Machine Learning in Healthcare

Bachelor in Data Science and Engineering

Matilde Sánchez Fernández
Pablo Martínez Olmos

Course content (partial)

Chapter I. Survival analysis

Chapter II. Point processes

- Point processes
- Examples of application
- Temporal Point Processes
- Intensity function
- Inference for spatial point patterns

Chapter III. Deep latent variable models for medical data

Chapter II. Point processes

Point processes
Examples of application
Temporal Point Processes
Intensity function
Inference for spatial point patterns

Point processes

Modes sequences of discrete events occurring in a continuous domain[1]:

- **Temporal processes:** model sequences of discrete (medical) events in continuous time $t \in \mathbb{R}$ (hospital visits, arrhythmia episodes, seizure onsets, infection spread)
  $\longrightarrow$ Models WHEN?
- **Spatial processes:** Models events in 2-D continuous space domain (locations of tumors within tissue, distribution of receptors on a cell membrane, geographic spread of disease cases, locations of ambulance dispatches in a city)
  $\longrightarrow$ Models WHERE?
- **Spatio-temporal processes:** Models events in 3-D continuous space-time domain (epidemic spread, hospital admissions across space and time, etc.)
  $\longrightarrow$ Models WHERE and WHEN?

[1] For this chapter we are using material from the course *Human-Centered ML*, of Saarland University, and ICML 2018 TUTORIAL *Learning with Temporal Point Processes*, given by A. De, U. Upadhyay, M. Gomez-Rodriguez, and I. Valera.

Point processes

Modes sequences of discrete events occurring in a continuous domain[1]:

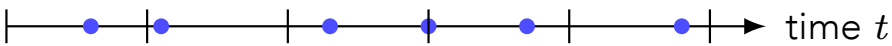
- **Temporal processes:** model sequences of discrete (medical) events in continuous time $t \in \mathbb{R}$ (hospital visits, arrhythmia episodes, seizure onsets, infection spread)
  $\longrightarrow$ Models WHEN?
- **Spatial processes:** Models events in 2-D continuous space domain (locations of tumors within tissue, distribution of receptors on a cell membrane, geographic spread of disease cases, locations of ambulance dispatches in a city)
  $\longrightarrow$ Models WHERE?
- **Spatio-temporal processes:** Models events in 3-D continuous space-time domain (epidemic spread, hospital admissions across space and time, etc.)
  $\longrightarrow$ Models WHERE and WHEN?

Includes Poisson processes, Hawkes processes, terminal point processes, neural point processes...

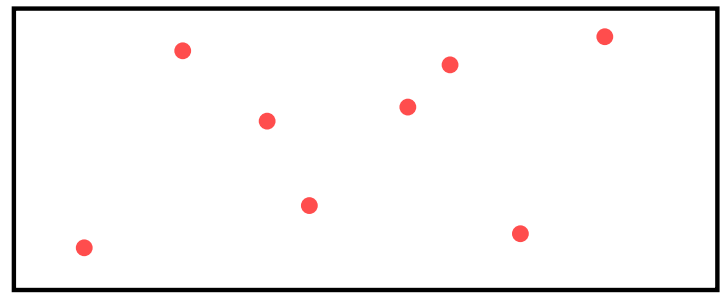
[1] For this chapter we are using material from the course *Human-Centered ML*, of Saarland University, and ICML 2018 TUTORIAL *Learning with Temporal Point Processes*, given by A. De, U. Upadhyay, M. Gomez-Rodriguez, and I. Valera.

Point processes

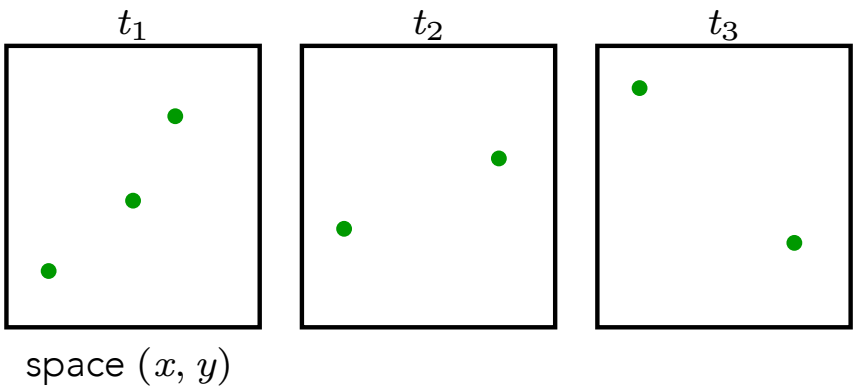
Temporal (time)



Spatial (2D space)



Spatio-temporal (space $\times$ time)



Legend

- time event
- spatial event
- spatio-temporal event

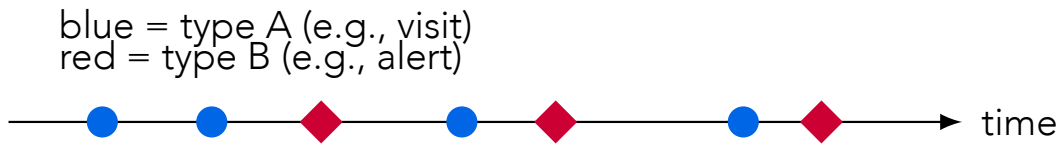
Point processes

Builds directly on Survival Analysis:

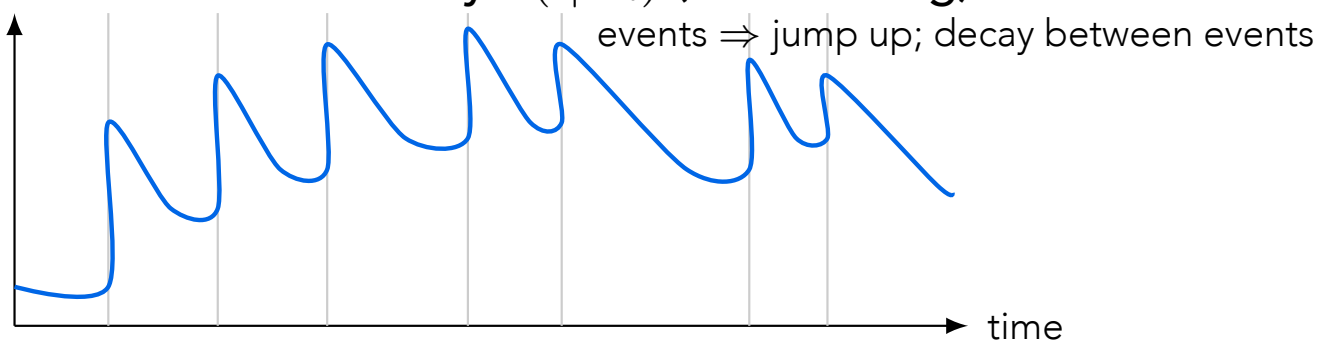
- **In the temporal domain...** both are about time-to-event modeling, but point processes allow *multiple* events per subject and mutually exciting or inhibiting patterns.
- The survival **hazard** becomes the point-process **conditional intensity** $\lambda(t|\mathcal{H}(t))$, now conditioned on the whole event history $\mathcal{H}(t)$ (not just covariates).
- Naturally handles **recurrent events**, **competing types** (marked processes), and **time-varying covariates**.
- Likelihoods look familiar: product over events of $\lambda(t_k)$ times an exponential of the integrated intensity $\int \lambda(u)\, du$ (an analogue of $\int h(u)\, du$).
- Captures **dependence between events**: self-excitation (Hawkes), inhibition (refractory periods), or renewal structure—phenomena survival models with a single event cannot express.
- Accounts for **high-frequency, irregular clinical data** (ICU monitors, EHR logs, alerts) where multiple time stamps per patient are the norm.

Survival: one event time

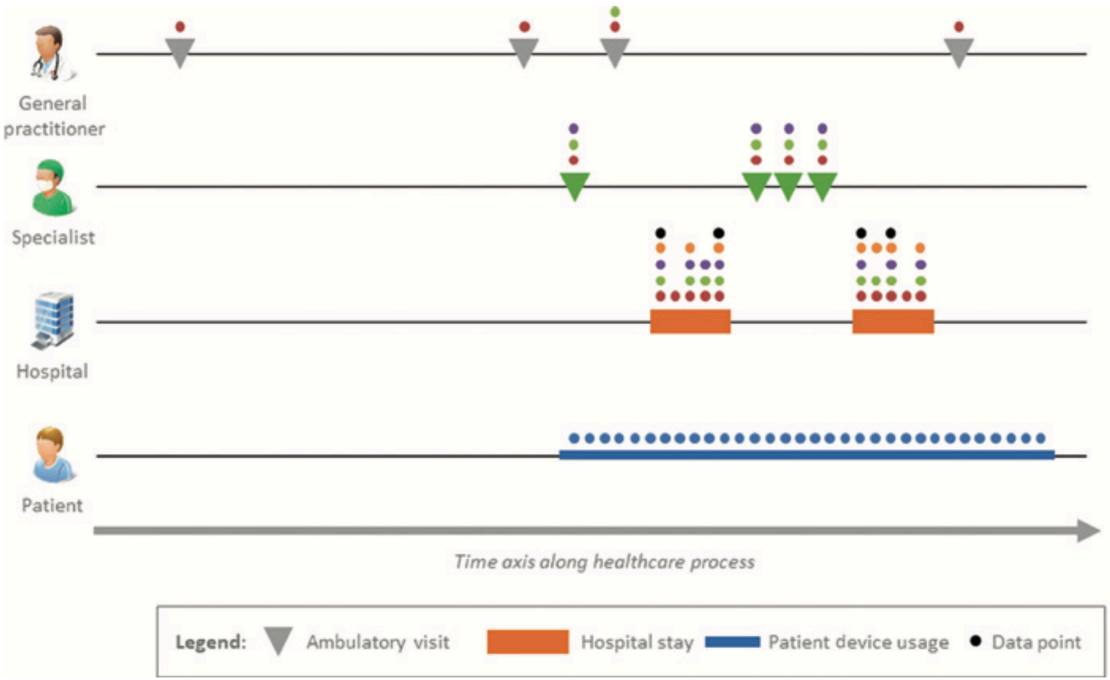
Point process: many events & types (marks)



Conditional intensity $\lambda(t|\mathcal{H}_t)$ (self-exciting)

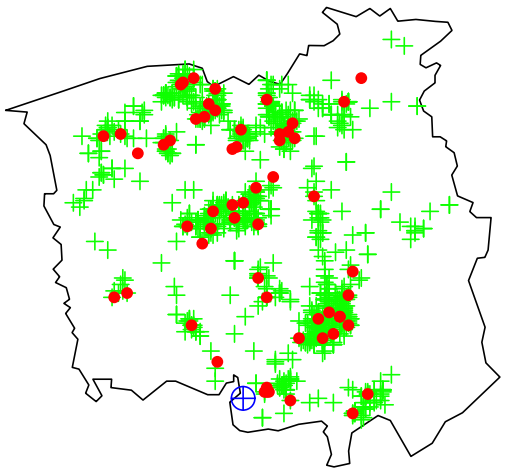


Point Processes to handle irregularly sampled time series



- Temporal point processes are a versatile framework for modeling event sequences in continuous time space

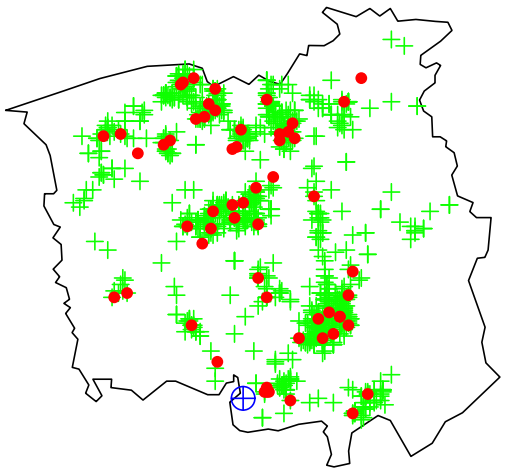
Point Processes to model locations



- Black boundary[a]: The Chorley-Ribble region in Britain (observation window)
- Red dots: Cases of larynx cancer (a point pattern)
- Green plusses: Cases of lung cancer (another point pattern)
- Blue plus/circle: An old incinerator

[a] Taken from: "Introduction to Spatial Point Processes and simulation based inference, Jesper Møller, De Aalborg University, Denmark

Point Processes to model locations



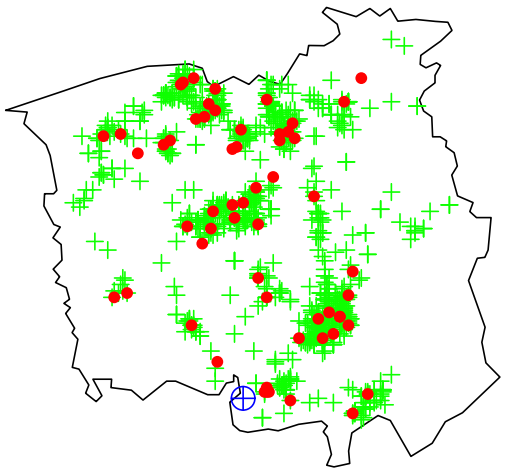
- Black boundary[a]: The Chorley-Ribble region in Britain (observation window)
- Red dots: Cases of larynx cancer (a point pattern)
- Green plusses: Cases of lung cancer (another point pattern)
- Blue plus/circle: An old incinerator

[a] Taken from: "Introduction to Spatial Point Processes and simulation based inference, Jesper Møller, De Aalborg University, Denmark

A spatial point process models the stochastic mechanism that generated the locations

where do events occur? are they randomly scattered or clustered?

Point Processes to model locations



- Black boundary[a]: The Chorley-Ribble region in Britain (observation window)
- Red dots: Cases of larynx cancer (a point pattern)
- Green plusses: Cases of lung cancer (another point pattern)
- Blue plus/circle: An old incinerator

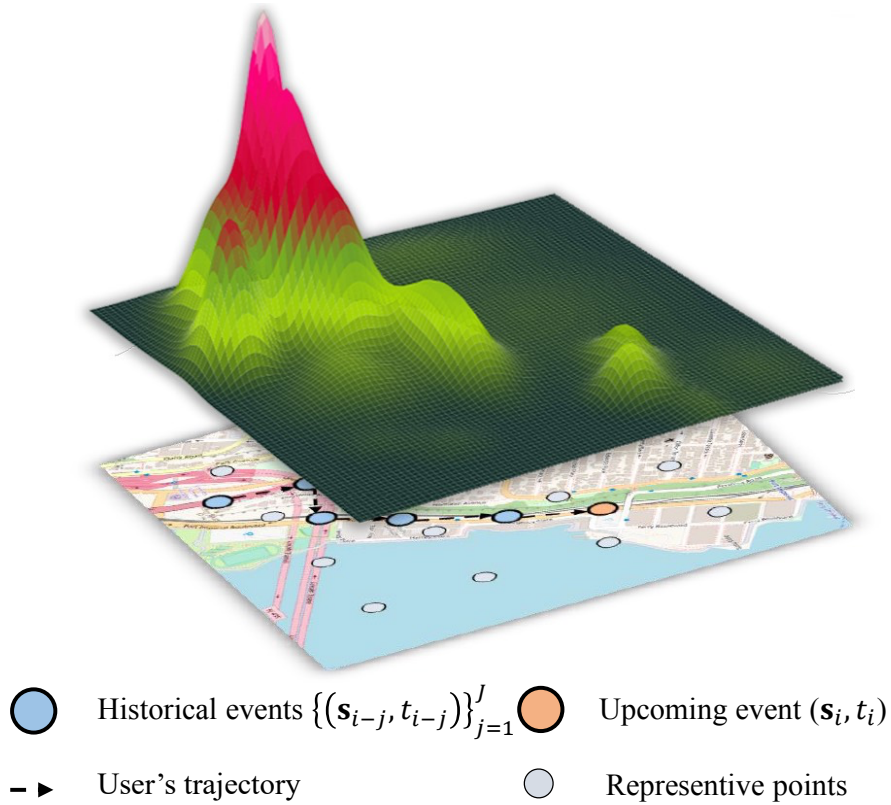
[a] Taken from: "Introduction to Spatial Point Processes and simulation based inference, Jesper Møller, De Aalborg University, Denmark

A spatial point process models the stochastic mechanism that generated the locations

where do events occur? are they randomly scattered or clustered?

- Are larynx cases clustered near the incinerator? [1, 2, 3]
- Do larynx and lung cases occur in the same areas (bivariate clustering)?
- Does risk intensity increase near a specific source?

