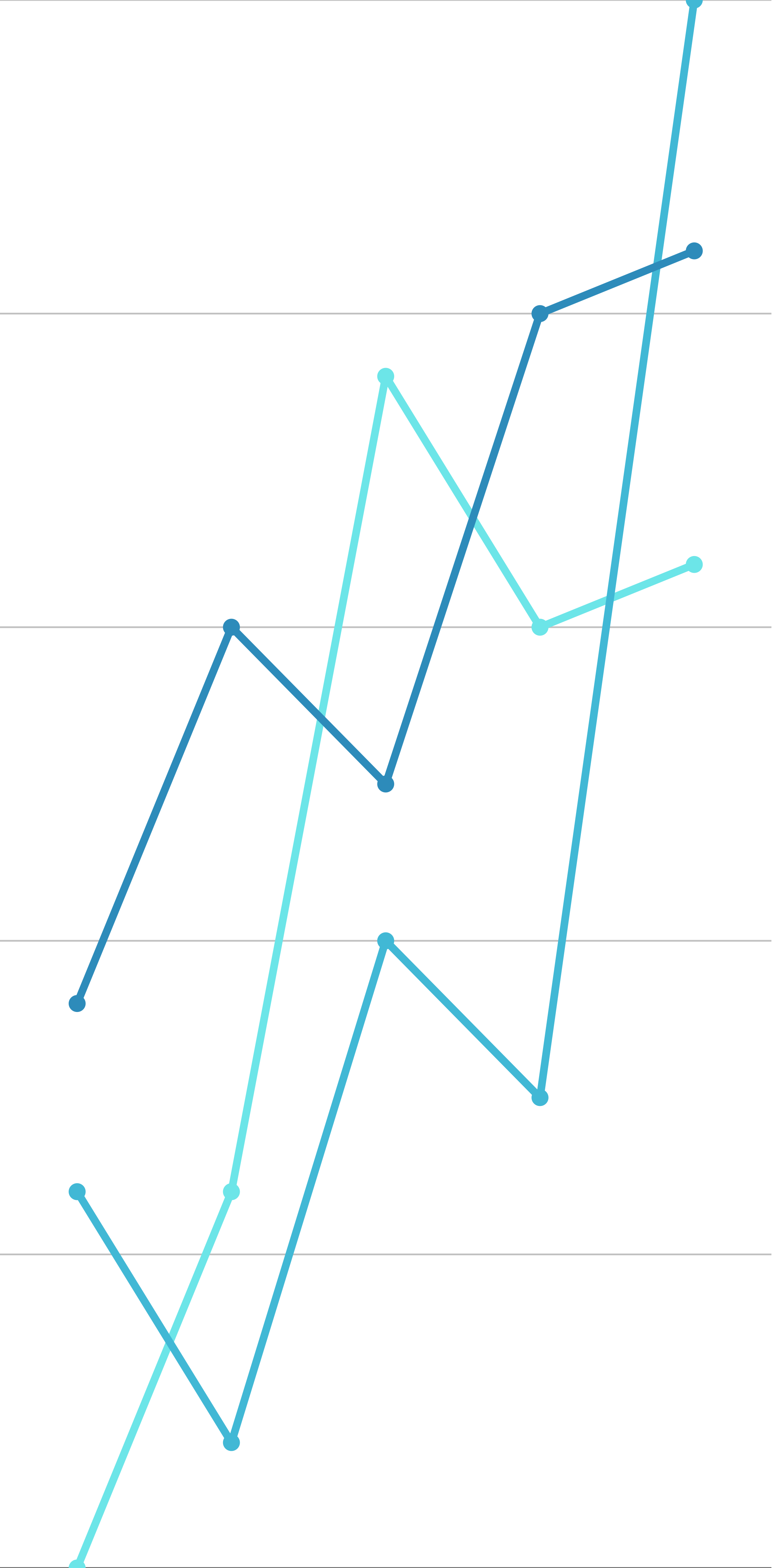


Временные точечные процессы

Владислав Жужель
Skoltech

Skolkovo Institute of Science and Technology

План семинара

- Последовательности событий
- Маркированные последовательности событий
- Классические модели
- Нейросетевые модели

