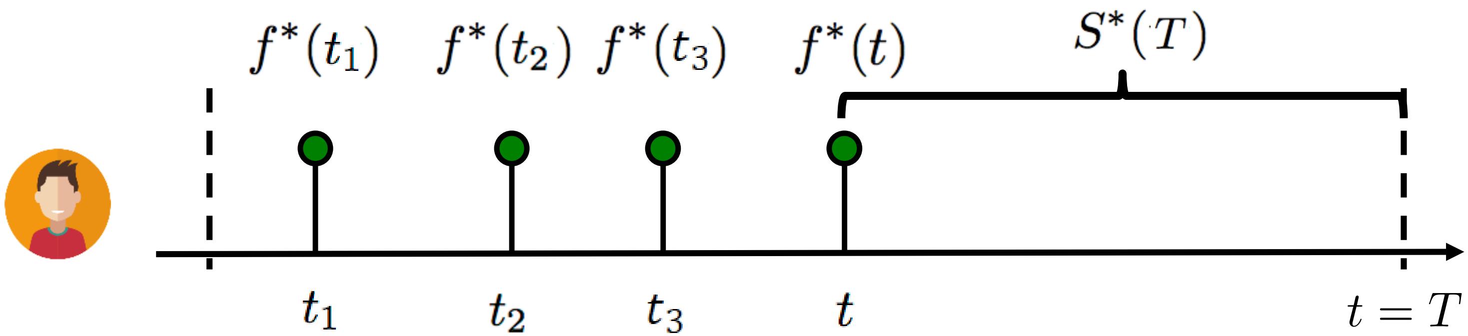
Skoltech

Distribution we are looking for the next event time



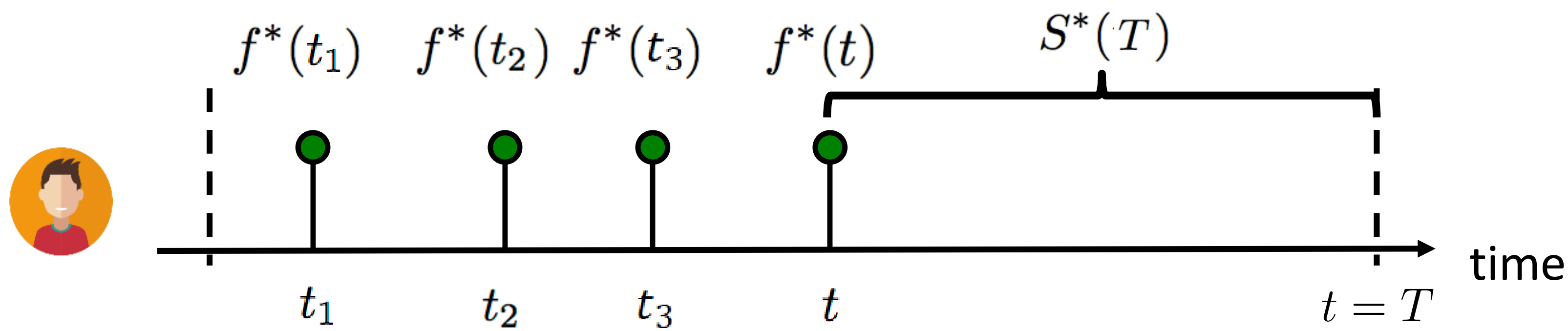
Skoltech

We can write and optimize a likelihood function



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

Density parametrization is hard



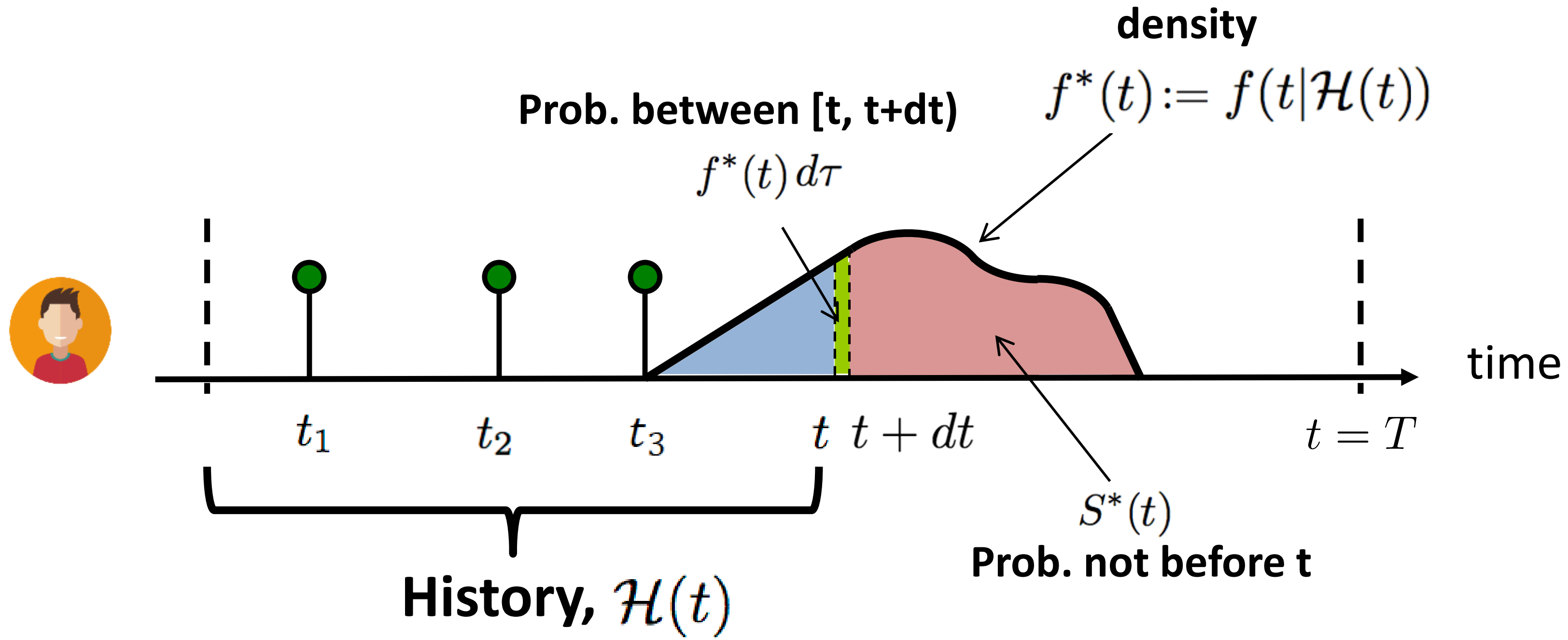
$$f^*(t_1) \quad f^*(t_2) \quad f^*(t_3) \quad f^*(t) \quad S^*(T)$$

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

It is <u>difficult for model design and interpretability</u>:

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

Intensity function is an alternative



Intensity:

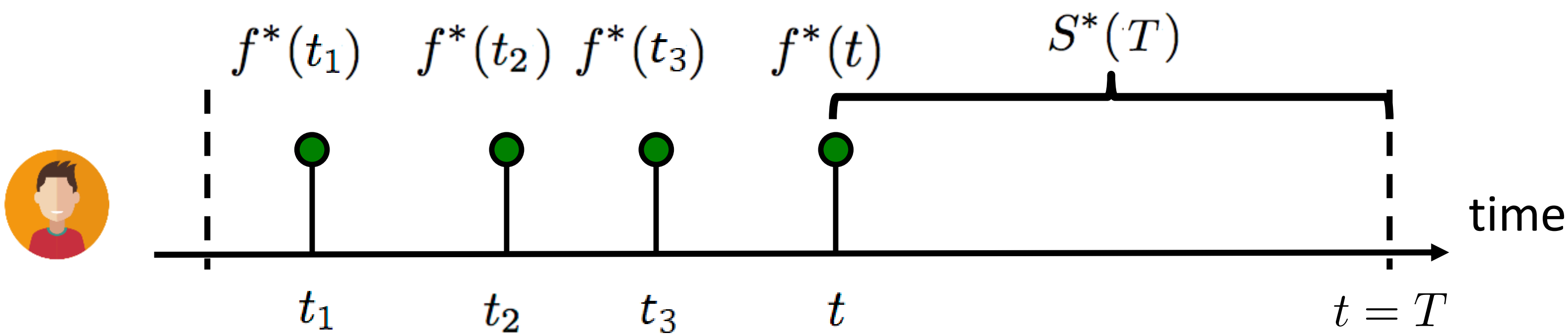
Probability between [t, t+dt) but not before t