Point Processes to model spatio-temporal events locations



– Illustration of learning spatio-temporal point process. We aim to learn the space-time intensity function given the historical event sequence and representative points as background.

From [4] Z. Zhou, X. Yang, R. Rossi, H. Zhao, and R. Yu, "Neural Point Process for Learning Spatiotemporal Event Dynamics," *arXiv*, 2021

Point Processes to model spatio-temporal events locations

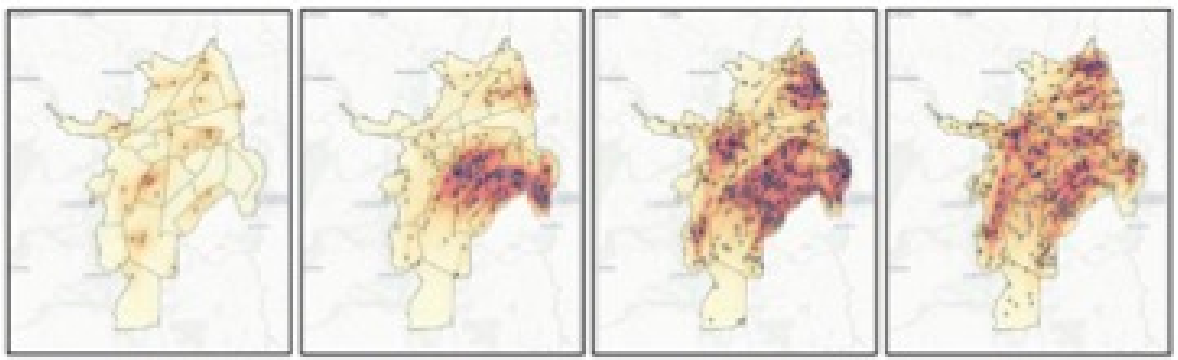


Figure 13. Predicted conditional intensity at four weeks: 22 March, 17 May, 28 June, and 30 August. The black dot represents an actual case reported in that week. The colour depth indicates the conditional intensity at the corresponding location, and a darker colour means a higher risk for citizens to be infected.

From [5] Z. Dong, S. Zhu, Y. Xie, J. Mateu, and F. J. Rodríguez-Cortés, "Non-stationary spatio-temporal point process modeling for high-resolution covid-19 data," *Journal of the Royal Statistical Society Series C: Applied Statistics*, vol. 72, no. 2, p. 368–386, 2023

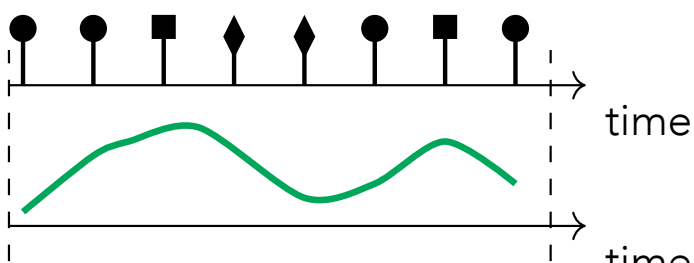
Events as time series

Definition

A stochastic, or random, process composed of a time-series of binary?? events that occur in continuous time [6] localized in $\mathcal{H} = \{t_i\}$.

- data are located at a discrete — countable— set of time points,
- opposed to continuous-valued processes, which can take on any of countless values at each point in time,
- a point process can take on only one of two possible values (binary??), indicating whether or not an event occurs at that time.

Discrete and continuous time series



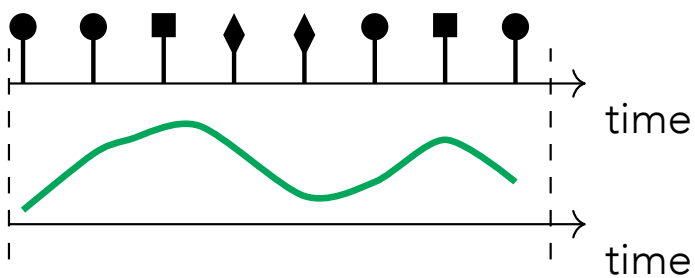
Events as time series

Definition

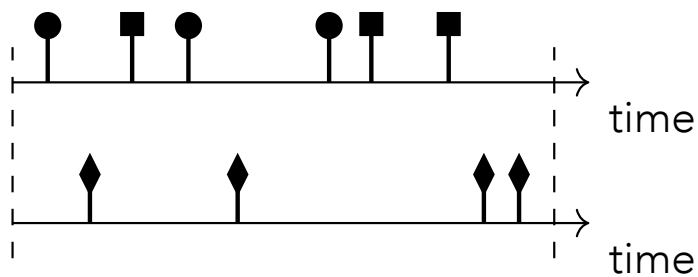
A stochastic, or random, process composed of a time-series of binary?? events that occur in continuous time [6] localized in $\mathcal{H} = \{t_i\}$.

- data are located at a discrete — countable— set of time points,
- opposed to continuous-valued processes, which can take on any of countless values at each point in time,
- a point process can take on only one of two possible values (binary??), indicating whether or not an event occurs at that time.

Discrete and continuous time series



Discrete events in continuous time



Events as time series

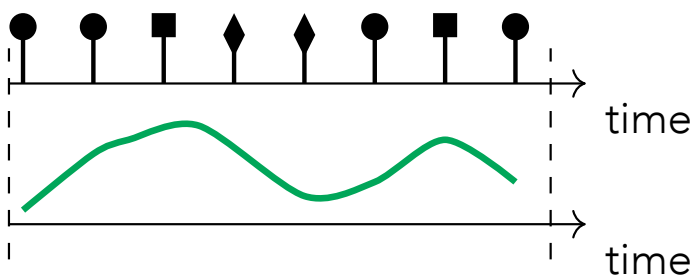
Definition

A stochastic, or random, process composed of a time-series of binary?? events that occur in continuous time [6] localized in $\mathcal{H} = \{t_i\}$.

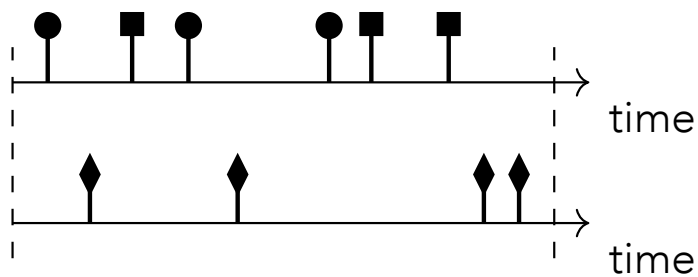
- data are located at a discrete — countable— set of time points,
- opposed to continuous-valued processes, which can take on any of countless values at each point in time,
- a point process can take on only one of two possible values (binary??), indicating whether or not an event occurs at that time.

Temporal point process (TPP) *data are often inappropriately analyzed, because most standard signal-processing techniques are designed primarily for continuous-valued data.*

Discrete and continuous time series

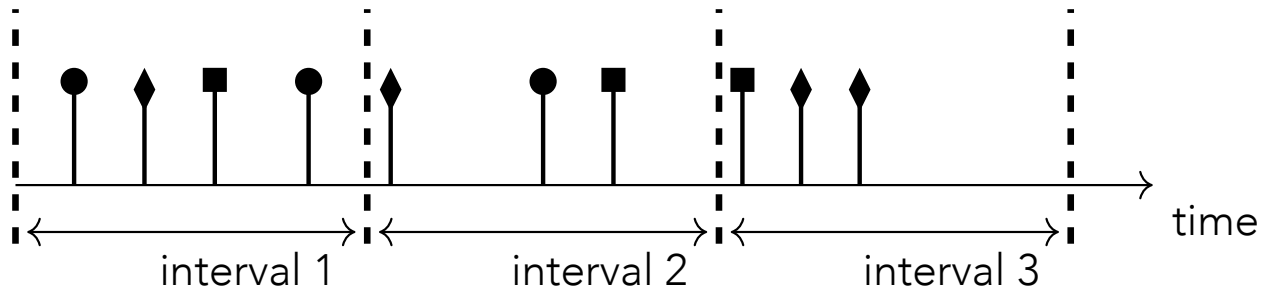


Discrete events in continuous time



Discretizing irregular samples

Aggregating events in intervals



- How to aggregate events per interval?
- What if no event in one interval?
- What about time-related interval?

Why discretizing irregular samples?

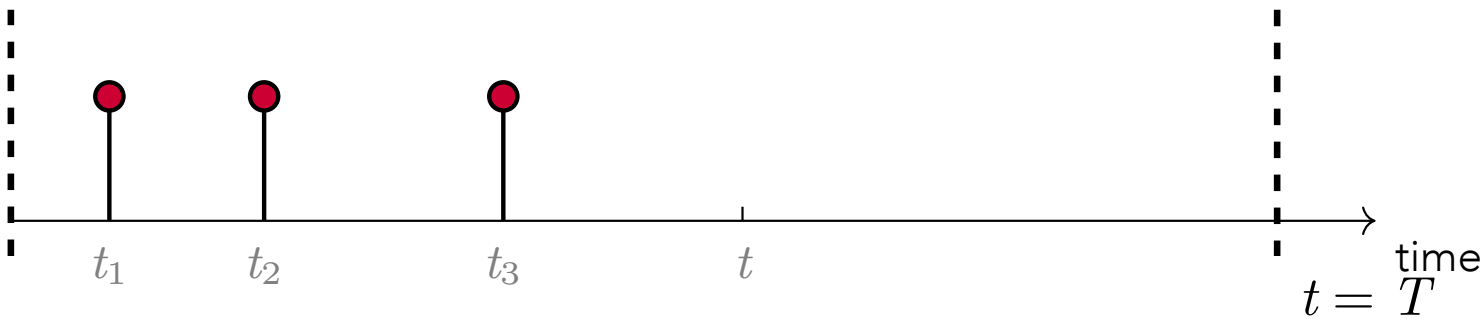
Reason	Detail
Computational tractability	Continuous-time models (like Hawkes or NeuralTPPs) are mathematically elegant but computationally heavy. Discretizing into intervals lets you use standard sequence models (RNNs, Transformers).
Data alignment	Electronic Health Records (EHRs) or sensors rarely produce perfectly time-stamped data; grouping into intervals aligns heterogeneous signals (labs, vitals, meds).
Aggregation & summarization	Within each interval you can compute summary features: counts, averages, min/max, presence indicators.
Handle missing data	Intervals also capture when nothing happened, which is informative for modeling.

Temporal Time processes

Definition

A stochastic, or random, process that models a discrete (countable) time-series of binary events that occur in continuous time [6].

– A realization is a list of events localized in discrete time in $\mathcal{H} = \{t_i\}$, where $t_i \in \mathbb{R}^+$ and $i \in \mathbb{Z}^+$.

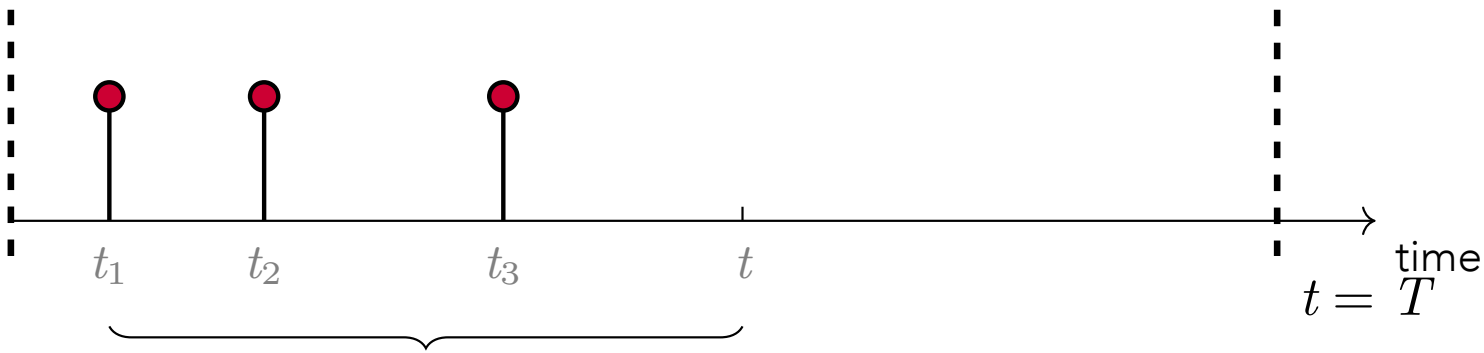


Temporal Time processes

Definition

A stochastic, or random, process that models a discrete (countable) time-series of binary events that occur in continuous time [6].

- A realization is a list of events localized in discrete time in $\mathcal{H} = \{t_i\}$, where $t_i \in \mathbb{R}^+$ and $i \in \mathbb{Z}^+$.



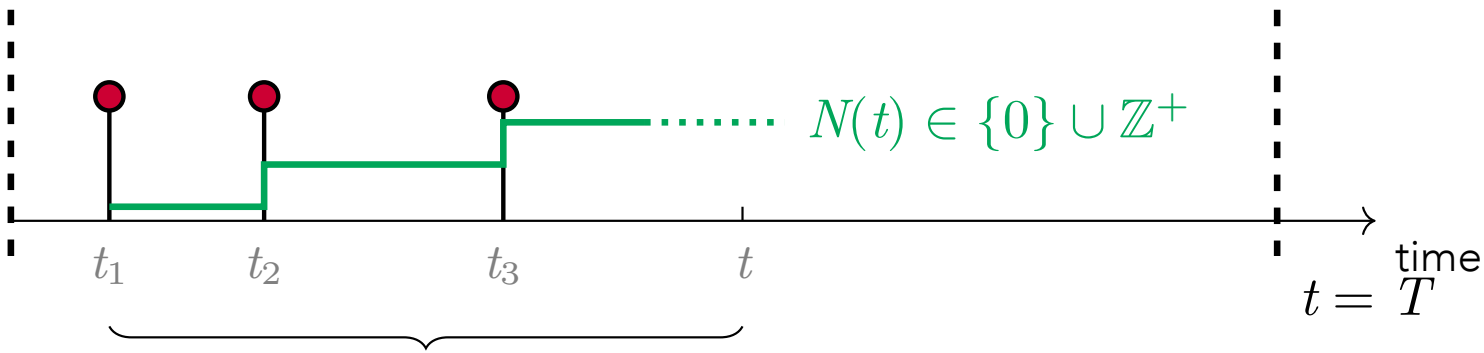
History $\mathcal{H}(t) = \{t_i | t_i < t\}$, up to t!!!!

Temporal Time processes

Definition

A stochastic, or random, process that models a discrete (countable) time-series of binary events that occur in continuous time [6].

- A realization is a list of events localized in discrete time in $\mathcal{H} = \{t_i\}$, where $t_i \in \mathbb{R}^+$ and $i \in \mathbb{Z}^+$.



History $\mathcal{H}(t) = \{t_i | t_i < t\}$, up to t!!!!

If we define $dN(t) = \sum_{t_i \in \mathcal{H}(t)} \delta\left(t - t_i\right) dt$ where $dN(t) \in \{0, 1\}$...

...Then we can formally define the number of events

$$N(t) = \int_0^t dN(s) = \sum_{t_i \in \mathcal{H}(t)} u(t - t_i)$$

- $u(t)$ is the Heaviside step function: $u(t) = 1$ if $t \geq 0$ and $u(t) = 0$ otherwise.

Model time as a random variable

Each event time t in a TPP is a random variable:

Given $\mathcal{H}(t) = \{t_1, \dots t_{i-1}\}$, where t_{i-1} is the last event in $\mathcal{H}(t)$ before t, we aim to characterize the time t of the next event, the i-th event:

Model time as a random variable

Each event time t in a TPP is a random variable:

Given $\mathcal{H}(t) = \{t_1, \dots t_{i-1}\}$, where t_{i-1} is the last event in $\mathcal{H}(t)$ before t, we aim to characterize the time t of the next event, the i-th event:

- A conditional pdf:

$$f^*(t) = f(t|\mathcal{H}(t)),$$

which is the probability that the next event will occur during the interval $[t, t + dt)$ conditioned on the history $\mathcal{H}(t)$.

Model time as a random variable

Each event time t in a TPP is a random variable:

Given $\mathcal{H}(t) = \{t_1, \dots t_{i-1}\}$, where t_{i-1} is the last event in $\mathcal{H}(t)$ before t, we aim to characterize the time t of the next event, the i-th event:

- A conditional pdf:

$$f^*(t) = f(t|\mathcal{H}(t)),$$

  which is the probability that the next event will occur during the interval $[t, t + dt)$ conditioned on the history $\mathcal{H}(t)$.