Часть 1

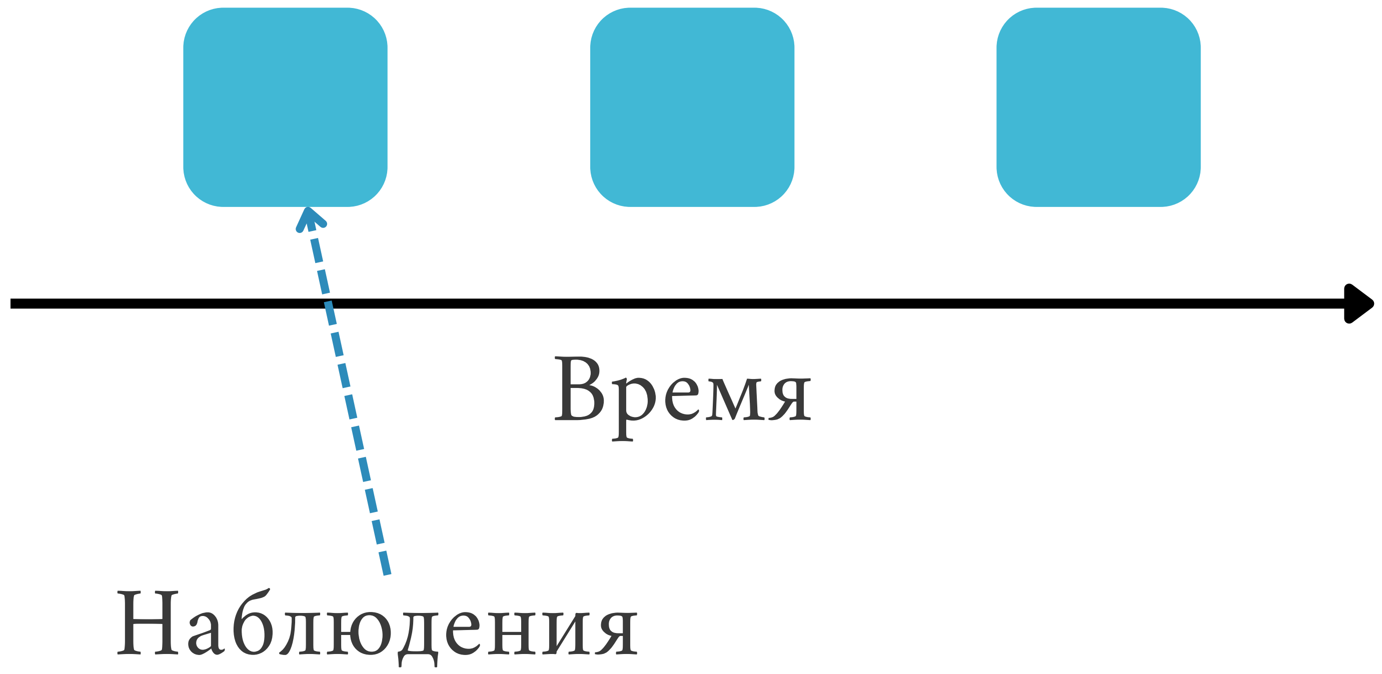
- Последовательности событий
- Маркированные последовательности событий
- Классические модели
- Нейросетевые модели

Последовательность событий

- Давайте обобщим классические временные ряды

Последовательность событий

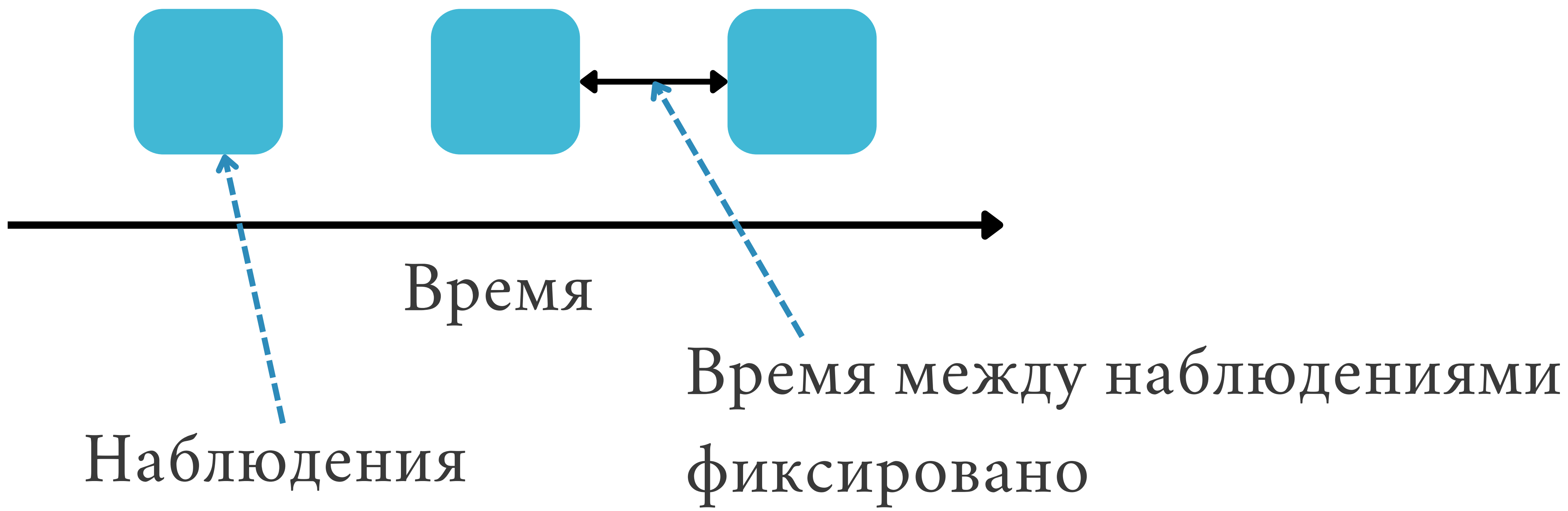
- Давайте обобщим классические временные ряды