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

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

Skoltech

Likelihood for intensity



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

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

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

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

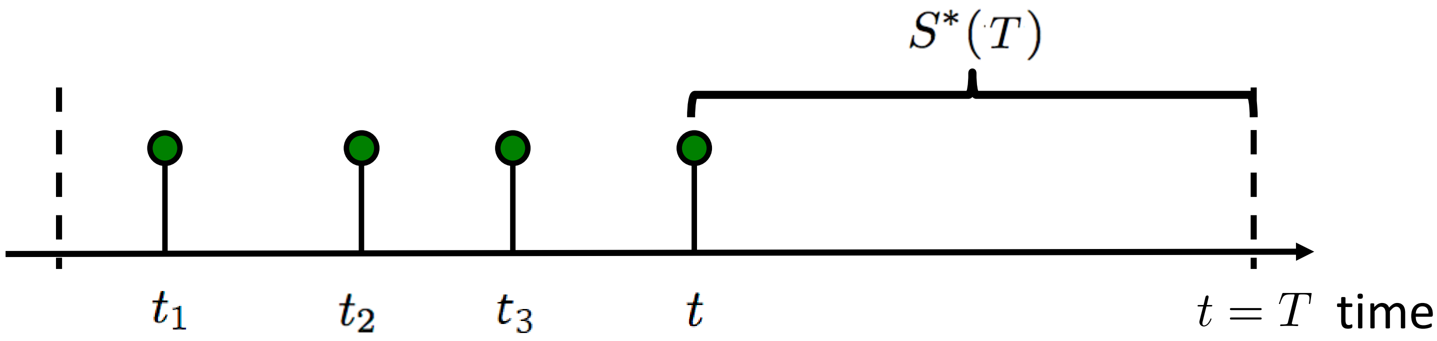
$\langle w, \phi^*(t)\rangle$

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

Suitable for model design and interpretable:

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

Log likelihood via intensity and density

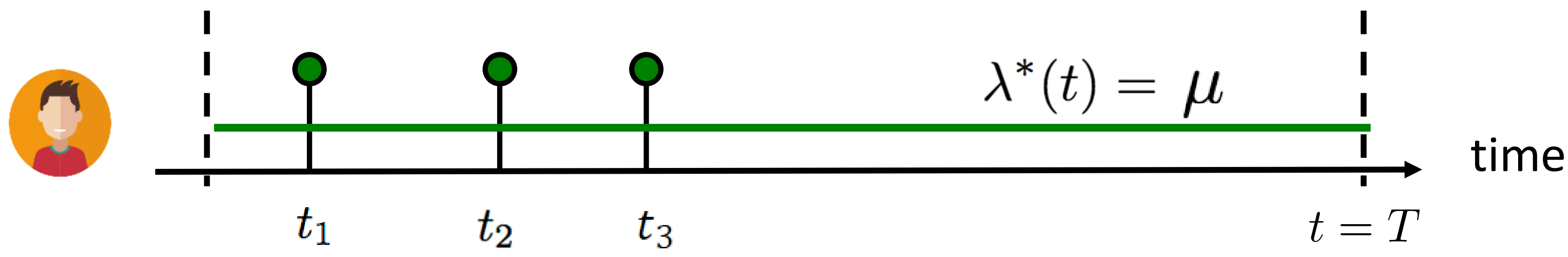


$$L = f(t_1|\mathcal{H}_0)f(t_2|\mathcal{H}_{t_1})\cdots f(t_n|\mathcal{H}_{t_{n-1}})(1 - F(T|\mathcal{H}_{t_n}))$$

$$\begin{aligned}
L &= \left(\prod_{i=1}^{n} f(t_i|\mathcal{H}_{t_{i-1}})\right)\frac{f(T|\mathcal{H}_{t_n})}{\lambda^*(T)} \\
&= \left(\prod_{i=1}^{n} \lambda^*(t_i)\exp\left(-\int_{t_{i-1}}^{t_i}\lambda^*(s)\mathrm{d}s\right)\right)\exp\left(-\int_{t_n}^{T}\lambda^*(s)\mathrm{d}s\right) \\
&= \left(\prod_{i=1}^{n}\lambda^*(t_i)\right)\exp\left(-\int_0^T\lambda^*(s)\mathrm{d}s\right),
\end{aligned}$$

Skoltech

Fitting & sampling for a Poisson process



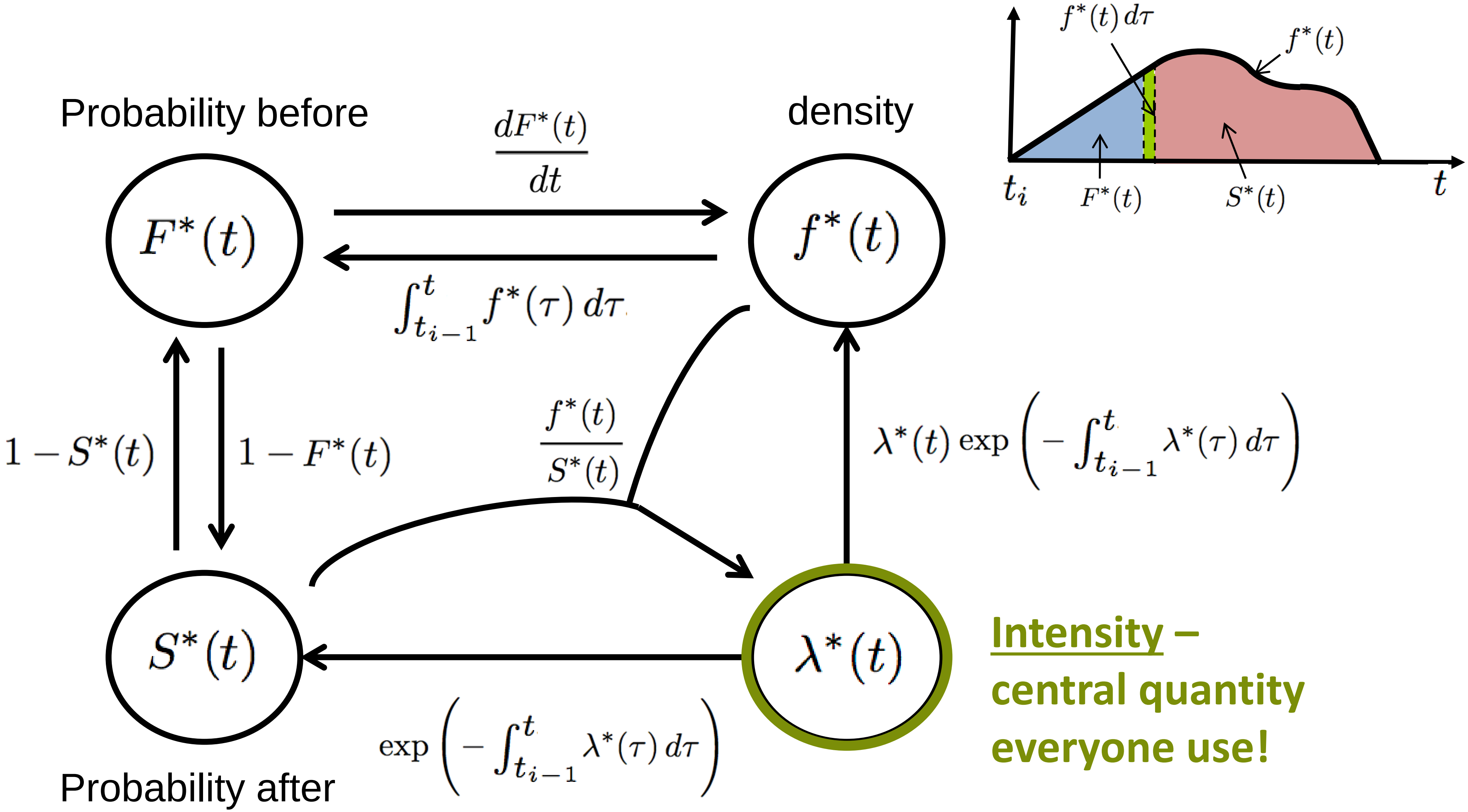
Fitting by maximum likelihood:

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

Sampling using inversion sampling:

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

Relation between f*, F*, S*, λ*



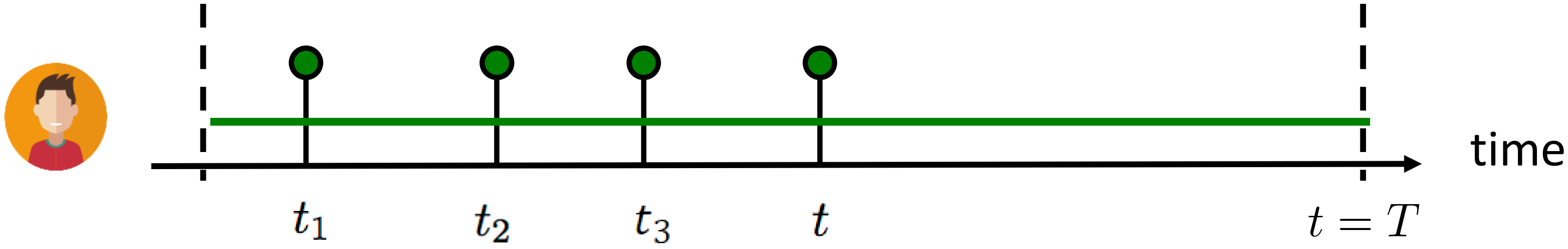
Intensity – central quantity everyone use!

Skoltech

Other models for temporal point processes

Poisson process

Coarse approximation of many real-life processes



Intensity of a Poisson process

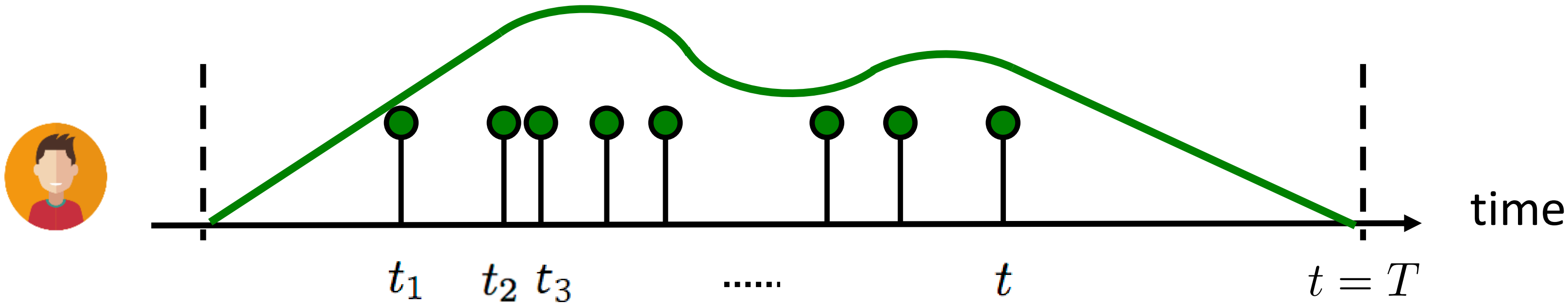
$$\lambda^*(t) = \mu$$

Observations:

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

Skoltech

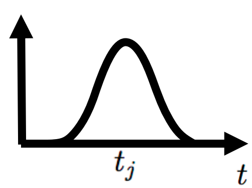
Inhomogeneous Poisson process



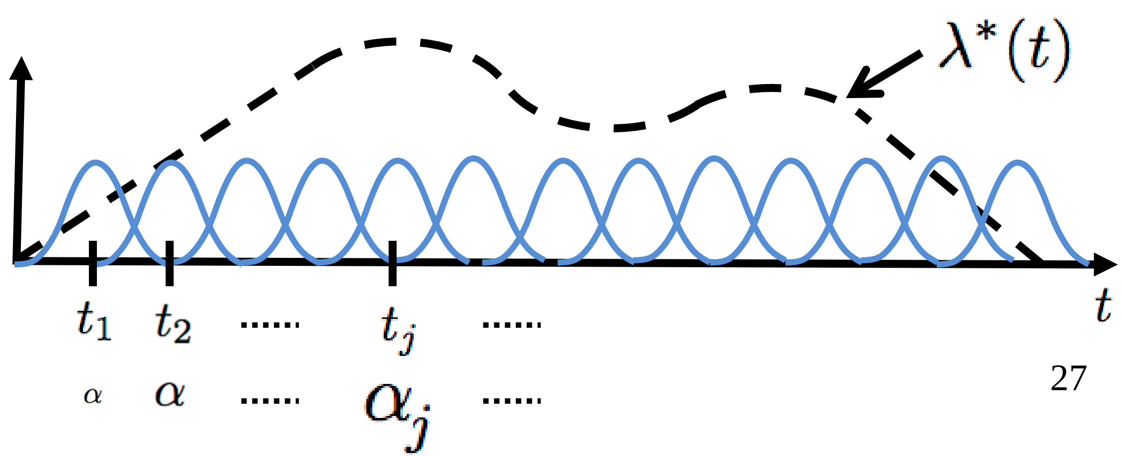
Intensity of an inhomogeneous Poisson process

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

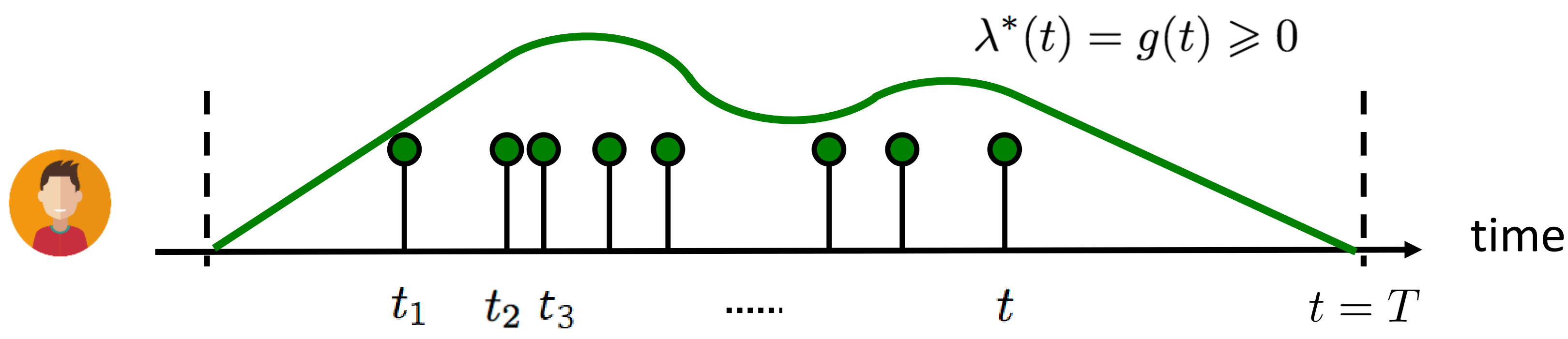
Example:



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



Fitting from inhomogeneous Poisson



Fitting by maximum log-likelihood:

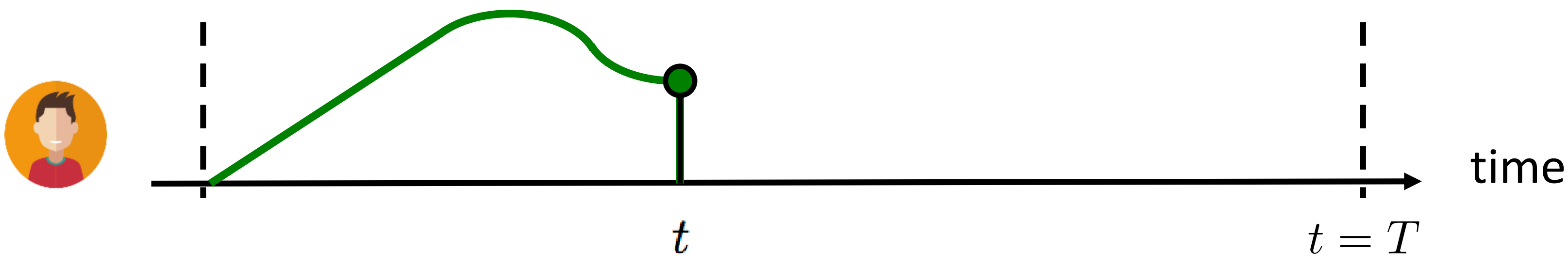
$$\underset{g(t)}{\text{maximize}} \sum_{i=1}^{n} \log g(t_i) - \int_0^T g(\tau)\, d\tau$$

Idea: we have additional features, so we can use a generalized linear model for it

Intensity is $g(t) = g(\boldsymbol{x}_t) = \exp(\boldsymbol{x}_t^T w)$

Skoltech

Terminating (or survival) process



Intensity of a terminating (or survival) process

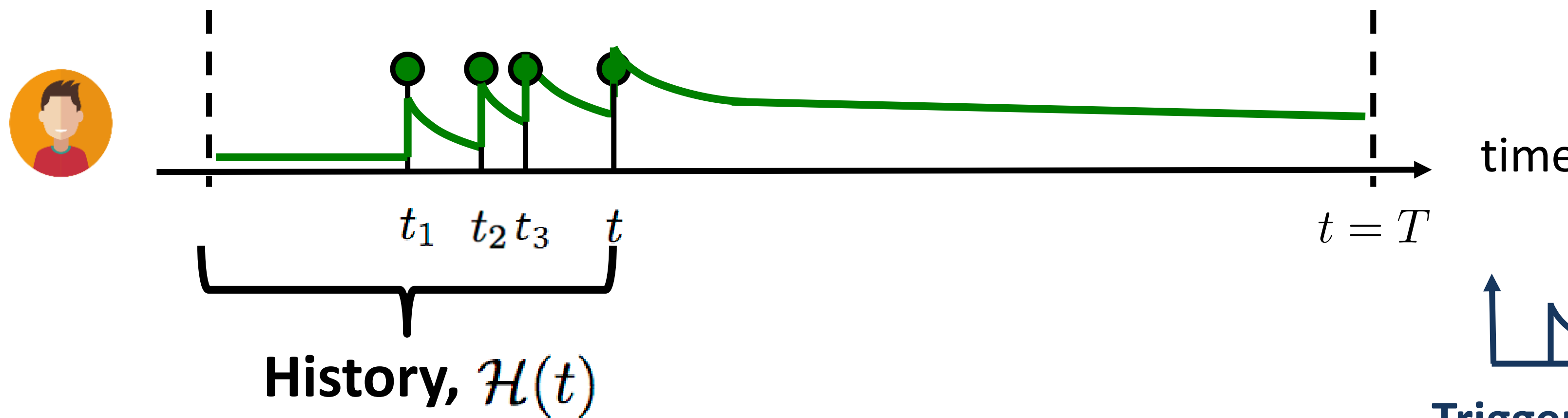
$$\lambda^*(t) = g^*(t)(1 - N(t)) \geqslant 0$$

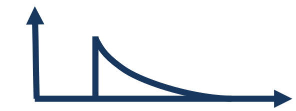
Observations:

1. Limited number of occurrences
2. Hazard function in actuarial science

Skoltech

Self-exciting Hawkes process





Intensity of self-exciting (or Hawkes) process:

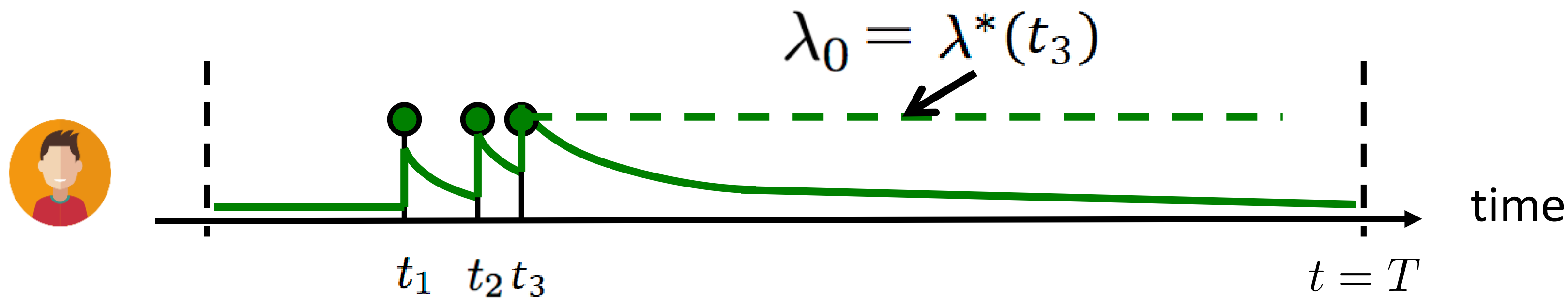
$$\lambda^*(t) = \mu + \alpha \sum_{t_i \in \mathcal{H}(t)} \overbrace{\kappa_\omega(t - t_i)}^{\text{Triggering kernel}}$$
$$= \mu + \alpha \kappa_\omega(t) \star dN(t)$$

Observations:

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

Hawkes, Alan G. "Spectra of some self-exciting and mutually exciting point processes." *Biometrika* 58.1 (1971): 83-90.