- A cumulative distribution function:

$$F^*(t) = F(t|\mathcal{H}(t)) = \int_{t_{i-1}}^{t} f^*(\tau)\, d\tau$$

  which is the probability that the next event will occur before time t conditioned on the history $\mathcal{H}(t)$.

Model time as a random variable

Each event time t in a TPP is a random variable:

Given $\mathcal{H}(t) = \{t_1, \dots t_{i-1}\}$, where t_{i-1} is the last event in $\mathcal{H}(t)$ before t, we aim to characterize the time t of the next event, the i-th event:

- A conditional pdf:

$$f^*(t) = f(t|\mathcal{H}(t)),$$

  which is the probability that the next event will occur during the interval $[t, t + dt)$ conditioned on the history $\mathcal{H}(t)$.

- A cumulative distribution function:

$$F^*(t) = F(t|\mathcal{H}(t)) = \int_{t_{i-1}}^{t} f^*(\tau)\, d\tau$$

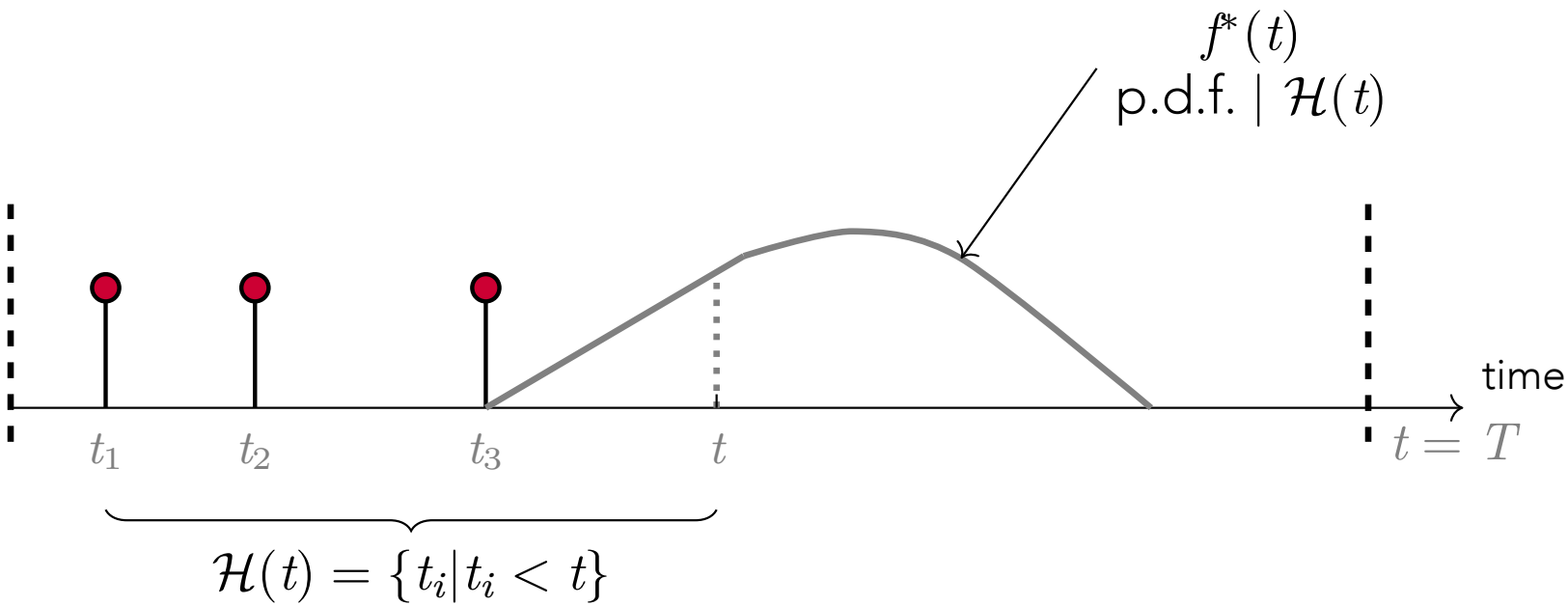
  which is the probability that the next event will occur before time t conditioned on the history $\mathcal{H}(t)$.

- A complementary cumulative distribution function, also called survival function

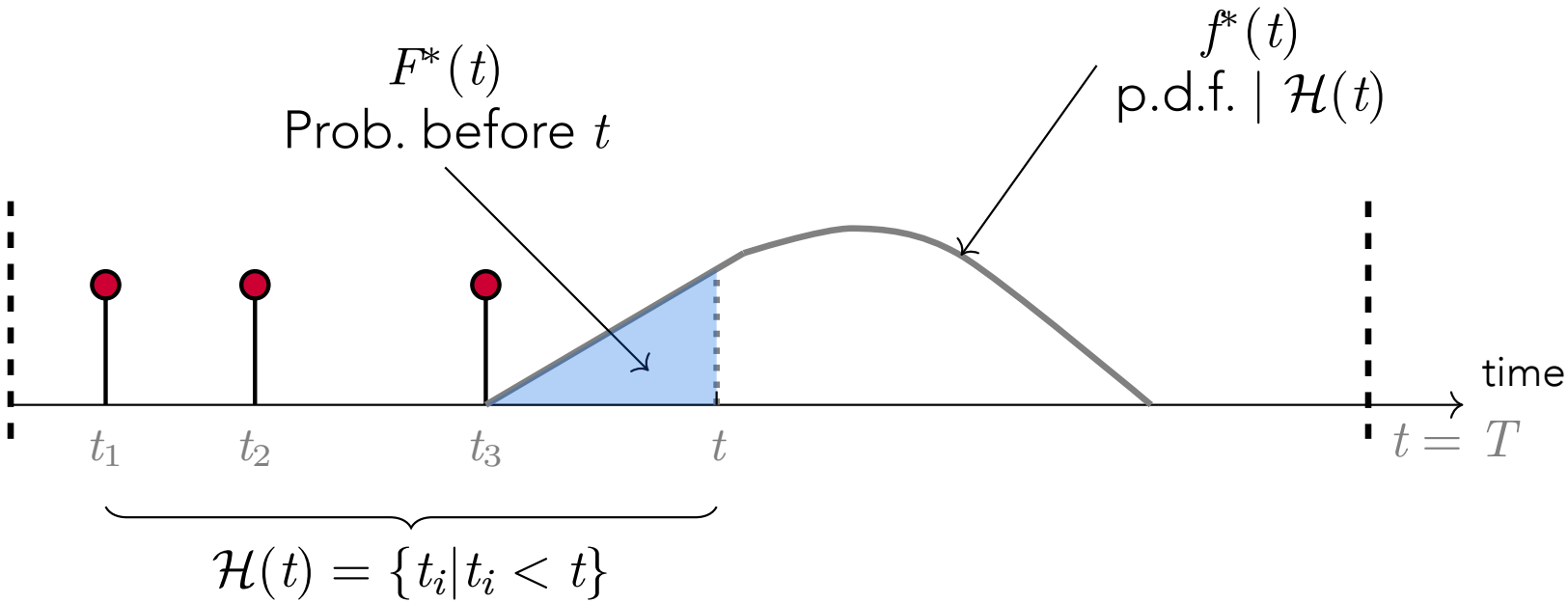
$$S^*(t) = S(t|\mathcal{H}(t)) = 1 - F^*(t),$$

  which is the probability that the next event will not occur before time t conditioned on the history $\mathcal{H}(t)$.