  - Курс валют
  - Средняя дневная температура

Последовательность событий

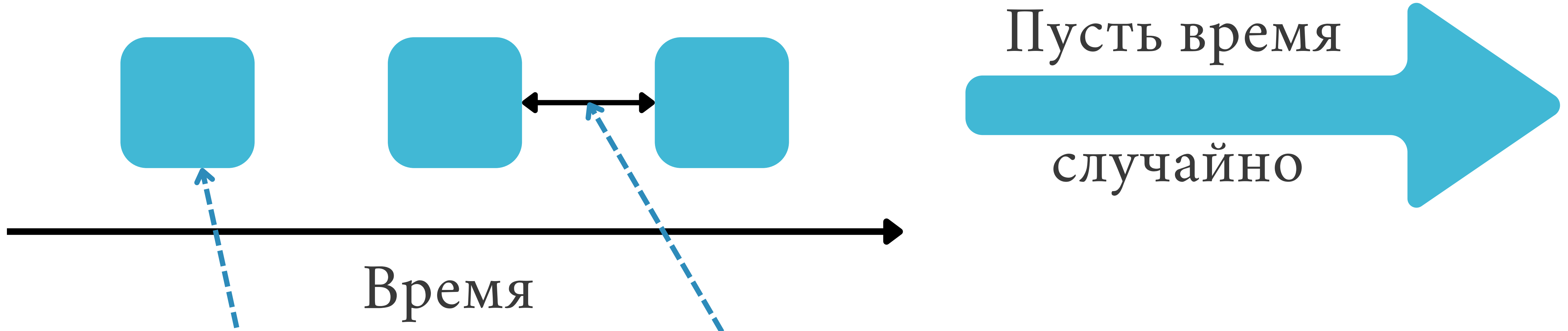
- Давайте обобщим классические временные ряды