Fitting a Hawkes process from a recorded timeline



Fitting by maximum likelihood:

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

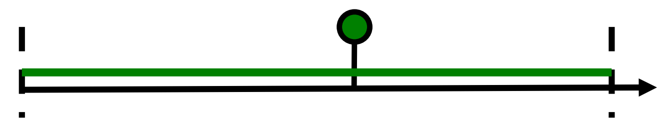
The max. likelihood is jointly convex in μ and α

Summary

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

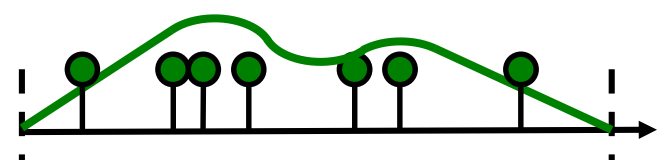
Poisson processes:

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



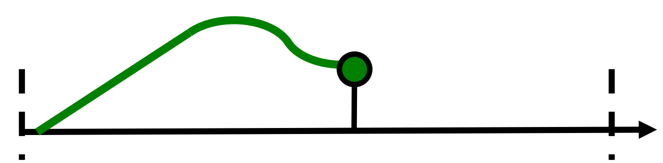
Inhomogeneous Poisson processes:

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



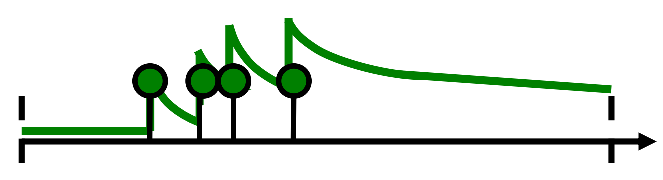
Terminating point processes:

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



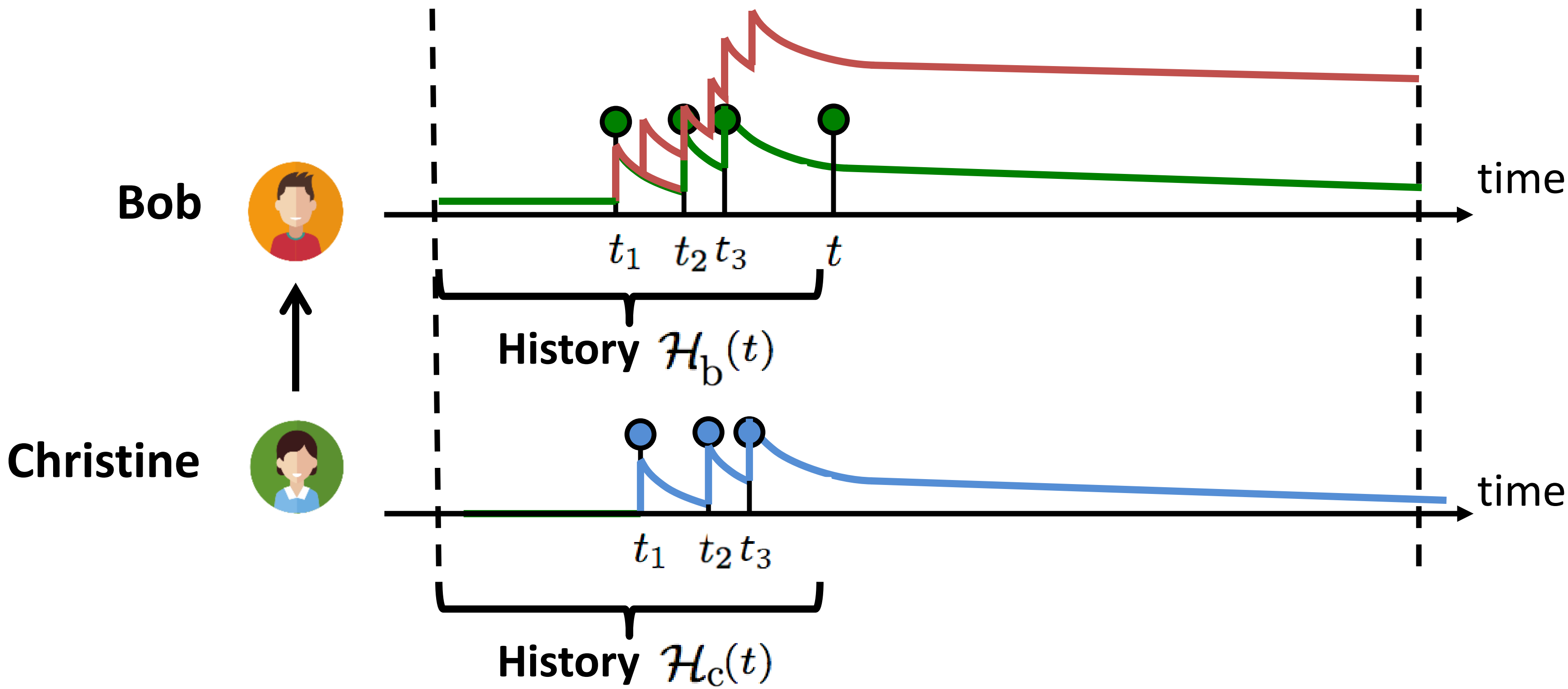
Self-exciting point processes:

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



Mutually excited point processes

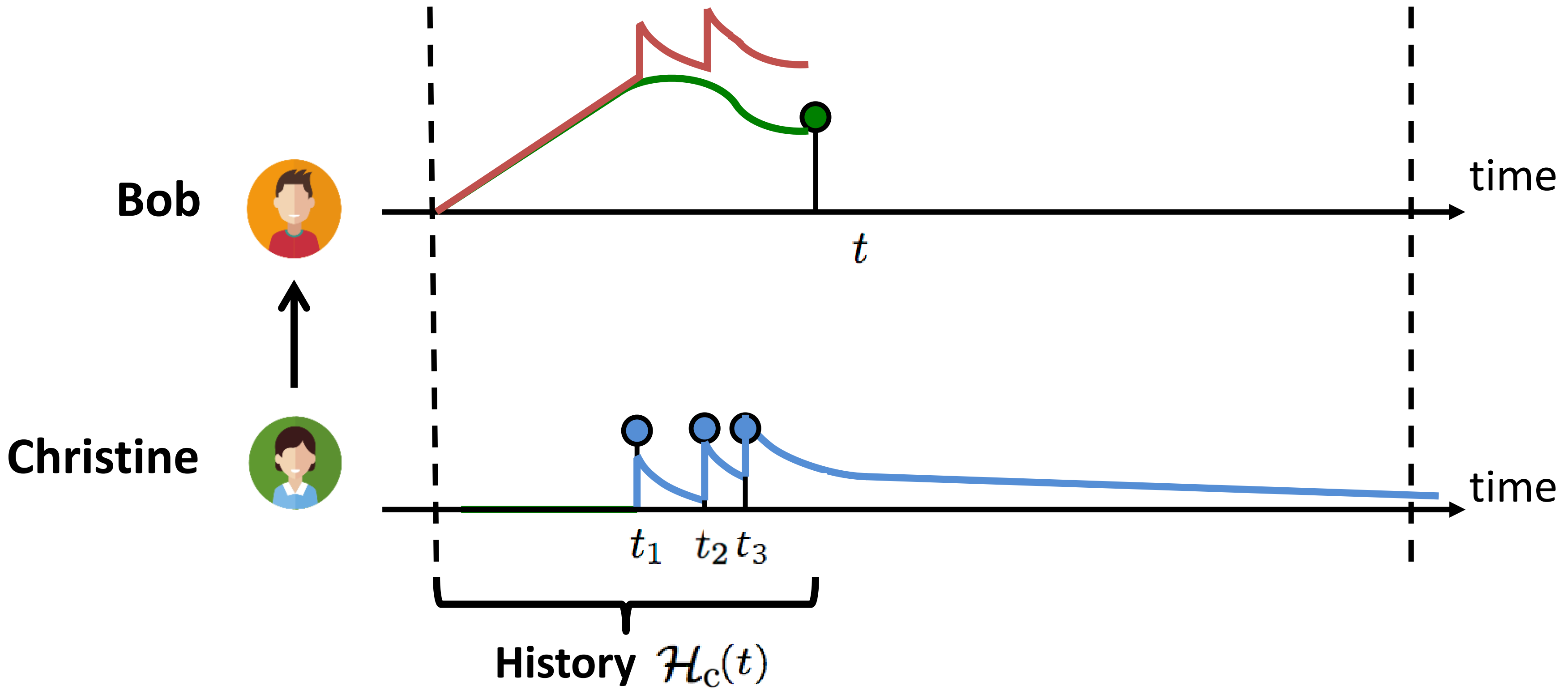
Mutually exciting process



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

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

Skoltech

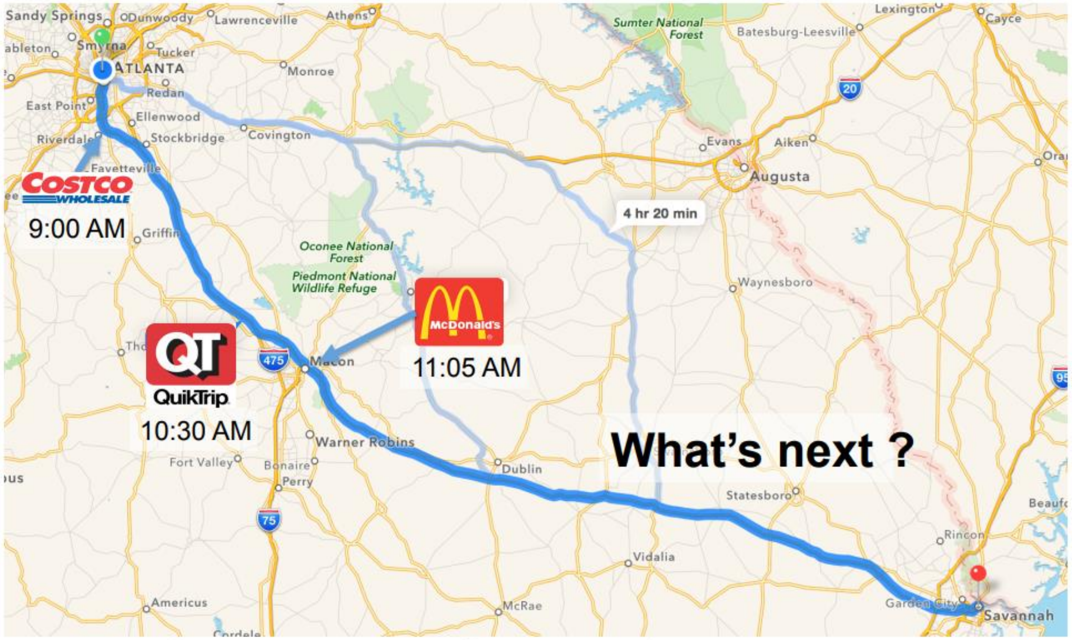
Marked temporal point processes

Skoltech

Marked temporal point processes

Marked temporal point process:

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



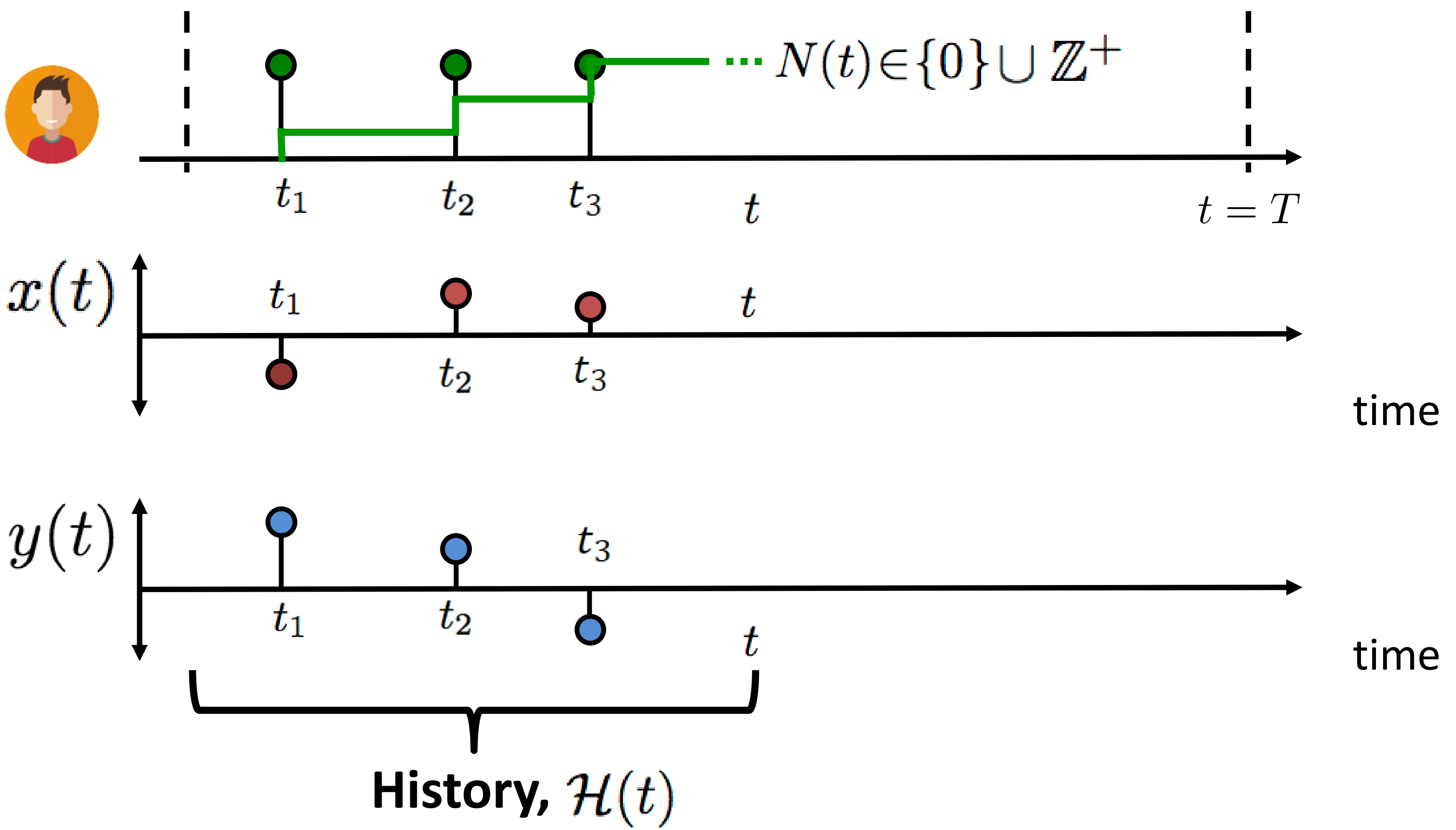
Given the trace of past locations and time, can we predict the location and time of the next stop?

Figure source: Du, Nan, et al. "Recurrent marked temporal point processes: Embedding event history to vector." KDD. 2016.