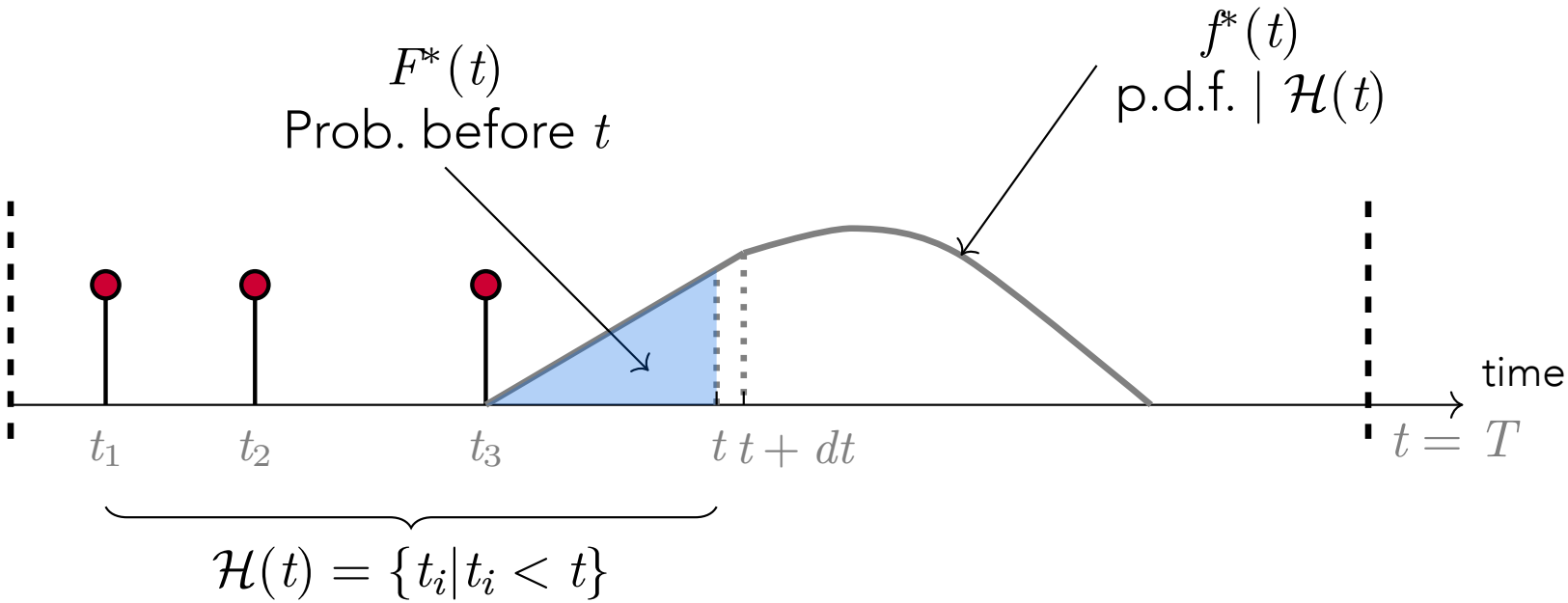
Model time as a random variable



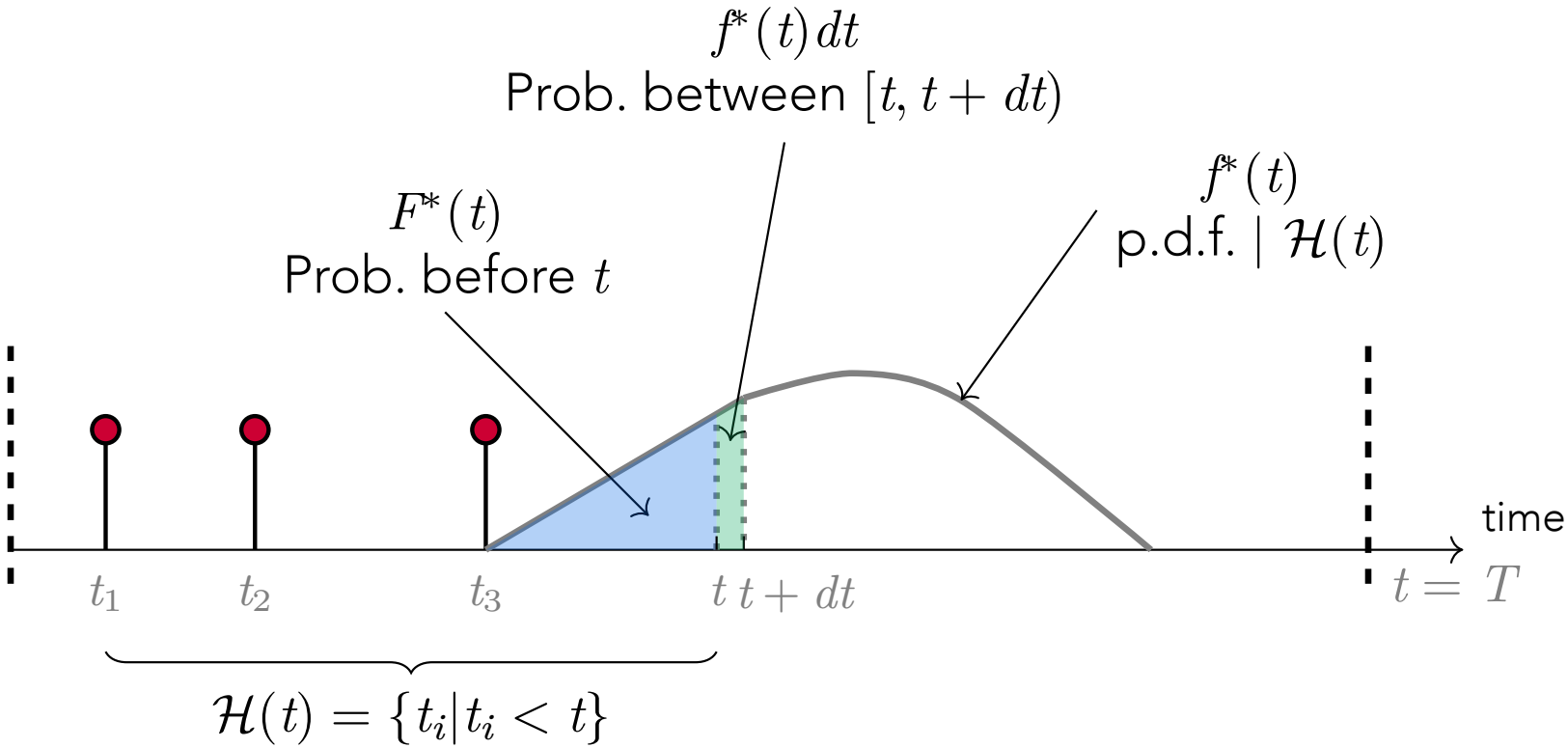
Model time as a random variable



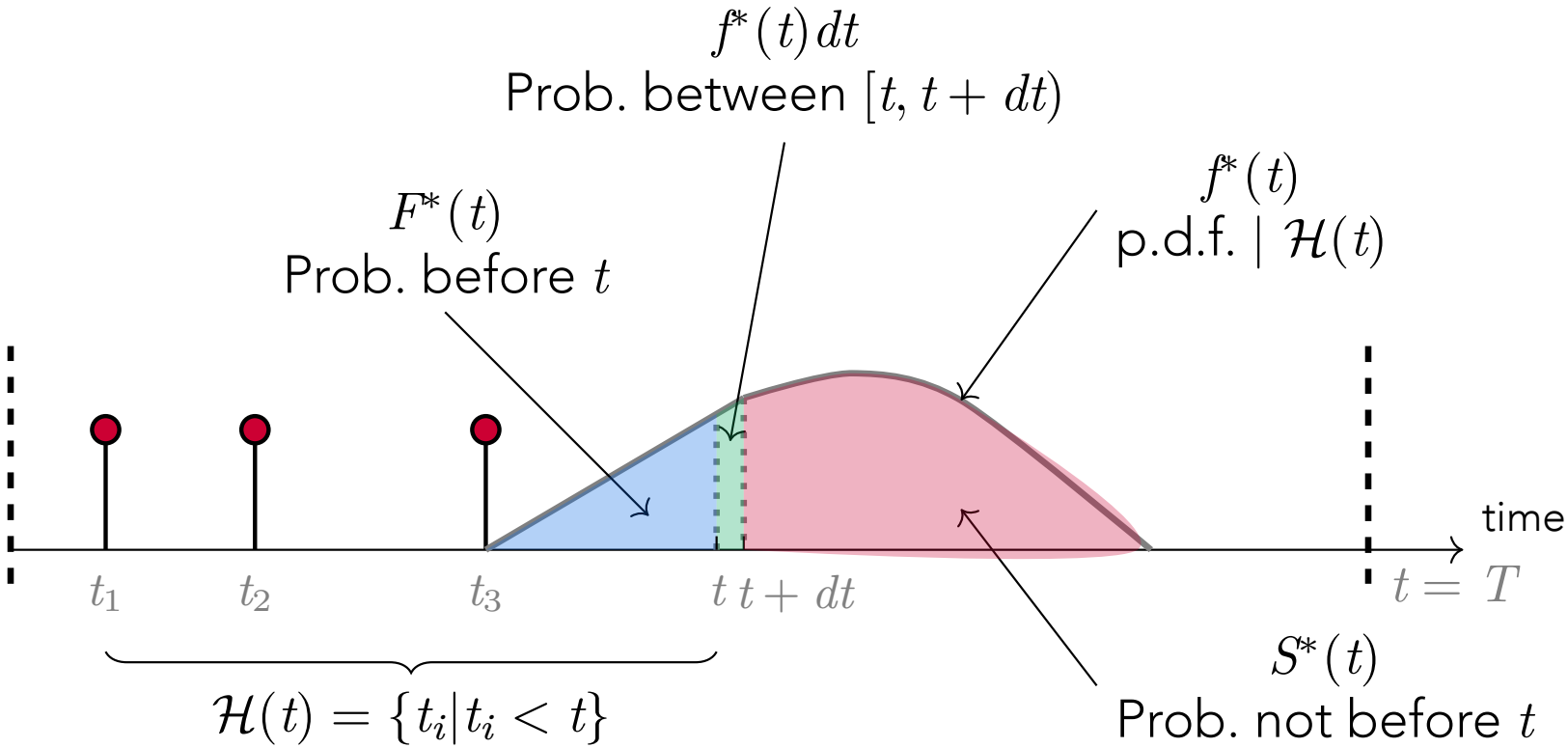
Model time as a random variable



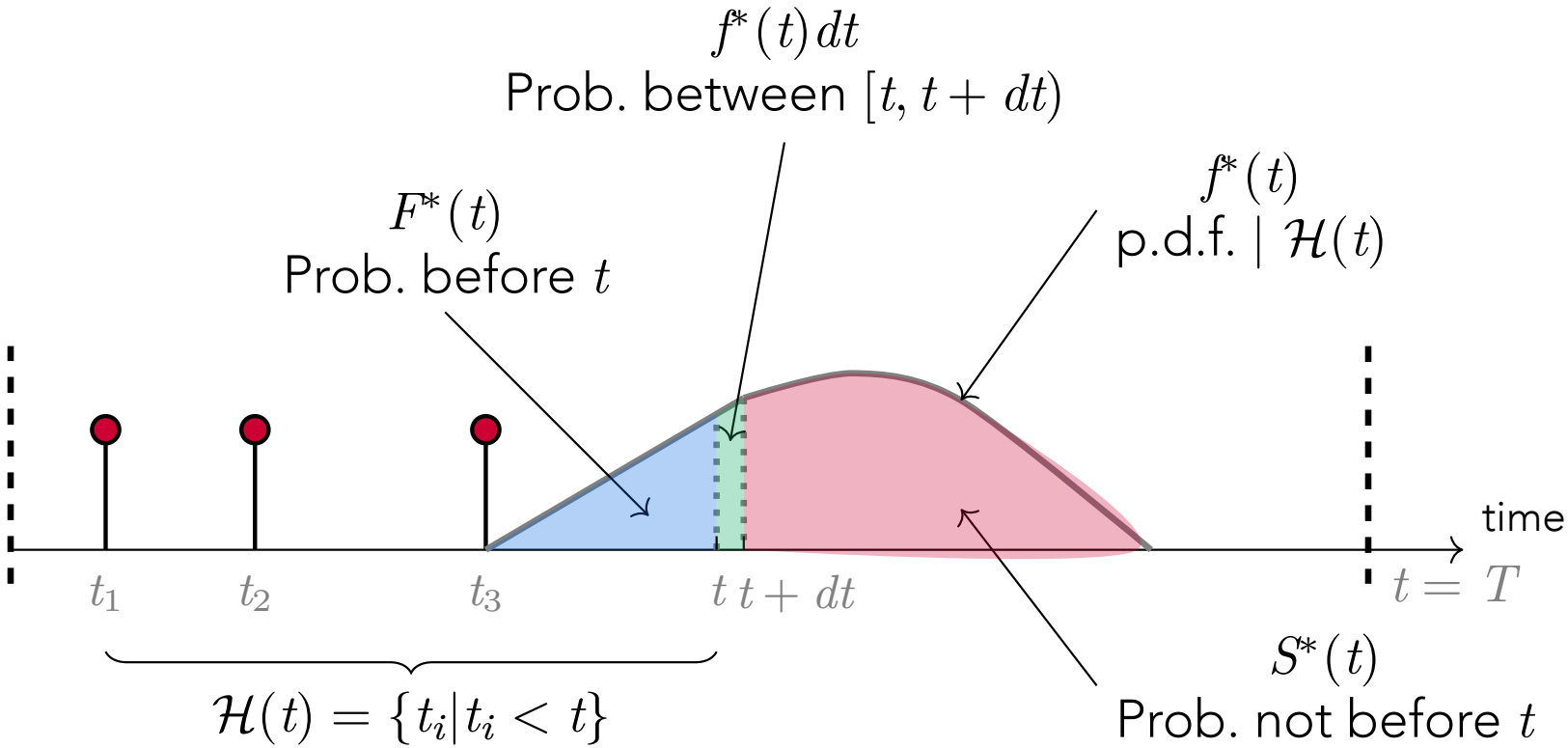
Model time as a random variable

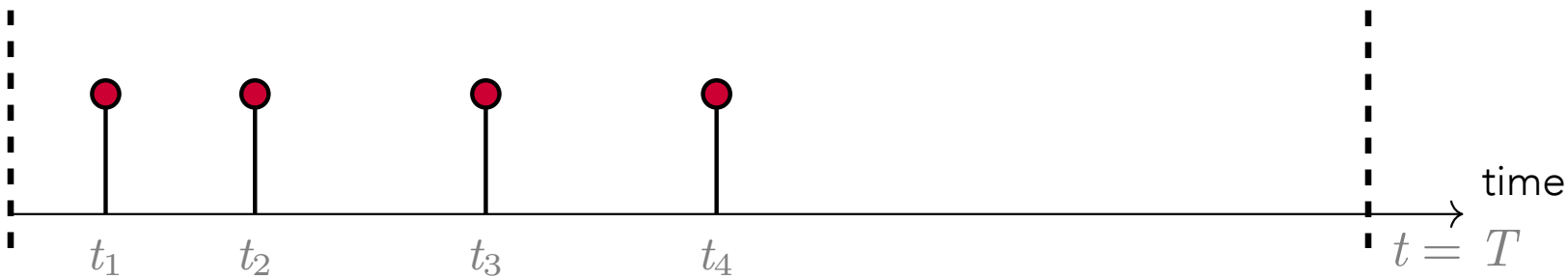


Model time as a random variable



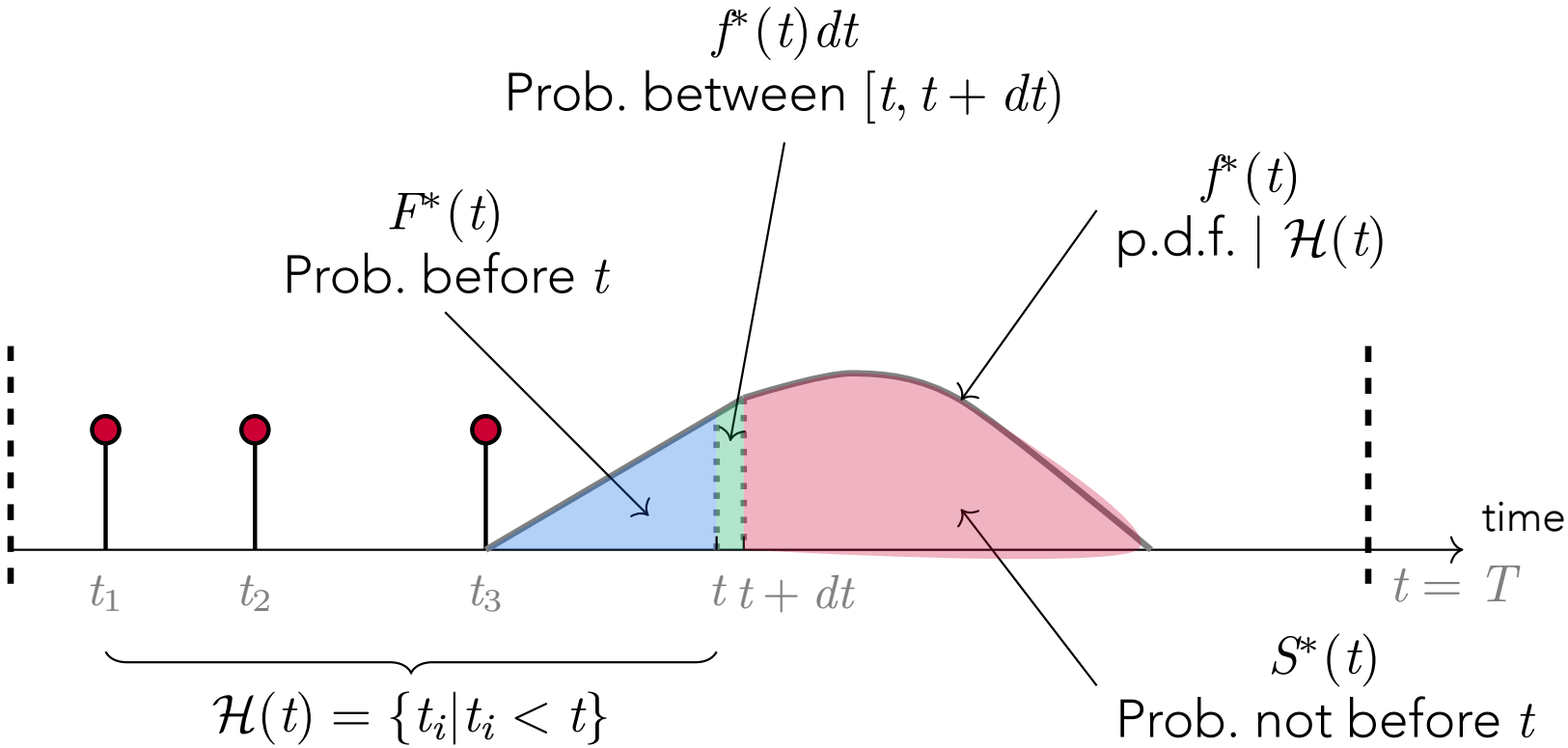
Model time as a random variable

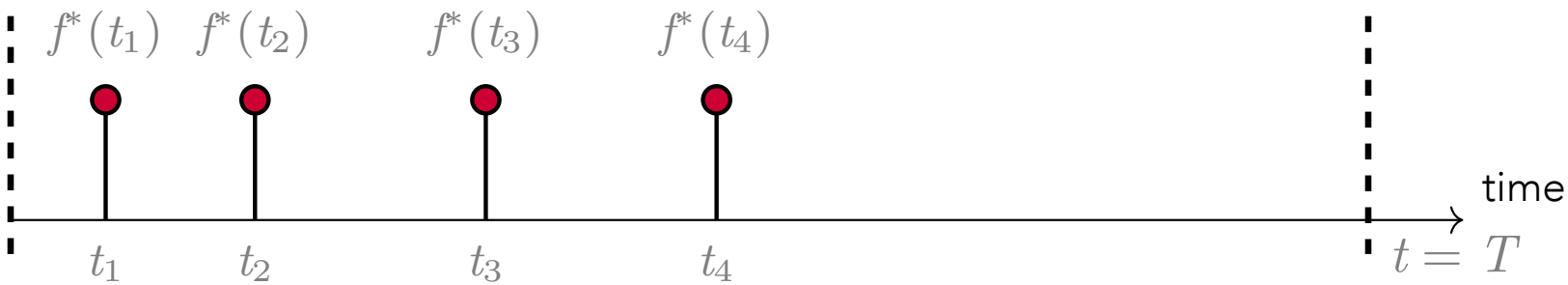




Likelihood of a timeline

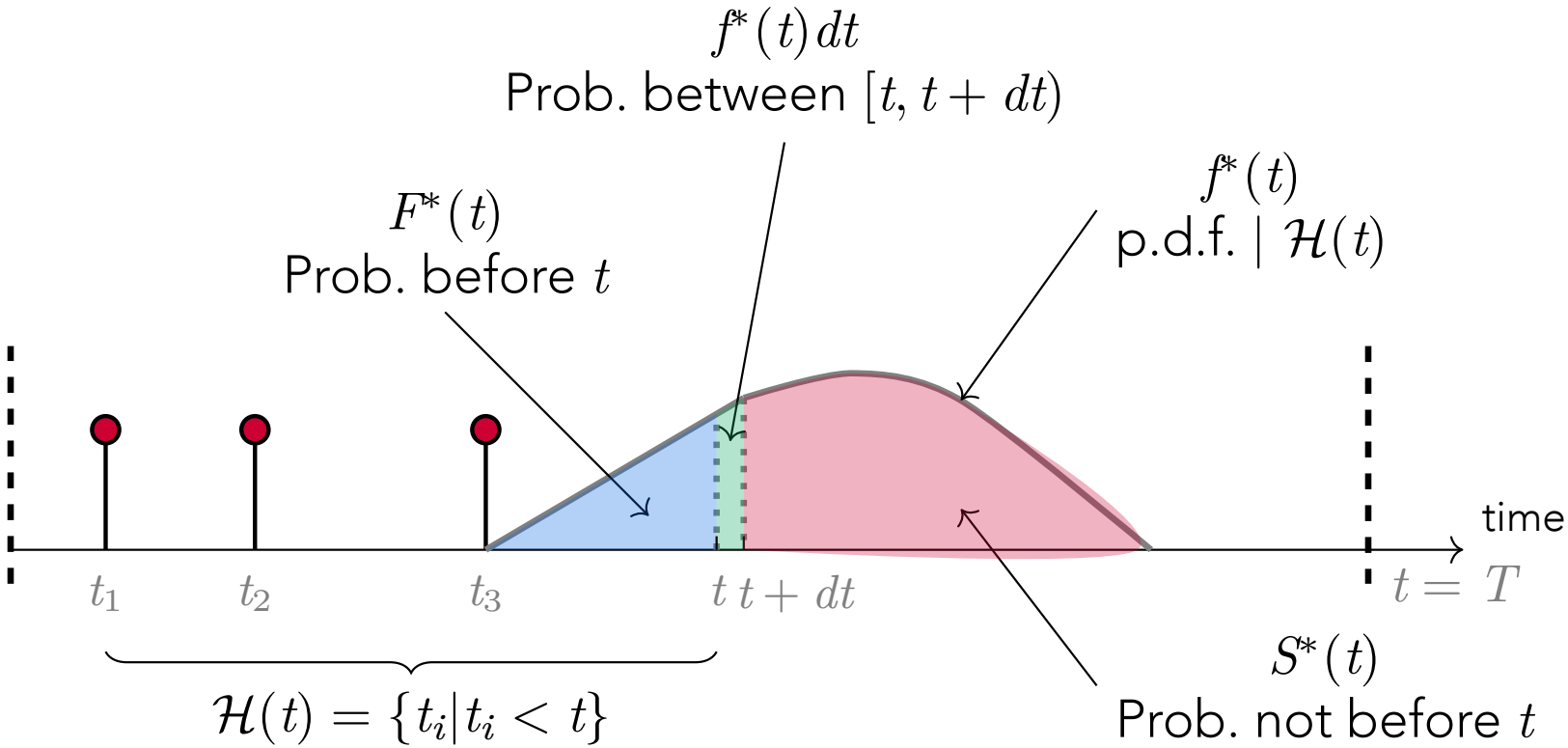
Model time as a random variable

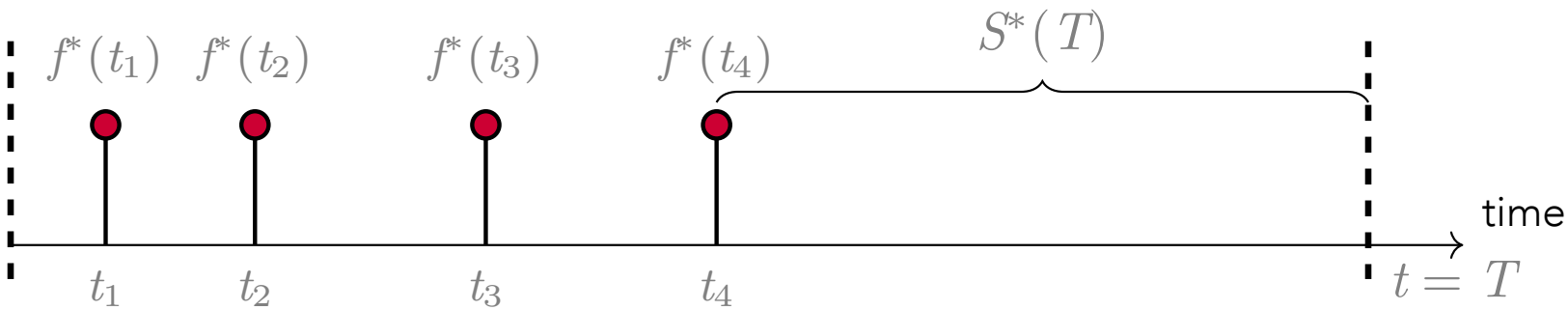




Likelihood of a timeline

Model time as a random variable





Likelihood of a timeline $\longrightarrow f^*(t_1) \times \cdots \times f^*(t_4) S^*(T)$

Model time as a random variable

Using the above functions, difficulties on the model design, interpretability and reusability arise:

- **Model design and interpretability**: difficult to build an intuition about how the functional form for $f^*(t)$ can be and how to associate specific regions of the function with interpretable mechanisms related to the events or temporal effects. Furthermore, we need that $f^*(t)$ is such that $\int_{t_{i-1}}^{\infty} f^*(\tau)d\tau = 1$ in order for it to be a valid pdf.
- **Model reusability**: difficult to combine several TPPs models or *timelines*.