- Курс валют
- Средняя дневная температура

Последовательность событий

- Давайте обобщим классические временные ряды



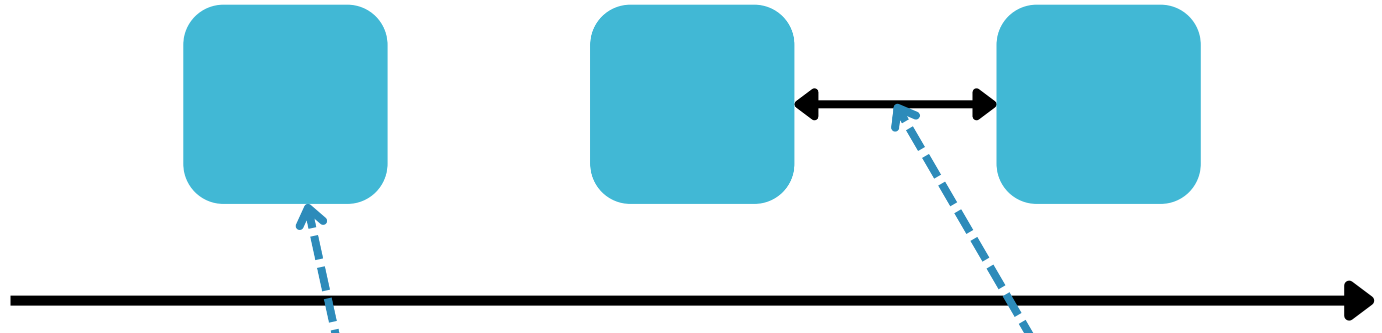
Время между наблюдениями фиксировано

Наблюдения

- Курс валют
- Средняя дневная температура

Последовательность событий

- Давайте обобщим классические временные ряды



Время

Наблюдения

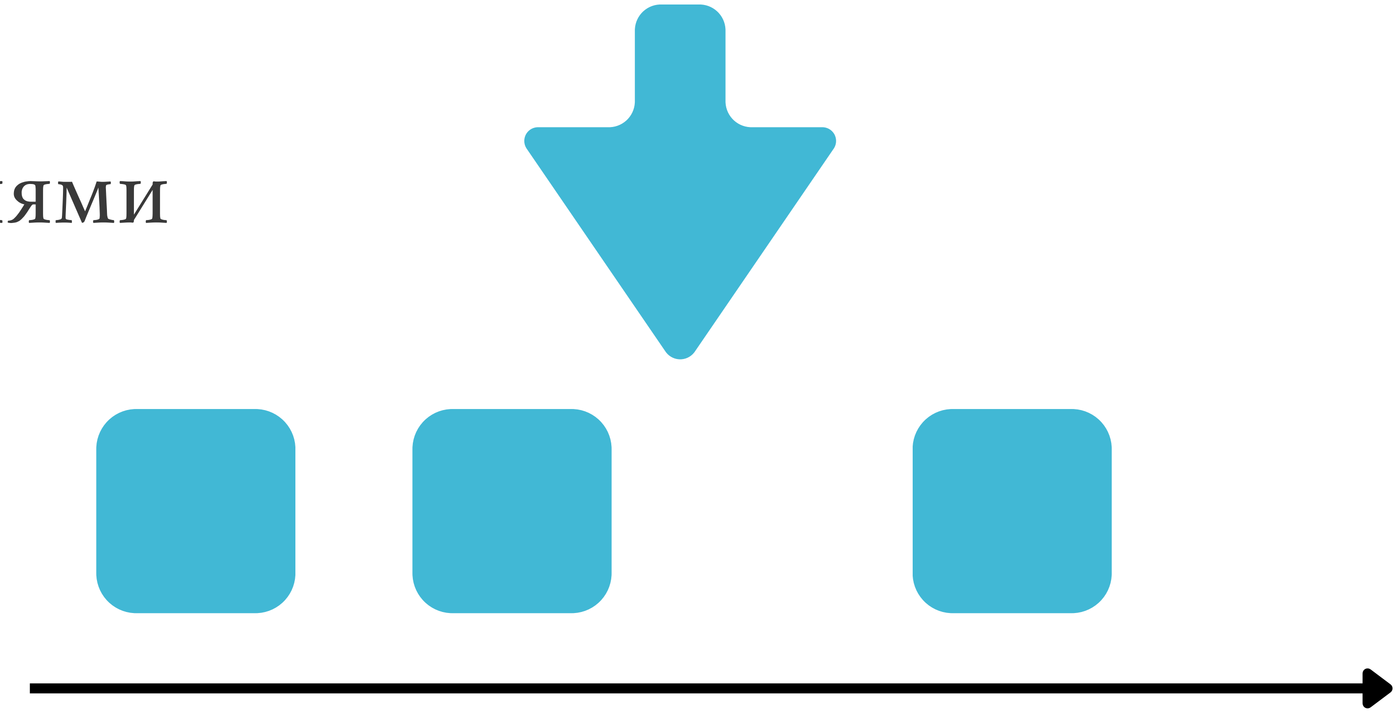
- Курс валют
- Средняя дневная температура