Marked temporal point processes

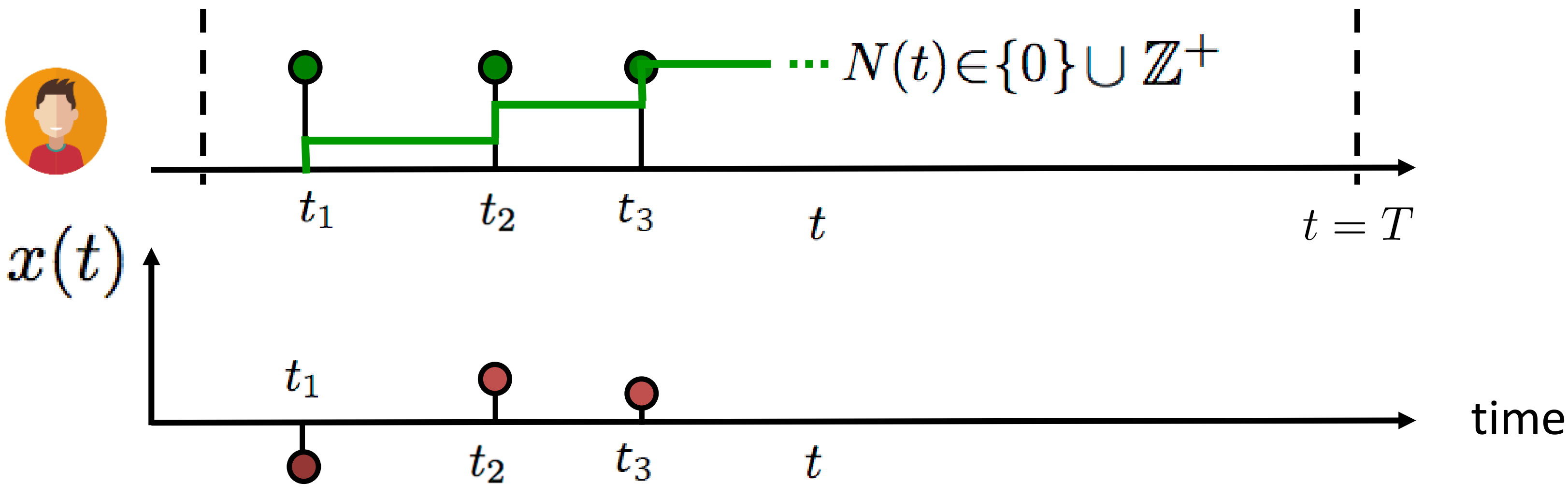
Marked temporal point process:

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



Skoltech

Independent identically distributed marks



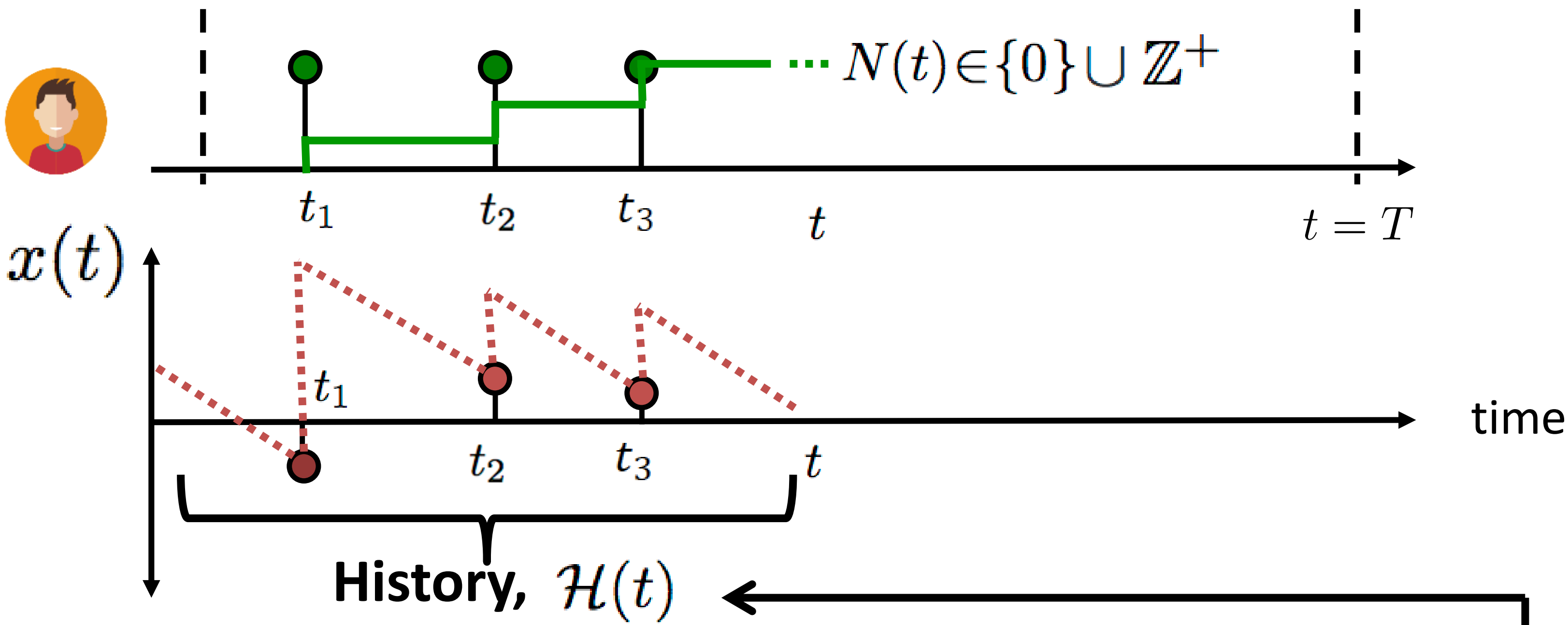
Distribution for the marks:

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

Observations:

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

Dependent marks: SDEs with jumps



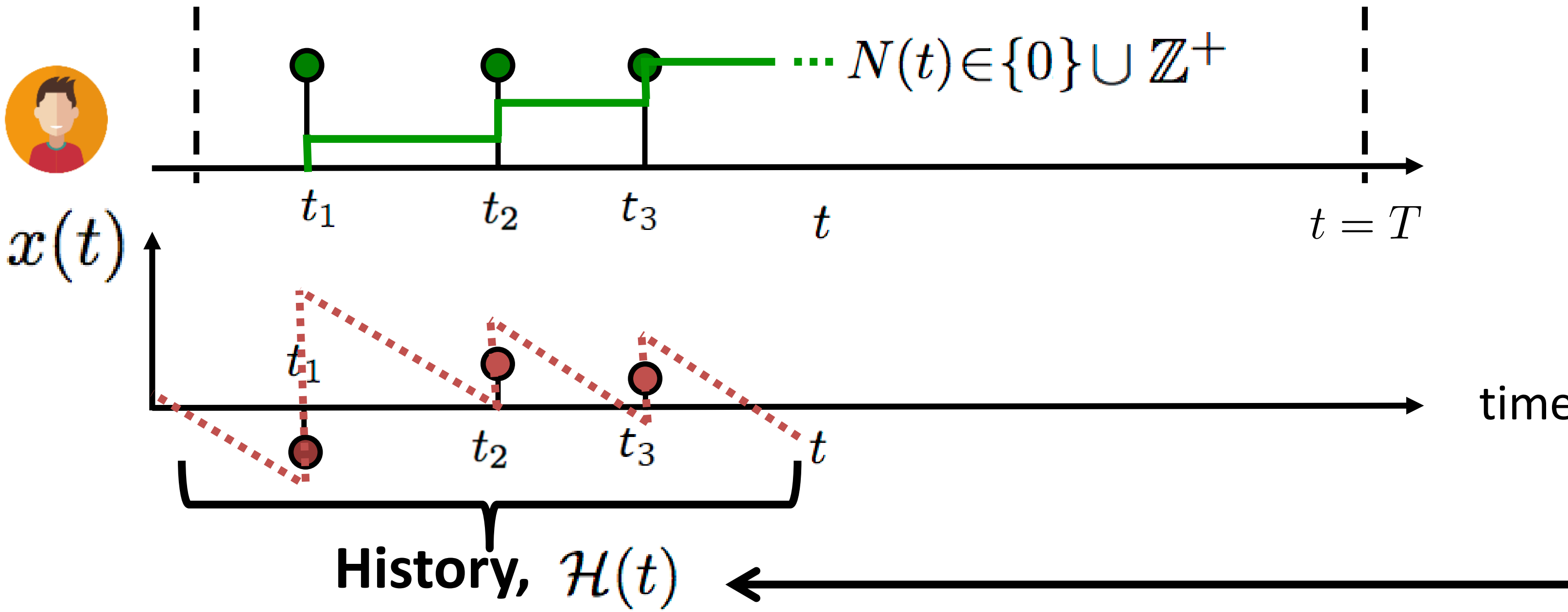
Marks given by stochastic differential equation with jumps:

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

Observations:

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

Dependent marks: distribution + SDE with jumps



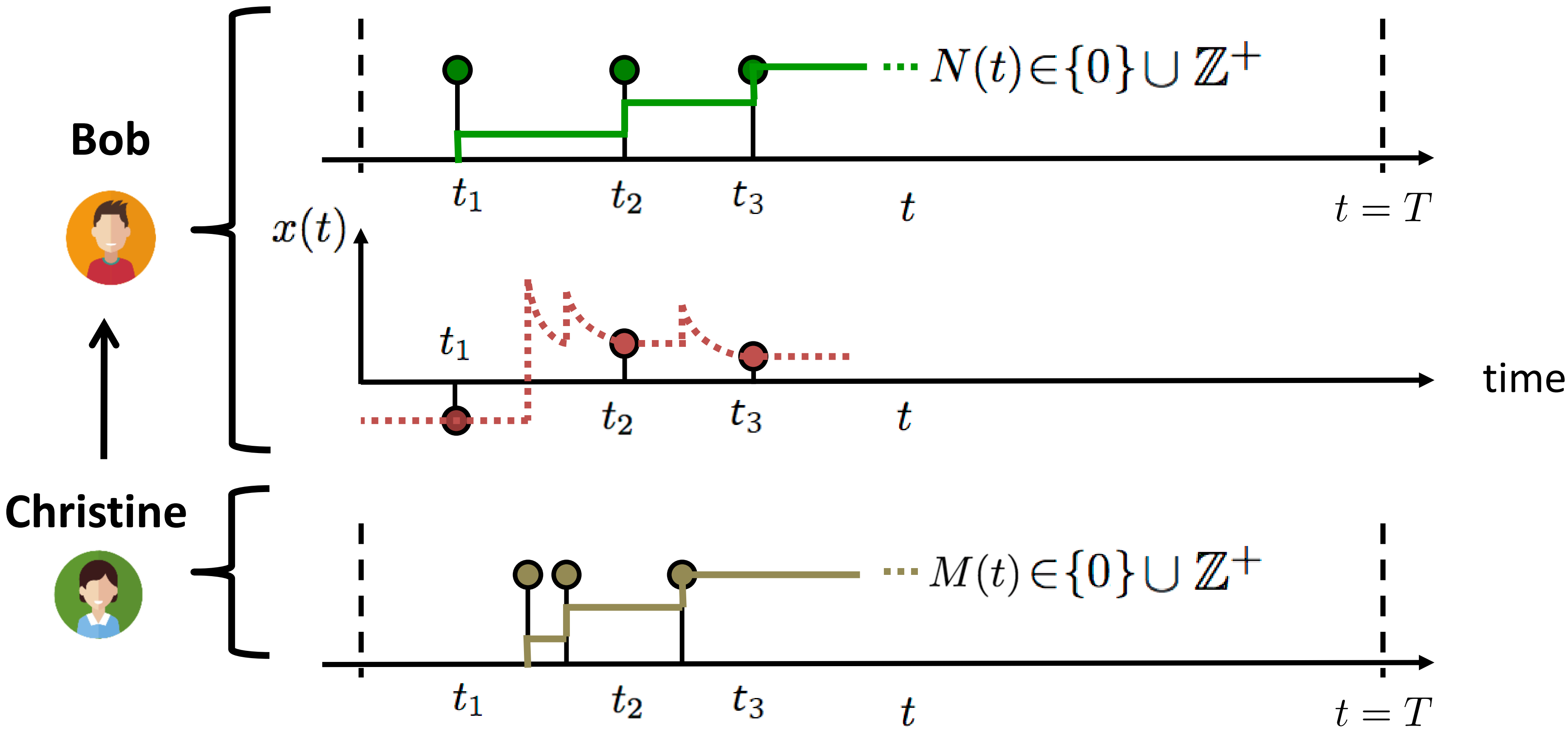
Distribution for the marks:

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

Observations:

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

Mutually exciting + marks



Marks affected by neighbors

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

Conclusions

Skoltech

Conclusions

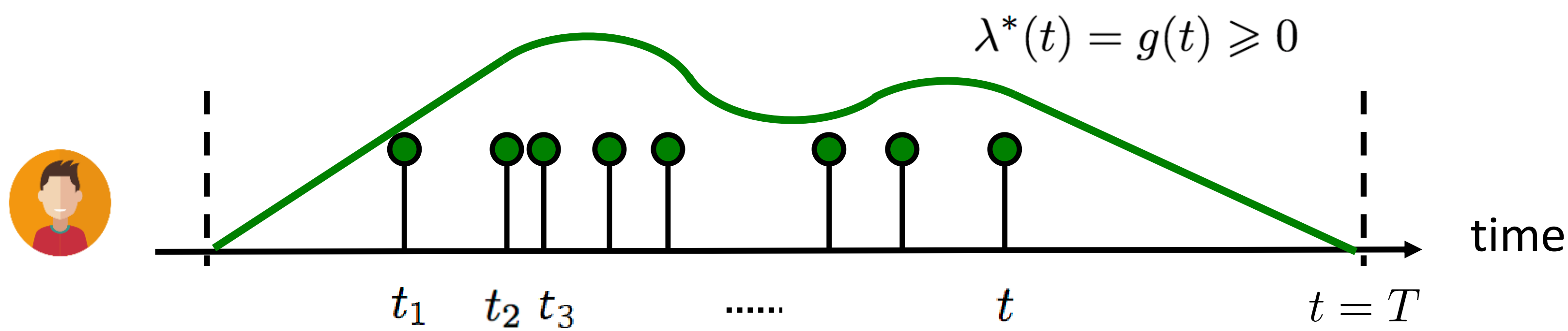
- Inhomogeneous Poisson point process provide a reasonable framework for the description of point processes

- Next time: how can we use neural networks inside this framework

Bonus II: Sampling

Skoltech

Fitting & sampling from inhomogeneous Poisson



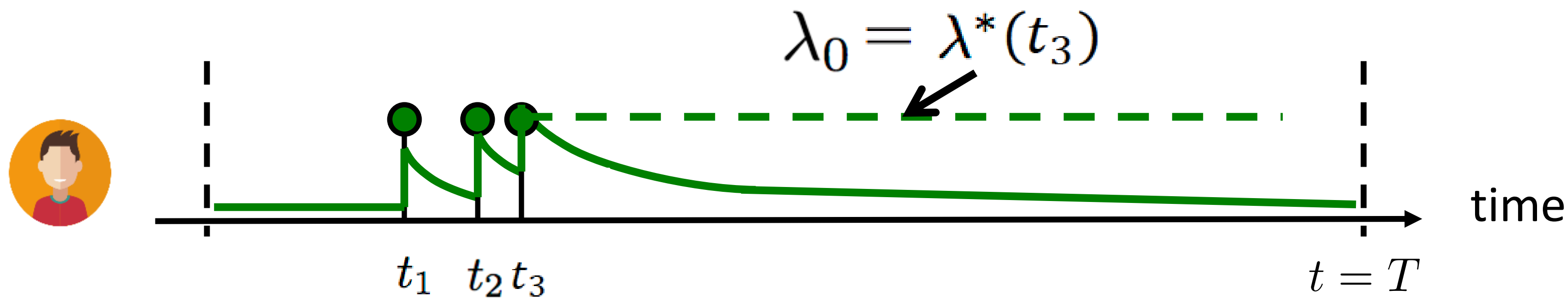
Fitting by maximum likelihood: $\underset{g(t)}{\text{maximize}} \sum_{i=1}^{n} \log g(t_i) - \int_0^T g(\tau)\, d\tau$

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. $g(t)/\mu$

Fitting a Hawkes process from a recorded timeline



Fitting by maximum likelihood:

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

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