Model time as a random variable

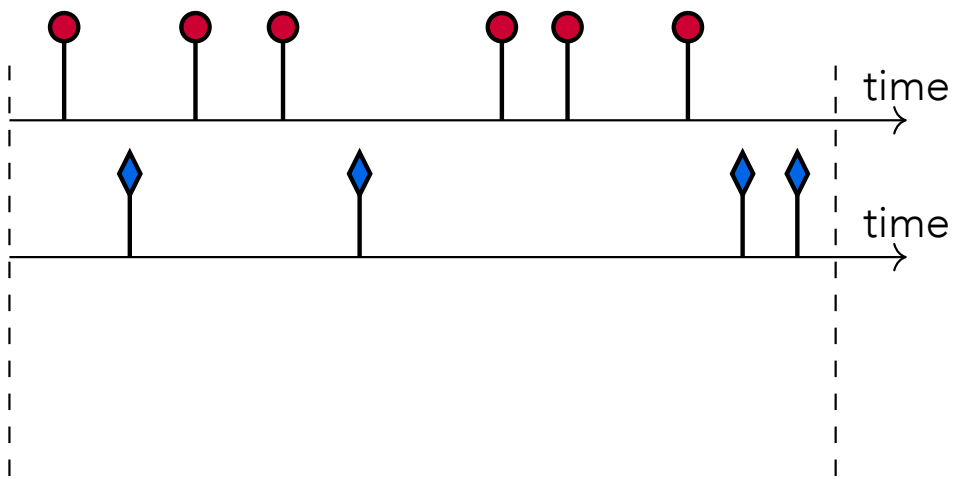
Example

We accurately characterize the time of the next event in two independent TPPs with histories $\mathcal{H}_1(t)$ and $\mathcal{H}_2(t)$ using $f_1^*(t)$ and $f_2^*(t)$.

- $\mathcal{H}_1(t)$: history of occurrences of atrial fibrillation episodes
- $\mathcal{H}_2(t)$: occurrences of ventricular tachycardia episodes

Highly nontrivial characterization of the time of the next event in the joint TPP with history $\mathcal{H}(t) = \mathcal{H}_1(t) \cup \mathcal{H}_2(t)$:

- $\mathcal{H}(t) = \mathcal{H}_1(t) \cup \mathcal{H}_2(t)$: history of all arrhythmic events.



Model time as a random variable

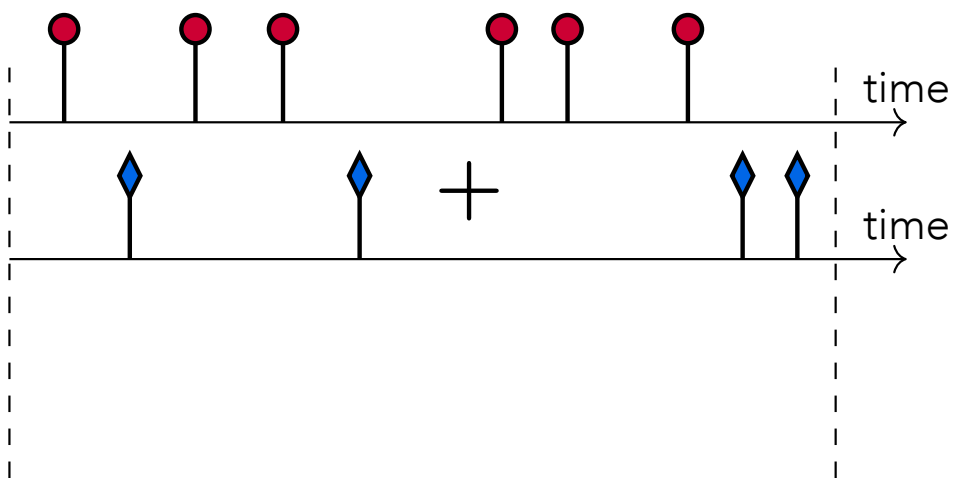
Example

We accurately characterize the time of the next event in two independent TPPs with histories $\mathcal{H}_1(t)$ and $\mathcal{H}_2(t)$ using $f_1^*(t)$ and $f_2^*(t)$.

- $\mathcal{H}_1(t)$: history of occurrences of atrial fibrillation episodes
- $\mathcal{H}_2(t)$: occurrences of ventricular tachycardia episodes

Highly nontrivial characterization of the time of the next event in the joint TPP with history $\mathcal{H}(t) = \mathcal{H}_1(t) \cup \mathcal{H}_2(t)$:

- $\mathcal{H}(t) = \mathcal{H}_1(t) \cup \mathcal{H}_2(t)$: history of all arrhythmic events.



Model time as a random variable

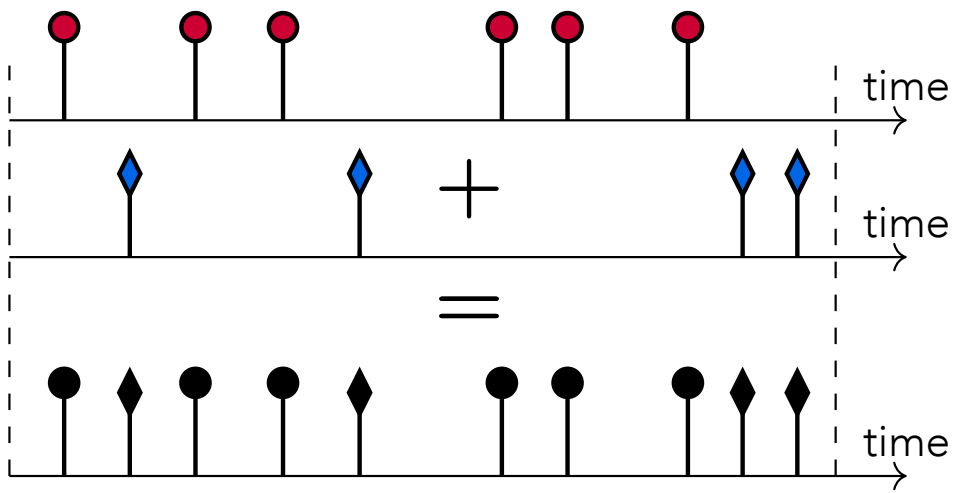
Example

We accurately characterize the time of the next event in two independent TPPs with histories $\mathcal{H}_1(t)$ and $\mathcal{H}_2(t)$ using $f_1^*(t)$ and $f_2^*(t)$.

- $\mathcal{H}_1(t)$: history of occurrences of atrial fibrillation episodes
- $\mathcal{H}_2(t)$: occurrences of ventricular tachycardia episodes

Highly nontrivial characterization of the time of the next event in the joint TPP with history $\mathcal{H}(t) = \mathcal{H}_1(t) \cup \mathcal{H}_2(t)$:

- $\mathcal{H}(t) = \mathcal{H}_1(t) \cup \mathcal{H}_2(t)$: history of all arrhythmic events.



Model time as a random variable

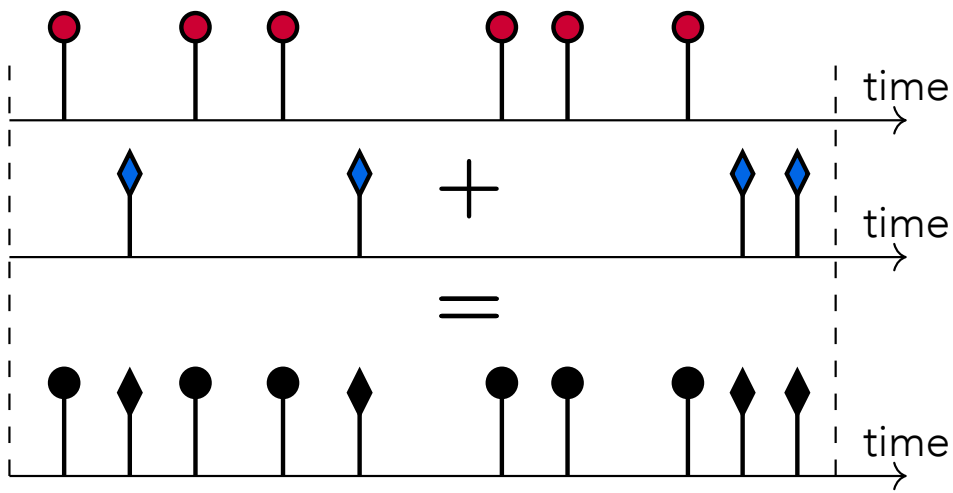
Example

We accurately characterize the time of the next event in two independent TPPs with histories $\mathcal{H}_1(t)$ and $\mathcal{H}_2(t)$ using $f_1^*(t)$ and $f_2^*(t)$.

- $\mathcal{H}_1(t)$: history of occurrences of atrial fibrillation episodes
- $\mathcal{H}_2(t)$: occurrences of ventricular tachycardia episodes

Highly nontrivial characterization of the time of the next event in the joint TPP with history $\mathcal{H}(t) = \mathcal{H}_1(t) \cup \mathcal{H}_2(t)$:

- $\mathcal{H}(t) = \mathcal{H}_1(t) \cup \mathcal{H}_2(t)$: history of all arrhythmic events.



$$f^*(t) \neq f_1^*(t_1) + f_2^*(t_2)$$
$$f^*(t) \neq f_1^*(t_1) \star f_2^*(t_2)$$

Intensity (hazard) function

Conditional intensity function:

Central modeling object in TPP: $\lambda^*(t) = \lambda(t|\mathcal{H}(t))$, conditional probability that the next event will happen during $[t, t+dt)$, given $\mathcal{H}(t) = \{t_1, \dots t_{i-1}\}$ and the fact that it has not happened before t, i.e.,

$$\lambda^*(t) = \lim_{dt \to 0} \frac{\Pr\left(N(t+dt) - N(t) = 1 | \mathcal{H}(t)\right)}{dt} = \lim_{dt \to 0} \frac{F^*(t+dt) - F^*(t)}{S^*(t)\, dt}$$

$$= \frac{f^*(t)}{S^*(t)} \geq 0$$

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

Given that $dN(t) \in \{0, 1\}$ and that it can only increase by one event at each dt, we also have

$$\Pr\left(dN(t) = 1 | \mathcal{H}(t)\right) = \lambda^*(t)\, dt$$

Furthermore:

$$\mathbb{E}\left\{dN(t) | \mathcal{H}(t)\right\} = 1 \times \Pr\left(dN(t) = 1 | \mathcal{H}(t)\right) + 0 \times \Pr\left(dN(t) = 0 | \mathcal{H}(t)\right) = \lambda^*(t)\, dt$$

Intensity (hazard) function

Advantages of using $\lambda^*(t)$ for the model design and interpretability:

- **Model design and interpretability**: using the 'ìdea" of $\lambda^*(t)$ being a rate, it is easy to build an intuition about the choice of its functional form. We only need to guarantee that $\lambda^*(t) \geq 0$.

Intensity (hazard) function

Advantages of using $\lambda^*(t)$ for the model design and interpretability:

- **Model design and interpretability**: using the 'ìdea" of $\lambda^*(t)$ being a rate, it is easy to build an intuition about the choice of its functional form. We only need to guarantee that $\lambda^*(t) \geq 0$.

Designing via the intensity function allows you to build models
locally $\rightarrow$ specify rate at each time t,
causally $\rightarrow$ naturally depends on history $\mathcal{H}(t)$ and
modularly $\rightarrow$ can add, remove, or constrain terms easily for interpretation.
Each component (baseline, covariate, excitation, decay) corresponds to a clear physiological or clinical mechanism.
In contrast, pdf-based designs are globally constrained and harder to interpret or update.

Intensity (hazard) function

Example

Model for readmissions of a chronic patient after hospital discharge.

$$\lambda^*(t|\mathcal{H}(t)) = \underbrace{\mu}_{\text{baseline rate}} + \underbrace{\alpha\, e^{-\beta(t-t_{\text{last}})}}_{\text{transient effect of last admission}} + \underbrace{\gamma^\top x(t)}_{\text{effect of covariates (labs, vitals, etc.)}}.$$

Intensity (hazard) function

Example

Model for readmissions of a chronic patient after hospital discharge.

$$\lambda^*(t|\mathcal{H}(t)) = \underbrace{\mu}_{\text{baseline rate}} + \underbrace{\alpha\, e^{-\beta(t-t_{\text{last}})}}_{\text{transient effect of last admission}} + \underbrace{\gamma^\top x(t)}_{\text{effect of covariates (labs, vitals, etc.)}}.$$

- μ: constant baseline hazard (patients always at some risk).
- $\alpha e^{-\beta(t-t_{\text{last}})}$: wen a patient is discharged from hospital after a serious event, they are not immediately back to baseline risk. There's usually a high short-term risk of being readmitted (relapse, infection, etc.) which then decreases gradually as the patient recovers.
- $x(t)$: time-varying covariates (e.g. inflammation markers, medication...).
- γ: weights (learned from data) showing covariate's contribution to the instantaneous readmission risk.

Intensity (hazard) function

Example

Model for readmissions of a chronic patient after hospital discharge.

$$\lambda^*(t|\mathcal{H}(t)) = \underbrace{\mu}_{\text{baseline rate}} + \underbrace{\alpha\, e^{-\beta(t-t_{\text{last}})}}_{\text{transient effect of last admission}} + \underbrace{\gamma^\top x(t)}_{\text{effect of covariates (labs, vitals, etc.)}}.$$

This intensity is:

- interpretable (each term maps to a physiological or behavioral factor),
- modular (you can add, remove, or constrain terms easily),
- and nonnegative (as required).

Intensity (hazard) function

Example

Modeling arrhythmia episodes (self-exciting process):

Cardiac arrhythmia episodes often cluster: one event increases short-term risk of another. Naturally, a Hawkes process captures this:

$$\lambda(t|\mathcal{H}(t)) = \mu + \sum_{t_i<t} \alpha\, e^{-\beta(t-t_i)}$$

Each previous arrhythmia $\mathcal{H}(t)$ temporarily raises the hazard; the effect decays exponentially.

- μ: baseline rate of spontaneous arrhythmias,
- α: magnitude of excitation after an episode (severity or persistence),
- β^{-1}: average duration of post-event vulnerability. example:
  - $\beta = 1/30$ per day $\rightarrow$ recovery time ≈ 30 days $\rightarrow$ risk returns to baseline after 1 month.
  - $\beta = 1/5$ per day $\rightarrow$ recovery time ≈ 5 days $\rightarrow$ risk drops faster.

Clinically, α and β relate to refractoriness and recovery time — interpretable physiological quantities.

General structure for an intensity model

The most common and interpretable way to design the conditional intensity in healthcare event data is an additive model:

$$\lambda(t|\mathcal{H}(t), x(t)) = \underbrace{\mu}_{\text{baseline}} + \underbrace{\text{history effects}}_{\text{self– or cross–excitation}} + \underbrace{\text{covariate effects}}_{\text{current or time-varying predictors}}$$

Why additive combination makes sense?

- It mirrors the superposition principle: independent sources of risk combine by addition.
- Each parameter controls one aspect of the physiology:
  - μ: baseline risk,
  - α, β: event-triggered transient effect,
  - γ: sensitivity to covariates.
- If new predictors or risk factors are discovered, you just add another term — no need to redesign or renormalize the whole pdf.

General structure for an intensity model

Alternative multiplicative (log-linear) version:

$$\lambda(t|\mathcal{H}(t), x(t)) = \lambda_0(t) \times \exp\left(\gamma^\top x(t) + \sum_{t_i<t} \alpha e^{-\beta(t-t_i)}\right)$$

- equivalent to applying an exponential link to the additive sum in the exponent.
- It keeps $\lambda(t|\mathcal{H}(t), x(t))$ positive automatically and allows interpreting γ as a log–relative risk.
- It's still "additive in the linear predictor," but multiplicative in the final rate. So conceptually you're still summing the effects before the link.