Время между наблюдениями фиксировано

Пусть время случайно

Время - фича

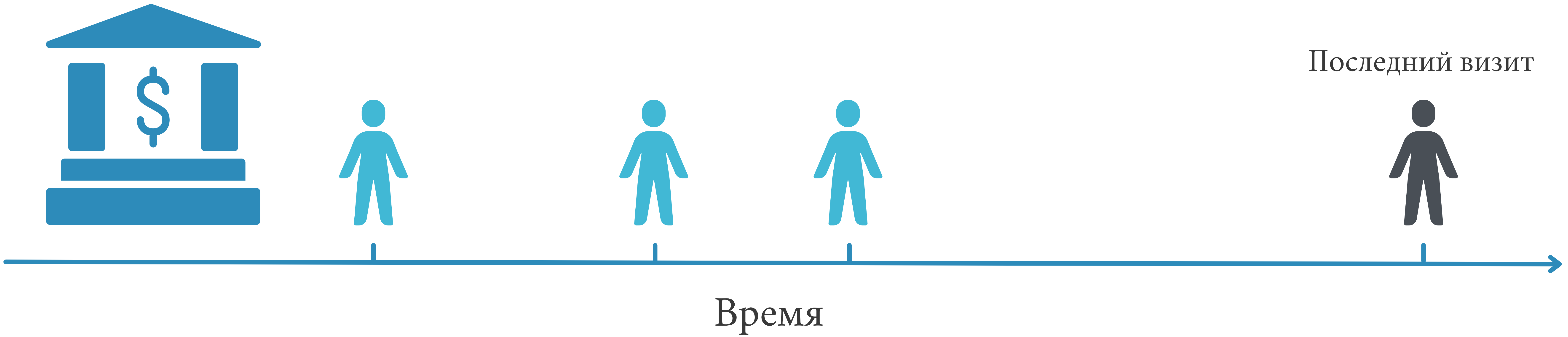
Классические подходы не работают



Время

Последовательность событий

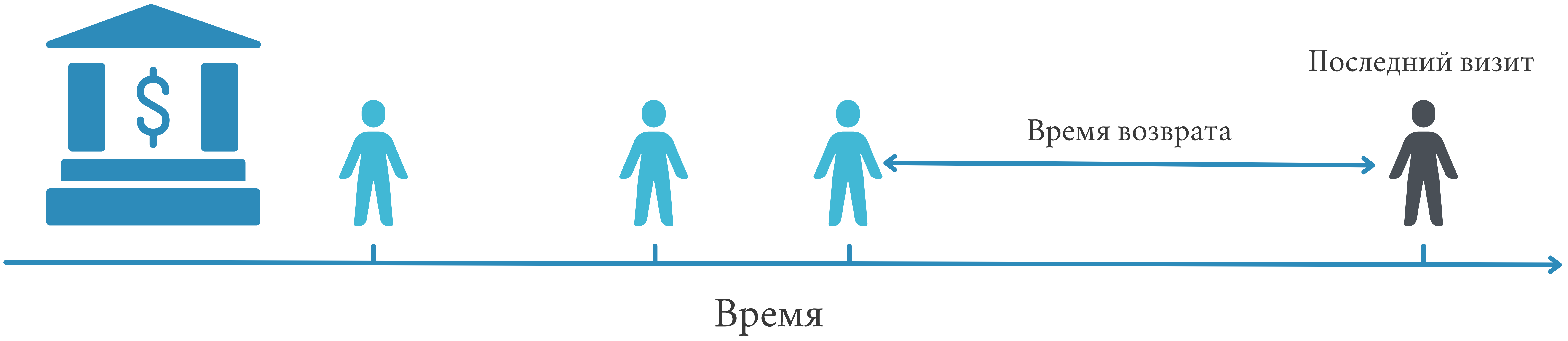
- Дискретные события в непрерывном времени
- Данные могут быть разреженными
- Неравные временные расстояния
- Примеры:
  - Посещения банка
  - Действия на сайте



Время возврата

- Когда пользователь в следующий раз вернется в банк?
- Мы можем еще предсказывать тип события
- Время возврата - это время между событий

$$\Delta t_{i+1} = t_{i+1} - t_i$$

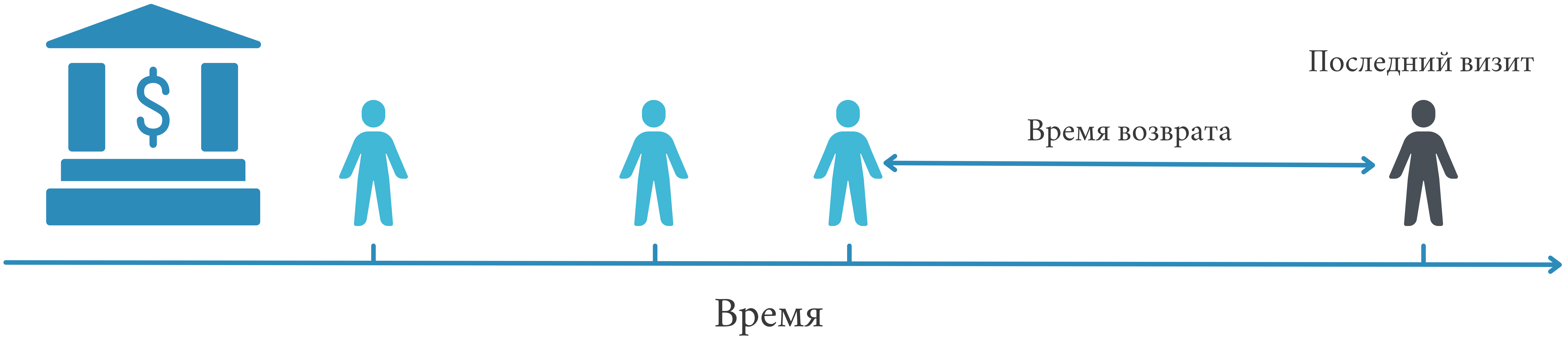


Время возврата

- Когда пользователь в следующий раз вернется в банк?
- Мы можем еще предсказывать тип события
- Время возврата - это время между событий

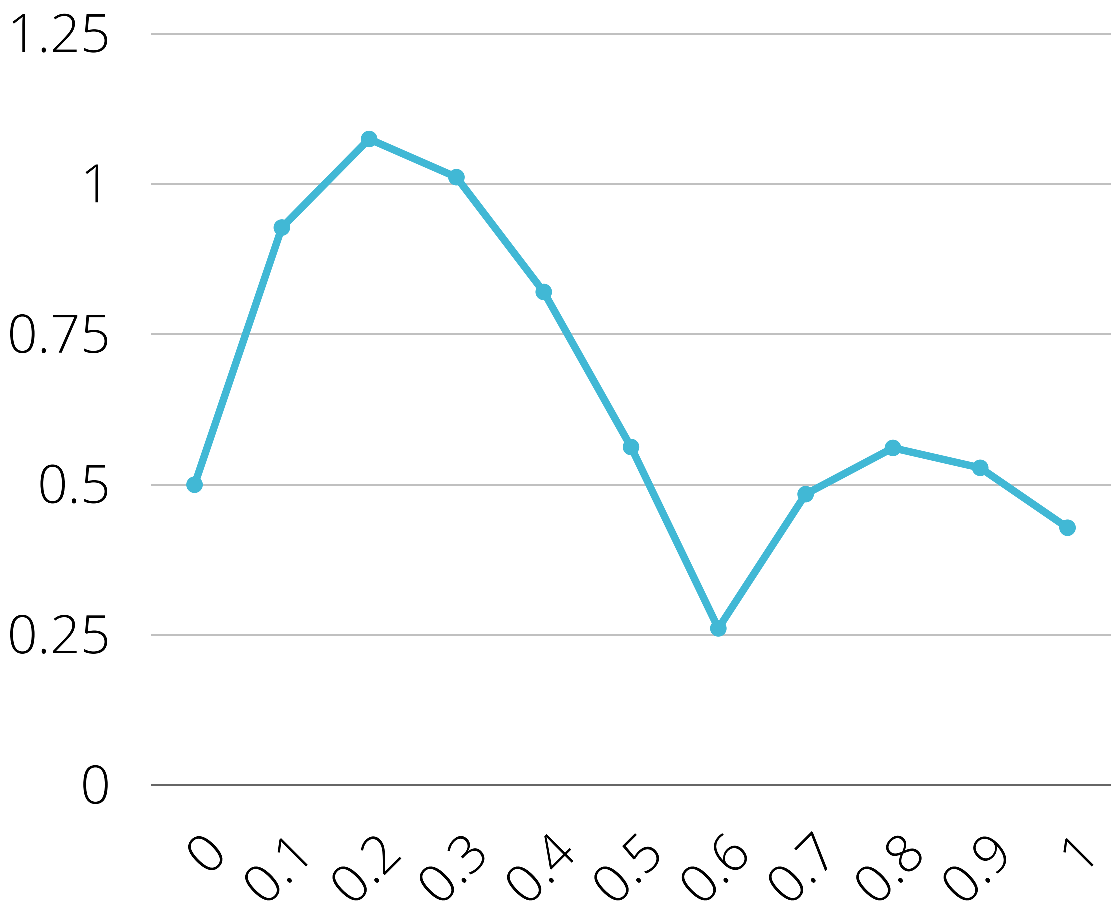
$$\Delta t_{i+1} = t_{i+1} - t_i$$

Подробнее в части 2



Время возврата как случайная величина

- Пусть T случайная величина, отвечающая за время возврата
- Случайные величины могут быть описаны с помощью плотности вероятности