Intensity (hazard) function

Advantage of using $\lambda^*(t)$ for the reusability:

- **Model reusability**: we can easily combine several TPPs models. Given two independent[2] TPP, with histories $\mathcal{H}_1(t)$ and $\mathcal{H}_2(t)$, using intensities $\lambda_1^*(t)$ and $\lambda_2^*(t)$, respectively, then, we can characterize the joint history $\mathcal{H}(t) = \mathcal{H}_1(t) \cup \mathcal{H}_2(t)$ by means of:

$$\begin{aligned}\lambda^*(t)\,dt &= \mathbb{E}\left\{dN(t)|\mathcal{H}(t)\right\} = \mathbb{E}\left\{dN_1(t) + dN_2(t)|\mathcal{H}(t)\right\} = \\ &= \mathbb{E}\left\{dN_1(t)|\mathcal{H}_1(t)\right\} + \mathbb{E}\left\{dN_2(t)|\mathcal{H}_2(t)\right\} = \\ &= \lambda_1^*(t)\,dt + \lambda_2^*(t)\,dt\end{aligned}$$

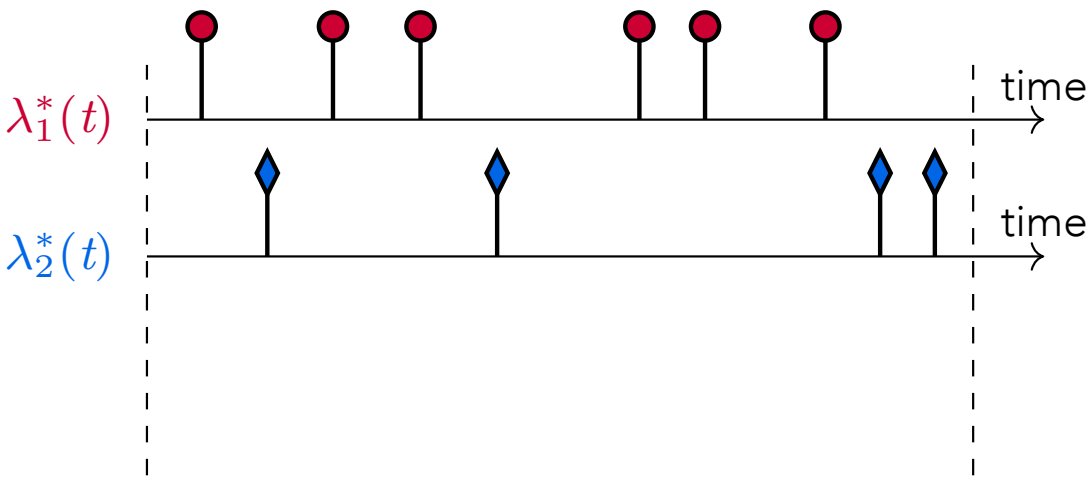
[2] In reality, if the conditional intensities $\lambda_1^*(t)$ and $\lambda_2^*(t)$ are conditioned on the joint history $\mathcal{H}(t) = \mathcal{H}_1(t) \cup \mathcal{H}_2(t)$, it is only necessary that both processes are conditionally independent given $\mathcal{H}(t)$. However, for ease of exposition, independence is assumed.

Intensity (hazard) function

Advantage of using $\lambda^*(t)$ for the reusability:

- **Model reusability**: we can easily combine several TPPs models. Given two independent[2] TPP, with histories $\mathcal{H}_1(t)$ and $\mathcal{H}_2(t)$, using intensities $\lambda_1^*(t)$ and $\lambda_2^*(t)$, respectively, then, we can characterize the joint history $\mathcal{H}(t) = \mathcal{H}_1(t) \cup \mathcal{H}_2(t)$ by means of:

$$\begin{aligned}\lambda^*(t)\,dt &= \mathbb{E}\left\{dN(t)|\mathcal{H}(t)\right\} = \mathbb{E}\left\{dN_1(t) + dN_2(t)|\mathcal{H}(t)\right\} = \\ &= \mathbb{E}\left\{dN_1(t)|\mathcal{H}_1(t)\right\} + \mathbb{E}\left\{dN_2(t)|\mathcal{H}_2(t)\right\} = \\ &= \lambda_1^*(t)\,dt + \lambda_2^*(t)\,dt\end{aligned}$$



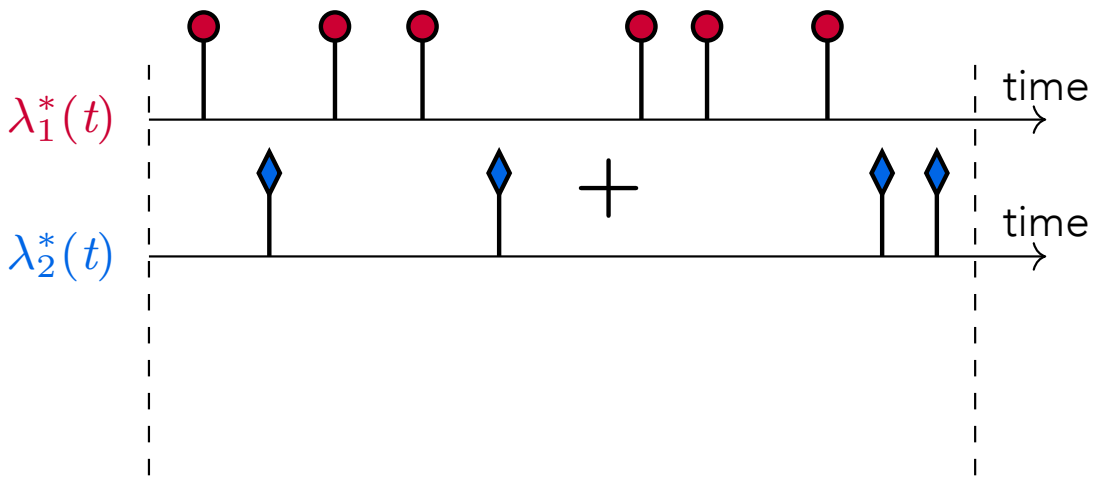
[2] In reality, if the conditional intensities $\lambda_1^*(t)$ and $\lambda_2^*(t)$ are conditioned on the joint history $\mathcal{H}(t) = \mathcal{H}_1(t) \cup \mathcal{H}_2(t)$, it is only necessary that both processes are conditionally independent given $\mathcal{H}(t)$. However, for ease of exposition, independence is assumed.

Intensity (hazard) function

Advantage of using $\lambda^*(t)$ for the reusability:

- **Model reusability**: we can easily combine several TPPs models. Given two independent[2] TPP, with histories $\mathcal{H}_1(t)$ and $\mathcal{H}_2(t)$, using intensities $\lambda_1^*(t)$ and $\lambda_2^*(t)$, respectively, then, we can characterize the joint history $\mathcal{H}(t) = \mathcal{H}_1(t) \cup \mathcal{H}_2(t)$ by means of:

$$\begin{aligned}\lambda^*(t)\,dt &= \mathbb{E}\left\{dN(t)|\mathcal{H}(t)\right\} = \mathbb{E}\left\{dN_1(t) + dN_2(t)|\mathcal{H}(t)\right\} = \\ &= \mathbb{E}\left\{dN_1(t)|\mathcal{H}_1(t)\right\} + \mathbb{E}\left\{dN_2(t)|\mathcal{H}_2(t)\right\} = \\ &= \lambda_1^*(t)\,dt + \lambda_2^*(t)\,dt\end{aligned}$$



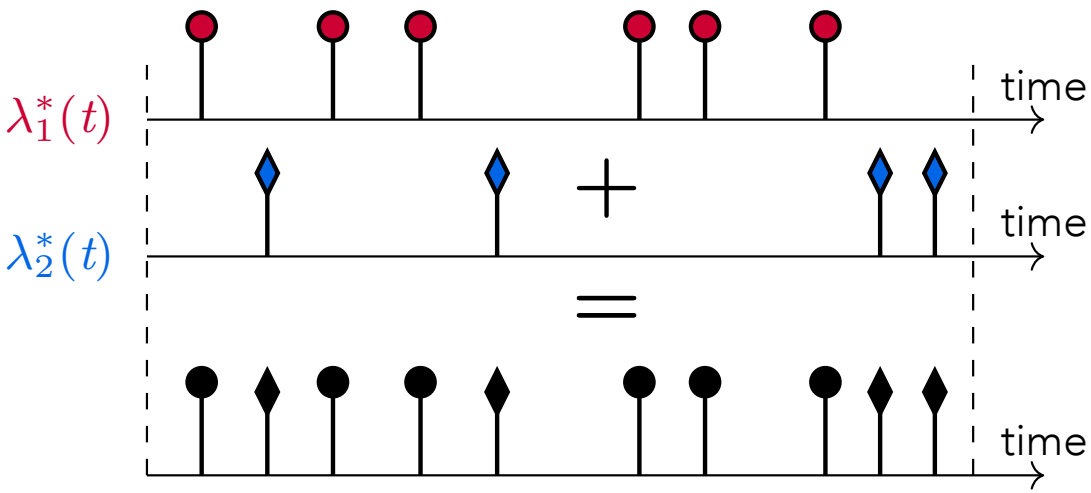
[2] In reality, if the conditional intensities $\lambda_1^*(t)$ and $\lambda_2^*(t)$ are conditioned on the joint history $\mathcal{H}(t) = \mathcal{H}_1(t) \cup \mathcal{H}_2(t)$, it is only necessary that both processes are conditionally independent given $\mathcal{H}(t)$. However, for ease of exposition, independence is assumed.

Intensity (hazard) function

Advantage of using $\lambda^*(t)$ for the reusability:

- **Model reusability**: we can easily combine several TPPs models. Given two independent[2] TPP, with histories $\mathcal{H}_1(t)$ and $\mathcal{H}_2(t)$, using intensities $\lambda_1^*(t)$ and $\lambda_2^*(t)$, respectively, then, we can characterize the joint history $\mathcal{H}(t) = \mathcal{H}_1(t) \cup \mathcal{H}_2(t)$ by means of:

$$\begin{aligned}\lambda^*(t)\,dt &= \mathbb{E}\left\{dN(t)|\mathcal{H}(t)\right\} = \mathbb{E}\left\{dN_1(t) + dN_2(t)|\mathcal{H}(t)\right\} = \\ &= \mathbb{E}\left\{dN_1(t)|\mathcal{H}_1(t)\right\} + \mathbb{E}\left\{dN_2(t)|\mathcal{H}_2(t)\right\} = \\ &= \lambda_1^*(t)\,dt + \lambda_2^*(t)\,dt\end{aligned}$$



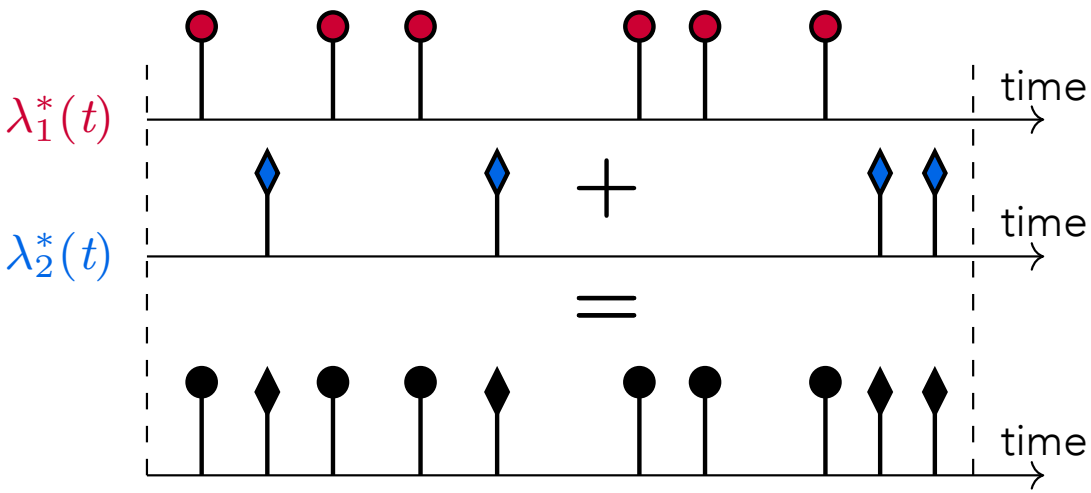
[2] In reality, if the conditional intensities $\lambda_1^*(t)$ and $\lambda_2^*(t)$ are conditioned on the joint history $\mathcal{H}(t) = \mathcal{H}_1(t) \cup \mathcal{H}_2(t)$, it is only necessary that both processes are conditionally independent given $\mathcal{H}(t)$. However, for ease of exposition, independence is assumed.

Intensity (hazard) function

Advantage of using $\lambda^*(t)$ for the reusability:

- **Model reusability**: we can easily combine several TPPs models. Given two independent[2] TPP, with histories $\mathcal{H}_1(t)$ and $\mathcal{H}_2(t)$, using intensities $\lambda_1^*(t)$ and $\lambda_2^*(t)$, respectively, then, we can characterize the joint history $\mathcal{H}(t) = \mathcal{H}_1(t) \cup \mathcal{H}_2(t)$ by means of:

$$\begin{aligned}\lambda^*(t)\,dt &= \mathbb{E}\left\{dN(t)|\mathcal{H}(t)\right\} = \mathbb{E}\left\{dN_1(t) + dN_2(t)|\mathcal{H}(t)\right\} = \\ &= \mathbb{E}\left\{dN_1(t)|\mathcal{H}_1(t)\right\} + \mathbb{E}\left\{dN_2(t)|\mathcal{H}_2(t)\right\} = \\ &= \lambda_1^*(t)\,dt + \lambda_2^*(t)\,dt\end{aligned}$$



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

[2] In reality, if the conditional intensities $\lambda_1^*(t)$ and $\lambda_2^*(t)$ are conditioned on the joint history $\mathcal{H}(t) = \mathcal{H}_1(t) \cup \mathcal{H}_2(t)$, it is only necessary that both processes are conditionally independent given $\mathcal{H}(t)$. However, for ease of exposition, independence is assumed.

$$\lambda^*(t) = \frac{f^*(t)}{S^*(t)} = -\frac{1}{S^*(t)}\frac{dS^*(t)}{dt} = -\frac{d\log S^*(t)}{dt} \rightarrow S^*(t) = \exp\left\{-\int_{t_{i-1}}^{t} \lambda^*(\tau)\,d\tau\right\}$$

$$f^*(t) = -\frac{d\left(\exp\left\{-\int_{t_{i-1}}^{t} \lambda^*(\tau)\,d\tau\right\}\right)}{dt} = \lambda^*(t) \times \exp\left\{-\int_{t_{i-1}}^{t} \lambda^*(\tau)\,d\tau\right\}$$

Likelihood for model fitting

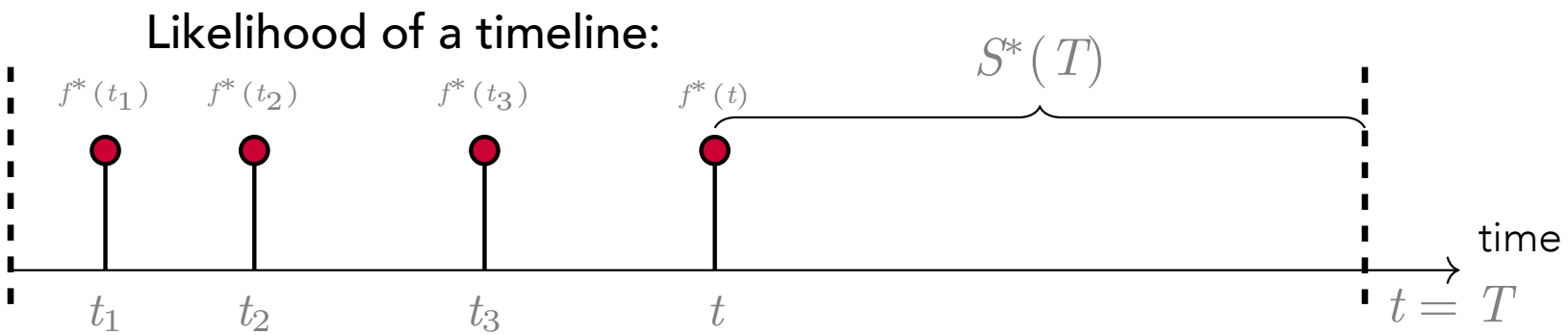
We aim to characterize the probability of an observed timeline $t \in [0, T] \longrightarrow$ used for model fitting!!!