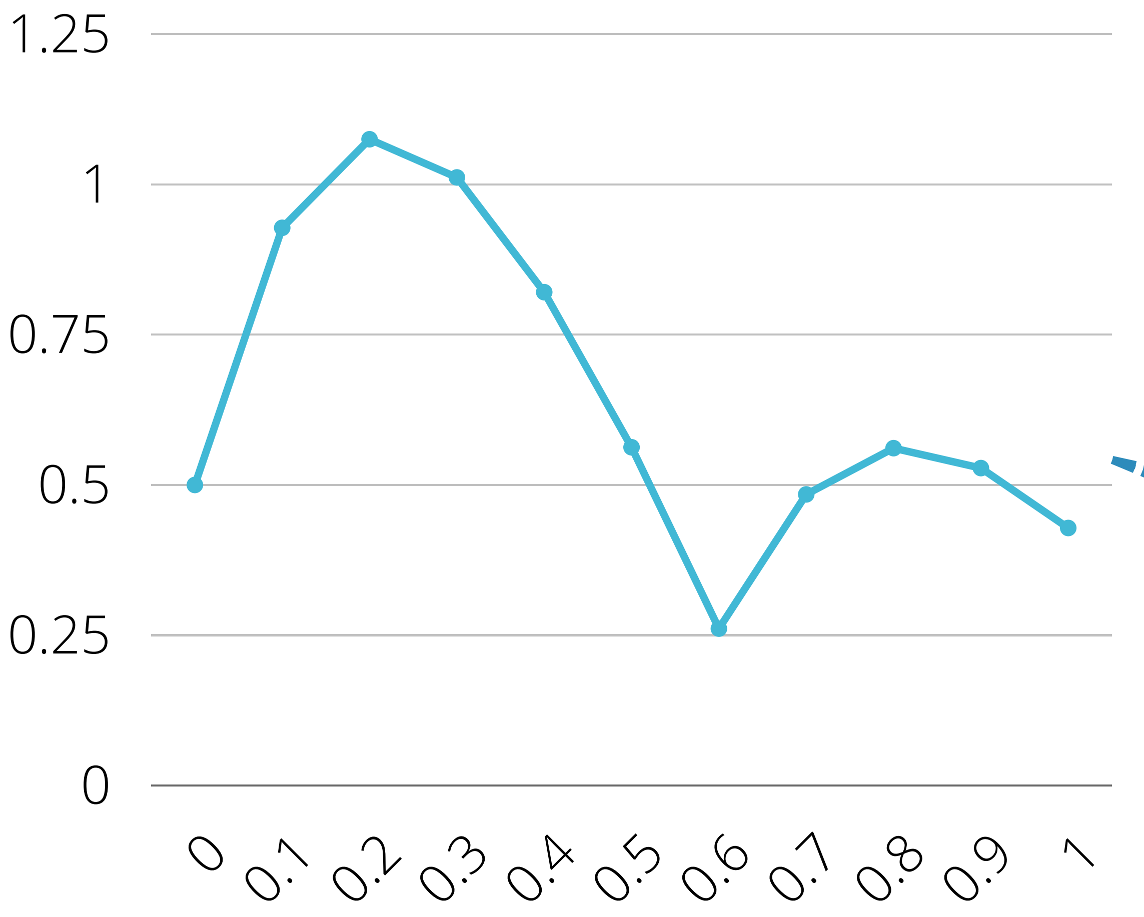
PDF

Плотность вероятности характеризует вероятность события в бесконечно малом интервале

$$f(t) = \lim_{\Delta t \to 0} \frac{P(t < T \leq t + \Delta t)}{\Delta t}$$

Время возврата как случайная величина

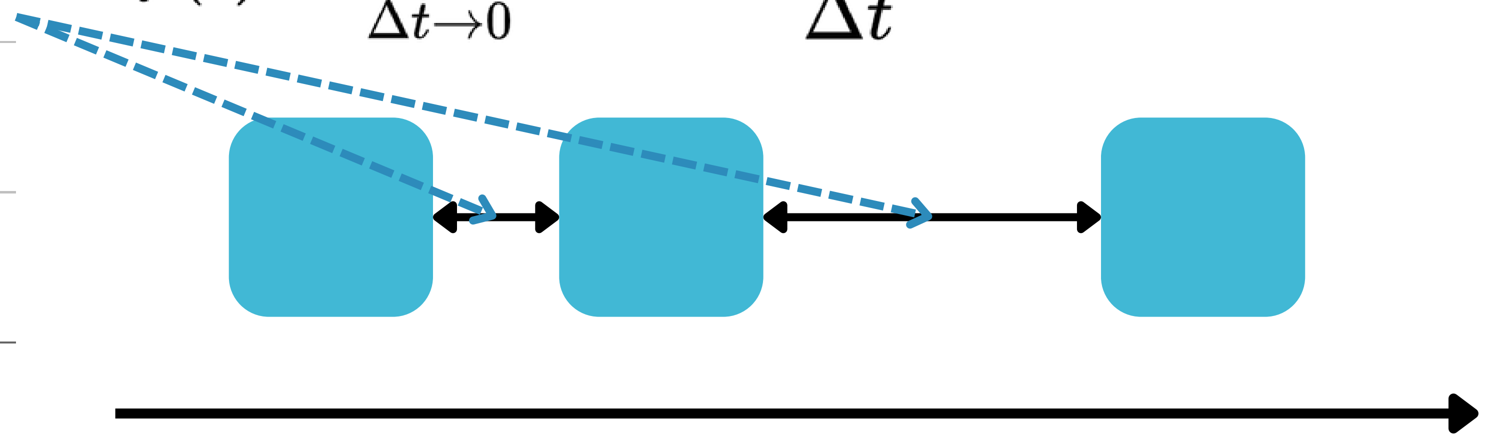
- Пусть T случайная величина, отвечающая за время возврата
- Случайные величины могут быть описаны с помощью плотности вероятности



PDF

Плотность вероятности характеризует вероятность события в бесконечно малом интервале