- We assume an observation during that timeline of a TPP $\rightarrow \mathcal{H}(T) = \{t_1, t_2, \ldots, t_n\}$
- Given a statistical model —$f^*_{\boldsymbol{\theta}}(t)$ or $\lambda^*_{\boldsymbol{\theta}}(t)$— for the TPP, parametrized in $\boldsymbol{\theta}$:

Definition

The likelihood function is defined as:

$$\mathcal{L}\left(\mathcal{H}(T); \boldsymbol{\theta}\right) = \prod_{i=1}^{n} f^*_{\boldsymbol{\theta}}(t_i) \times S^*_{\boldsymbol{\theta}}(T)$$



Likelihood for model fitting

Since the Likelihood is:

- the probability of the observed data given a parameter value, ...
- .... we can interpret it as the mathematical expression that measures how well a particular model (with parameters $\boldsymbol{\theta}$) explains the observed data

Objective: infer structure in temporal distribution of events

We will use the **MLL** to find the model parameters $\boldsymbol{\theta}$ under which an the model chosen is more likely to generate $\mathcal{H}(T)$,

$$\hat{\boldsymbol{\theta}} = \operatorname{argmax} \{\log \mathcal{L}(\mathcal{H}(T); \boldsymbol{\theta})\}$$

Note: proof that

$$\mathcal{L}(\mathcal{H}(T); \boldsymbol{\theta}) = \prod_{i=1}^{n} f^*_{\boldsymbol{\theta}}(t_i) \times S^*_{\boldsymbol{\theta}}(T) = \prod_{i=1}^{n} \lambda^*_{\boldsymbol{\theta}}(t_i) \times \exp\left(-\int_0^T \lambda^*_{\boldsymbol{\theta}}(\tau)\, d\tau\right)$$

- The first term rewards the model for assigning high intensity where real events occurred. Each observed event t_i contributes $\lambda^*_{\boldsymbol{\theta}}(t_i)$.
- The second term penalizes the model for predicting too many events overall through the integral of λ over time.

So it balances "predicting events where they happened" vs. "not overpredicting events everywhere."

Basic building blocks (intensity functions)

Set of basic intensity functions, which are often the building blocks of more complex models. For each of these intensity functions, we will learn how to perform two common operations:

- Estimate the model parameters that best fits a specific history of events $\mathcal{H}(T)$, i.e., the model parameters under which the intensity function is more likely to generate that specific history of events.
- Sample new events from the intensity function.

These are:

- Homogeneous Poisson process
- Inhomogeneous Poisson Process
- Hawkes Process
- Terminating point process

Homogeneous Poisson process

The simplest temporal point process, where the intensity, or the rate of events, is given by a constant parameter μ , i.e.,

$$\lambda^*_\mu(t) = \mu \geq 0$$

Note that:

- The intensity is independent of the history $\mathcal{H}(t)$.
- The occurrence of events happens uniformly at random and the inter-event time, i.e., $t_i - t_{i-1}$ for any i, is exponentially distributed with mean $1/\mu$.

$$\lambda^*_\mu(t) = \mu \quad \longrightarrow$$

$$\begin{aligned} \mathrm{f}^*_\mu(t) &= \mu \times \exp\left(-\int_{t_{i-1}}^{t} \mu d\tau\right) \\ &= \mu \times \exp\left(-\mu(t - t_i)\right) \\ F^*_\mu(t) &= 1 - \exp\left(-\mu(t - t_i)\right) \end{aligned}$$

Homogeneous Poisson process

Why is it called "Poisson process"?

A temporal point process describes random points in time (events): If the rate of events per unit time is constant — say, always 5 events per hour — then events happen:

- independently of each other,
- uniformly over time,
- with no memory of the past.

That is exactly what a Poisson process is:
a process in which the number of events $N(t)$ — not the event times themselves — in any time interval of length dt follows a Poisson distribution:

$$N(t) \sim \text{Poisson}(\mu t).$$

That is,

$$\Pr[N(t) = k] = \frac{(\mu t)^k e^{-\mu t}}{k!}.$$

Homogeneous Poisson process

The log-likelihood (LL):

$$\log \mathcal{L}\left(\mathcal{H}(T); \mu\right) = \sum_{i=1}^{n} \log \lambda^*_{\boldsymbol{\theta}}(t_i) - \int_0^T \lambda^*_{\boldsymbol{\theta}}(\tau)\, d\tau = n \log \mu - \mu T$$

Homogeneous Poisson process

The LL:

$$\log \mathcal{L}\left(\mathcal{H}(T); \mu\right) = \sum_{i=1}^{n} \log \lambda^*_{\boldsymbol{\theta}}(t_i) - \int_0^T \lambda^*_{\boldsymbol{\theta}}(\tau)\, d\tau = n \log \mu - \mu T$$

The model satisfying the MLL:

$$\hat{\mu} = \operatorname{argmax}\left\{\log \mathcal{L}\left(\mathcal{H}(T); \mu\right)\right\} = \operatorname{argmax}\left\{n \log \mu - \mu T\right\} = \frac{n}{T}$$

Homogeneous Poisson process

The LL:

$$\log \mathcal{L}\left(\mathcal{H}(T); \mu\right) = \sum_{i=1}^{n} \log \lambda^*_{\boldsymbol{\theta}}(t_i) - \int_0^T \lambda^*_{\boldsymbol{\theta}}(\tau)\, d\tau = n \log \mu - \mu T$$

The model satisfying the MLL:

$$\hat{\mu} = \operatorname{argmax}\left\{\log \mathcal{L}\left(\mathcal{H}(T); \mu\right)\right\} = \operatorname{argmax}\left\{n \log \mu - \mu T\right\} = \frac{n}{T}$$

And to generate new sample events from $F^*_\mu(t)$:

Algorithm 1: HomogenousPoisson(μ; t_{i-1})

1: **Input:** μ (parameter), t_{i-1} (time of the last event)
2: **Output:** t (time of the next event)
3: $u \sim \text{Unif}[0, 1]$
4: $t \leftarrow -\frac{\log(1-u)}{\mu} + t_{i-1}$
5: **return** t

Fitting an Homogeneous Poisson process

Likelihood of a timeline:

$$\hat{\mu}????$$

Homogeneous Poisson process

> Epidemiol Perspect Innov. 2007 Feb 16;4:1. doi: 10.1186/1742-5573-4-1.

Applying the compound Poisson process model to the reporting of injury-related mortality rates

Scott R Kegler [1]

> Med Biol Eng Comput. 2010 Aug;48(8):799-810. doi: 10.1007/s11517-010-0638-6. Epub 2010 Jun 4.

A poisson process model for hip fracture risk

Zvi Schechner [1], Gangming Luo, Jonathan J Kaufman, Robert S Siffert

> Ann Emerg Med. 1999 Apr;33(4):409-17. doi: 10.1016/s0196-0644(99)70305-7.

Modeling the occurrence of cardiac arrest as a poisson process

E Skogvoll [1], B H Lindqvist

> Cytometry A. 2015 May;87(5):385-92. doi: 10.1002/cyto.a.22620. Epub 2015 Jan 8.

Poisson-event-based analysis of cell proliferation

Huw D Summers [1], John W Wills, M Rowan Brown, Paul Rees

Inhomogeneous Poisson process

The intensity is defined by a time-varying function $g_{\boldsymbol{\theta}}(t) \geq 0$, with model parameters $\boldsymbol{\theta}$, i.e.,

$$\lambda^*_{\boldsymbol{\theta}}(t) = g_{\boldsymbol{\theta}}(t) \quad \longrightarrow$$

$$\mathrm{f}^*_{\boldsymbol{\theta}}(t) = g_{\boldsymbol{\theta}}(t) \times \exp\left(-\int_{t_{i-1}}^{t} g_{\boldsymbol{\theta}}(t)\, d\tau\right)$$

$$F^*_{\boldsymbol{\theta}}(t) = 1 - \exp\left(-\int_{t_{i-1}}^{t} g_{\boldsymbol{\theta}}(t)\, d\tau\right)$$

- By definition, the intensity is independent of the history $\mathcal{H}(t)$.

Inhomogeneous Poisson process

The LL:

$$\log \mathcal{L}\left(\mathcal{H}(T); \boldsymbol{\theta}\right) = \sum_{i=1}^{n} \log g_{\boldsymbol{\theta}}(t_i) - \int_0^T g_{\boldsymbol{\theta}}(\tau)\, d\tau$$

Inhomogeneous Poisson process

The LL:

$$\log \mathcal{L}\left(\mathcal{H}(T); \boldsymbol{\theta}\right) = \sum_{i=1}^{n} \log g_{\boldsymbol{\theta}}(t_i) - \int_0^T g_{\boldsymbol{\theta}}(\tau)\, d\tau$$

The model satisfying the MLL:

$$\hat{\boldsymbol{\theta}} = \operatorname{argmax}\left\{\log \mathcal{L}\left(\mathcal{H}(T); \boldsymbol{\theta}\right)\right\} = \operatorname{argmax}\left\{\sum_{i=1}^{n} \log g_{\boldsymbol{\theta}}(t_i) - \int_0^T g_{\boldsymbol{\theta}}(\tau)\, d\tau\right\}$$

Inhomogeneous Poisson process

The LL:

$$\log \mathcal{L}\left(\mathcal{H}(T); \boldsymbol{\theta}\right) = \sum_{i=1}^{n} \log g_{\boldsymbol{\theta}}(t_i) - \int_0^T g_{\boldsymbol{\theta}}(\tau)\, d\tau$$

The model satisfying the MLL:

$$\hat{\boldsymbol{\theta}} = \operatorname{argmax}\left\{\log \mathcal{L}\left(\mathcal{H}(T); \boldsymbol{\theta}\right)\right\} = \operatorname{argmax}\left\{\sum_{i=1}^{n} \log g_{\boldsymbol{\theta}}(t_i) - \int_0^T g_{\boldsymbol{\theta}}(\tau)\, d\tau\right\}$$

The solution to the above optimization depend on the specific parametrization $\boldsymbol{\theta}$ (or functional form) of $g_{\boldsymbol{\theta}}(t)$. Some options (convex):

- Piece-constant intensity: $g_{\boldsymbol{\theta}}(t) = \sum_{j=1}^{n} \theta_j \mathbb{I}\left(\tau_{j-1} \leq t \leq \tau_j\right)$
- Mixture of RBF kernels: $g_{\boldsymbol{\theta}}(t) = \sum_{j=1}^{n} \theta_j \exp\left(-\beta(t-\tau_i)^2\right)$
  - where $\{\tau_i\}$ and β are given constants, which may be chosen using cross-validation

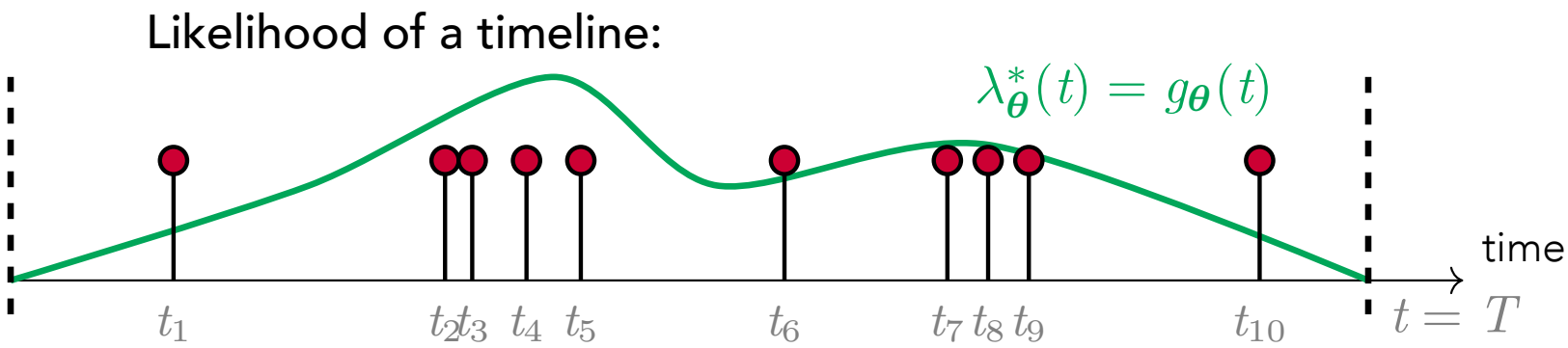
Inhomogeneous Poisson process

And to generate new sample events from $F^*_{\boldsymbol{\theta}}(t)$:

Algorithm 2: InhomoPoissonOnline($g_\theta(t)$; t_{i-1})

Input: $g_\theta(t)$ (intensity function), t_{i-1} (time of the last event)
Output: t (time of the next event)
$g_{max} \leftarrow \max_\tau g_\theta(\tau)$
$t \leftarrow t_{i-1}$
repeat
    $t \leftarrow \text{HomogenousPoisson}(g_{max}; t)$
    $u \sim \text{Unif}[0, 1]$
until $u \leq \frac{g(t)}{g_{max}}$
return t

Fitting and Inhomogeneous Poisson process with RBF



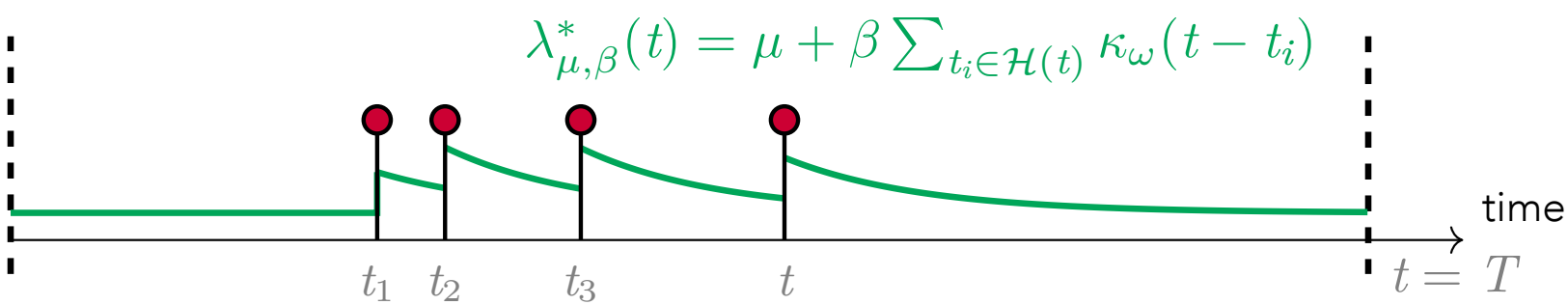
$$\hat{\boldsymbol{\theta}}????$$

Hawkes Process

Hawkes process is defined by a history dependent intensity $\lambda^*_{\mu,\beta}(t)$:

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

where $\kappa_\omega(t) = \exp(-\omega t)\mathbb{I}(t \geq 0)$[3]



- Clustered (or bursty) occurrence of events: each event temporarily increases the probability of future events nearby in time (self-exciting)
- Intensity is stochastic — $\lambda^*_{\mu,\beta}(t)$ evolves randomly, because it depends on the random past event times—and history dependent

[3] ω will be obtained by cross-validation.

Fitting Hawkes process

The LL:

$$\log \mathcal{L}\left(\mathcal{H}(T); \mu, \beta\right) = \sum_{i=1}^{n} \log \lambda^*_{\mu,\beta}(t_i) - \int_0^T \lambda^*_{\mu,\beta}(t))\, d\tau$$

Fitting Hawkes process

The LL:

$$\log \mathcal{L}\left(\mathcal{H}(T); \mu, \beta\right) = \sum_{i=1}^{n} \log \lambda^*_{\mu,\beta}(t_i) - \int_0^T \lambda^*_{\mu,\beta}(t))\, d\tau$$

The model satisfying the MLL:

$$\hat{\boldsymbol{\theta}} = \operatorname{argmax} \left\{\log \mathcal{L}\left(\mathcal{H}(T); \mu, \beta\right)\right\} = \operatorname{argmax} \left\{ \sum_{i=1}^{n} \log \lambda^*_{\mu,\beta}(t_i) - \int_0^T \lambda^*_{\mu,\beta}(t)\, d\tau \right\}$$

Fitting Hawkes process

The LL:

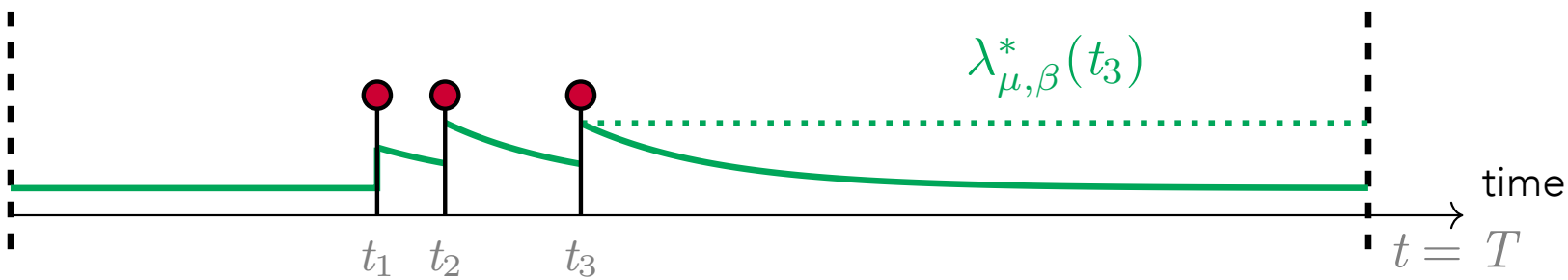
$$\log \mathcal{L}\left(\mathcal{H}(T); \mu, \beta\right) = \sum_{i=1}^{n} \log \lambda^*_{\mu,\beta}(t_i) - \int_0^T \lambda^*_{\mu,\beta}(t))\, d\tau$$

The model satisfying the MLL:

$$\hat{\boldsymbol{\theta}} = \operatorname{argmax}\left\{\log \mathcal{L}\left(\mathcal{H}(T); \mu, \beta\right)\right\} = \operatorname{argmax}\left\{\sum_{i=1}^{n} \log \lambda^*_{\mu,\beta}(t_i) - \int_0^T \lambda^*_{\mu,\beta}(t)\, d\tau\right\}$$

The MLL is jointly convex in μ and β

– Do we know $\lambda^*_{\mu,\beta}(t)\ \forall t \in [0, T]$?

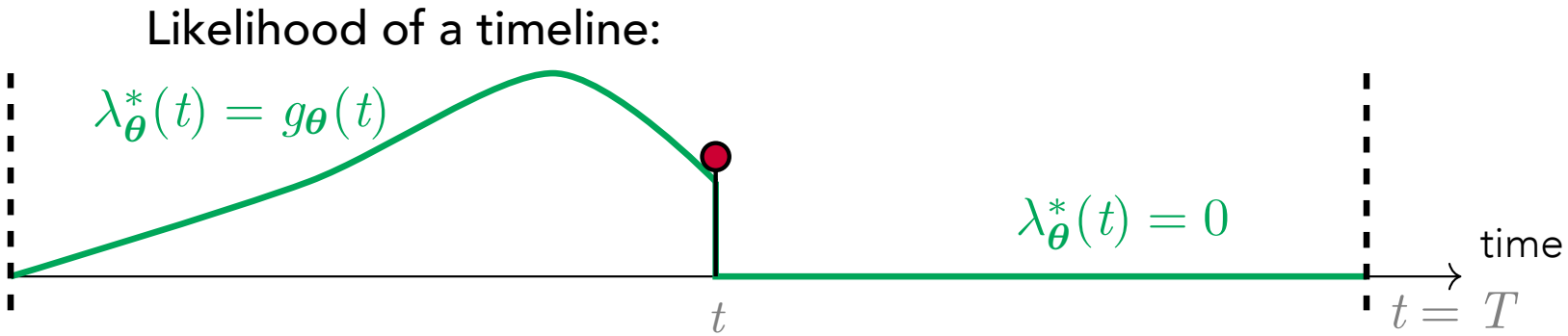


Terminating (or survival) point process

A terminating point process finishes once an event happens:

$$\lambda^*_{\boldsymbol{\theta}}(t) = g_{\boldsymbol{\theta}}(t) \times (1 - N(t))$$

where $N(t)$ is the corresponding counting process, $g_{\boldsymbol{\theta}}(t)$ is a nonnegative intensity function, with parameters $\boldsymbol{\theta}$, and the intensity $\lambda^*_{\theta}(t)$ becomes zero if an events happens.

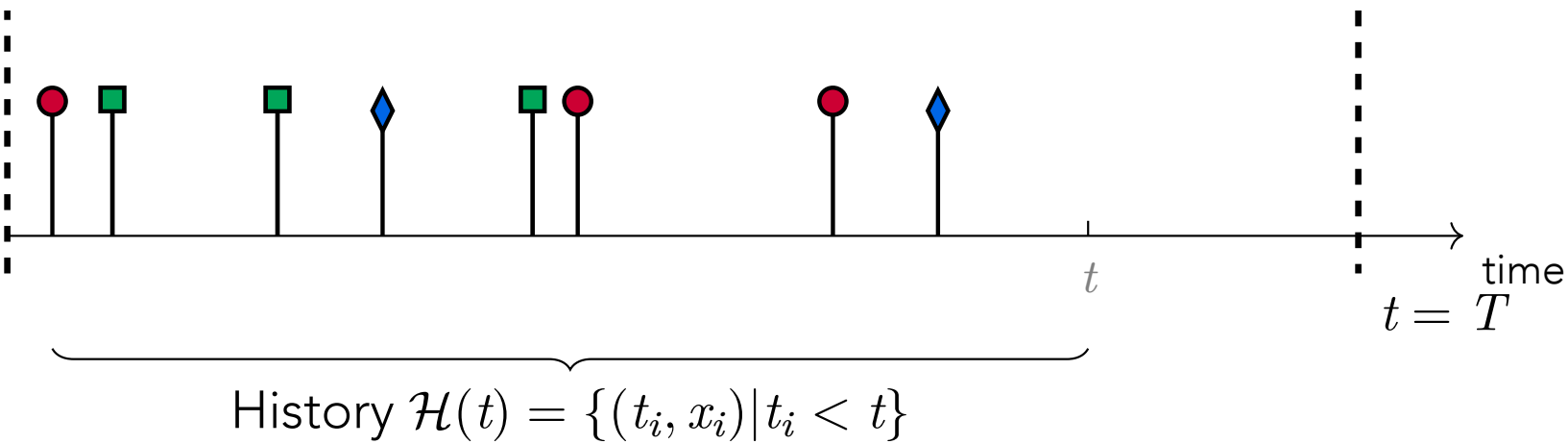


Marked Temporal Time processes

Definition

A stochastic, or random, process that models a discrete (countable) time-series of marked events that occur in continuous time $\{(t_i, x_i)\}$ with $t_i \in \mathbb{R}^+$, $x_i \in \mathcal{X}$ and $i \in \mathbb{Z}^+$.

- The domain of the mark $\mathcal{X}$ is application dependent. In the example $\mathcal{X} = \{$ ■ , ● , ♦ $\}$
- The history of all events $\mathcal{H}(t)$ up to, but not including, time t contains both times and marks, i.e., $\mathcal{H}(t) = \{(t_i, x_i) | t_i < t\}$.



Inference for spatial point patterns

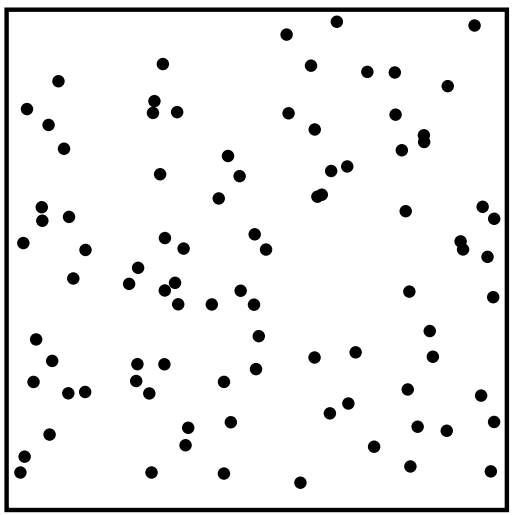
Objective is to infer structure in spatial distribution of points:

- interaction between points: regularity or clustering ('random')
- inhomogeneity linked to covariates ('systematic')
- to investigate an hypothesis (e.g. the minicolumn hypothesis)

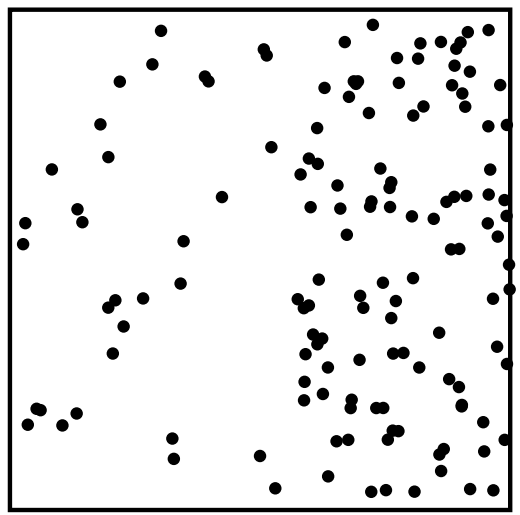
Spatial point process (SPP) *are stochastic models for spatial point patterns. Such models and various statistical tools have been developed depending on the problem and the type of spatial point pattern dataset.*

Spatial Point Process Patterns

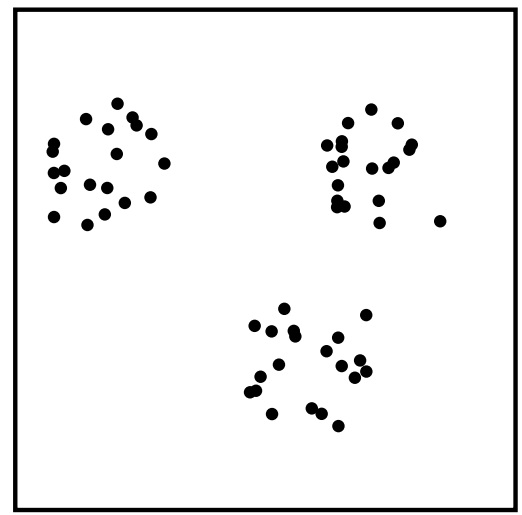
Complete Spatial Randomness



Inhomogeneous pattern



Clustered pattern



Complete spatial randomness (CSR)

CSR: *Each point occurs independently of all others, and the expected density (intensity) is constant across space.*

Intuition:

- Every region B has the same expected number of points: $\lambda|B|$.
- The number of points in disjoint regions are independent.
- The pattern looks uniformly "sprinkled".

Formal definition:

A **homogeneous Poisson process** with constant intensity $\mu > 0$ on a region $A \subset \mathbb{R}^2$:

– For any bounded region $B \subset A$, we model the number of points randomly placed in B as $N(B) \sim \text{Poisson}(\lambda|B|)$, where $|B|$ is the area:

$$\Pr(N(B) = k) = \frac{(\mu|B|)^k e^{-\mu|B|}}{k!}$$

– for disjoint regions $B_1, B_2, \ldots, B_k$, the number of points in each $N(B_1), N(B_2), \ldots, N(B_k)$, are independent.

Inhomogeneous Point Process (non-stationary Poisson)

Points are still independent, but their expected density varies across space.

Intuition:

- There are regions of higher and lower point density.
- Points are still "Poisson"—no clustering or inhibition—just varying intensity.

Formal definition:

A non-homogeneous Poisson process on A with intensity function $\mu(x, y) \geq 0, (x, y) \in A$, is a random process such that for any $B \subset A$

$$N(B) \sim \text{Poisson}(\Lambda(B)) \qquad \text{where } \Lambda(B) = \int_B \mu(u, v)\, du dv$$

and for disjopint regions $B_1, B_2, \ldots, B_k$, the number of points in each $N(B_1), N(B_2), \ldots, N(B_k)$, are independent.

Example

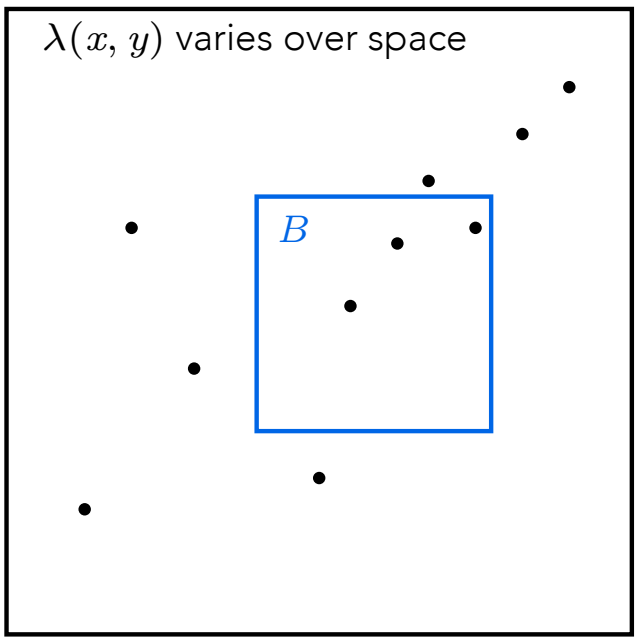
If you assume disease risk decreases with distance d from an incinerator:

$$\mu(x, y) = \lambda_0 \, e^{-\alpha \, d(x,y)},$$

then nearby regions have higher expected counts — an inhomogeneous Poisson model of the Chorley–Ribble data.

Poisson processes: space vs. time (analogy)

Spatial Poisson process

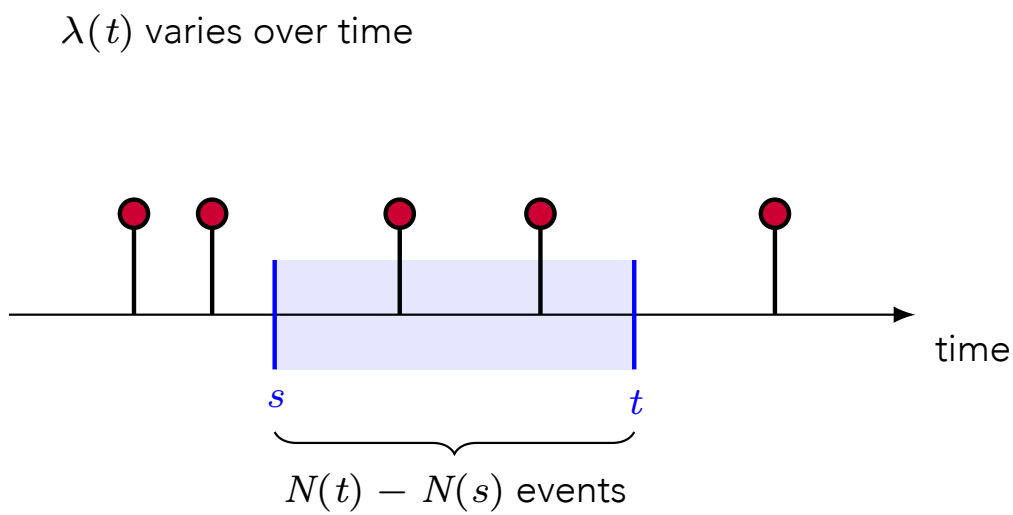


$A \subset \mathbb{R}^2$

$$N(B) \sim \text{Poisson}\Big(\int_B \lambda(x, y)\, dx\, dy \Big)$$

$$E[N(B)] = \int_B \lambda(x, y)\, dx\, dy$$

Temporal Poisson process



$$N(t) - N(s) \sim \text{Poisson}\Big(\int_s^t \lambda(u)\, du \Big)$$

$$E[N(t) - N(s)] = \int_s^t \lambda(u)\, du$$

Bibliography I

[1] P. J. Diggle, "A Point Process Modelling Approach to Raised Incidence of a Rare Phenomenon in the Vicinity of a Prespecified Point," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 153, no. 3, pp. 349–362, 1990.

[2] P. Elliott, M. Hills, J. Beresford, I. Kleinschmidt, D. Jolley, S. Pattenden, L. Rodrigues, A. Westlake, and G. Rose, "Incidence of cancers of the larynx and lung near incinerators of waste solvents and oils in Great Britain," *The Lancet*, vol. 339, no. 8797, pp. 854–858, 1992.

[3] P. J. Diggle and B. S. Rowlingson, "A Conditional Approach to Point Process Modelling of Elevated Risk," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 157, no. 3, pp. 433–440, 1994.

[4] Z. Zhou, X. Yang, R. Rossi, H. Zhao, and R. Yu, "Neural Point Process for Learning Spatiotemporal Event Dynamics," *arXiv*, 2021.

[5] Z. Dong, S. Zhu, Y. Xie, J. Mateu, and F. J. Rodríguez-Cortés, "Non-stationary spatio-temporal point process modeling for high-resolution covid-19 data," *Journal of the Royal Statistical Society Series C: Applied Statistics*, vol. 72, no. 2, p. 368–386, 2023.

[6] D. J. Daley and D. Vere-Jones, "An introduction to the theory of point processes, volume i: Elementary theory and methods," *Probability and its Applications*, 2003.