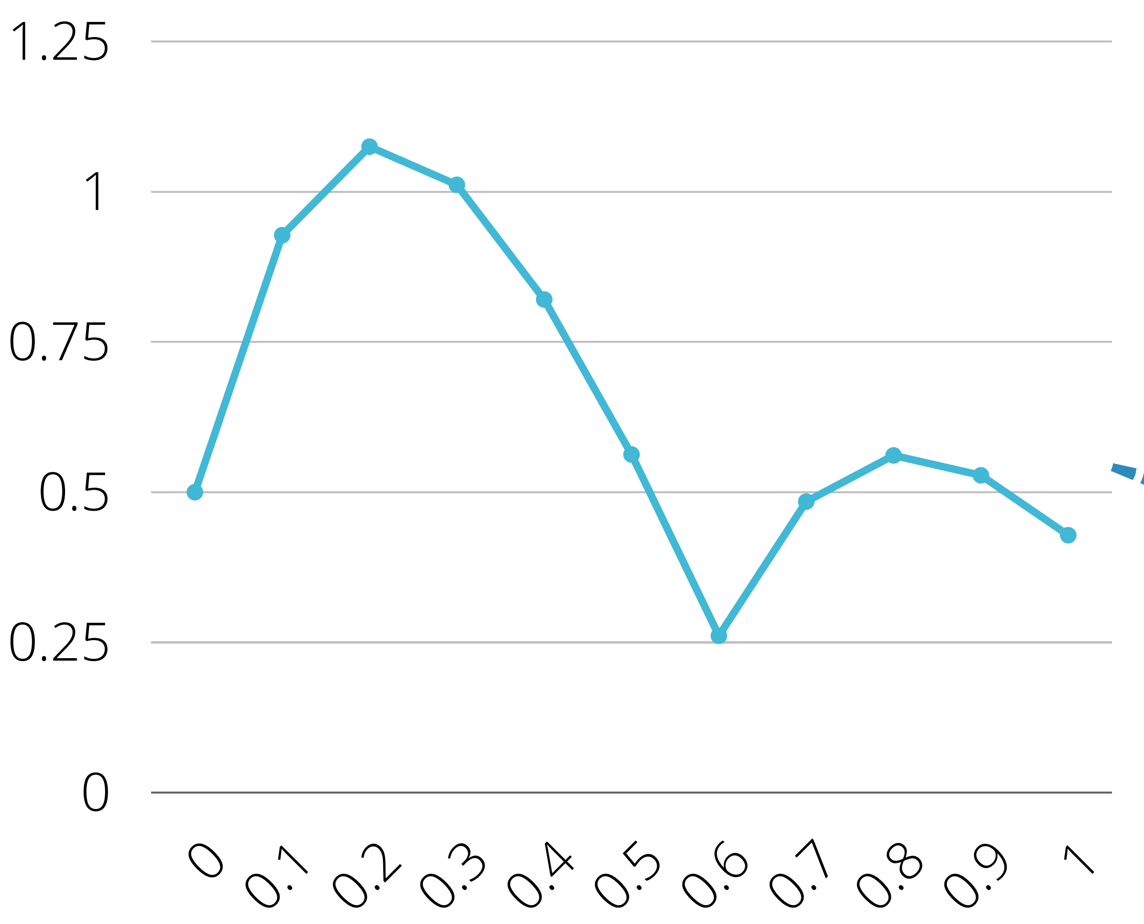
$$f(t) = \lim_{\Delta t \to 0} \frac{P(t < T \leq t + \Delta t)}{\Delta t}$$



Время

Время возврата как случайная величина

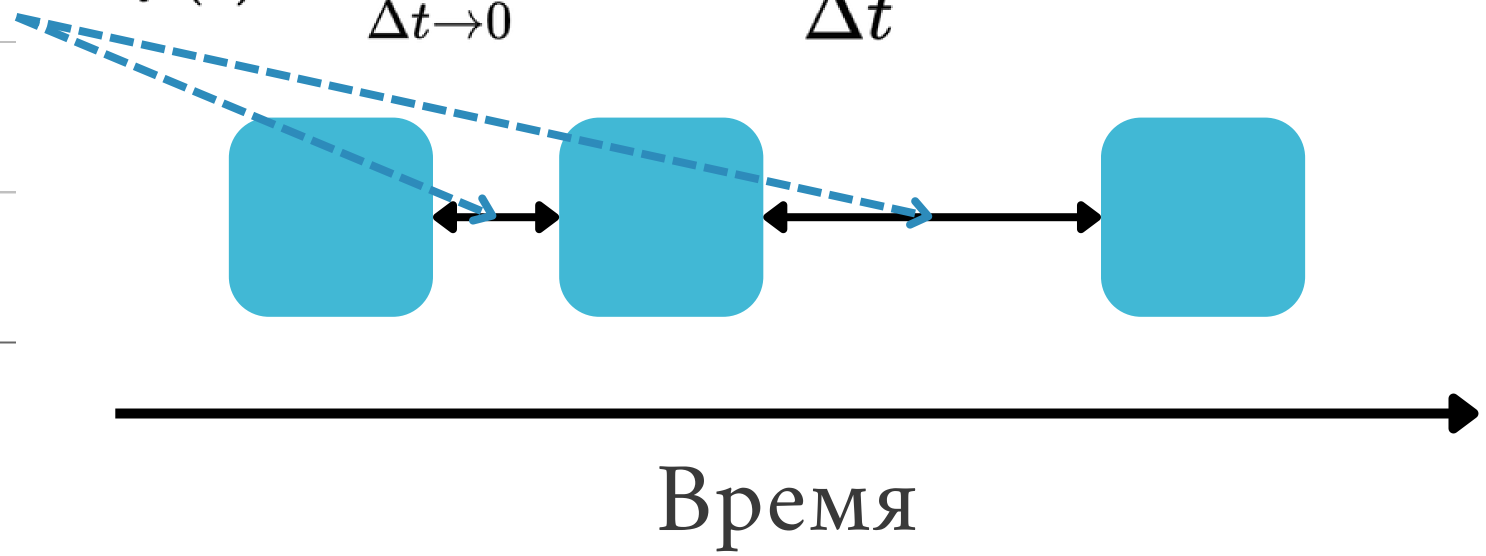
- Пусть T случайная величина, отвечающая за время возврата
- Случайные величины могут быть описаны с помощью плотности вероятности



PDF

Плотность вероятности характеризует вероятность события в бесконечно малом интервале

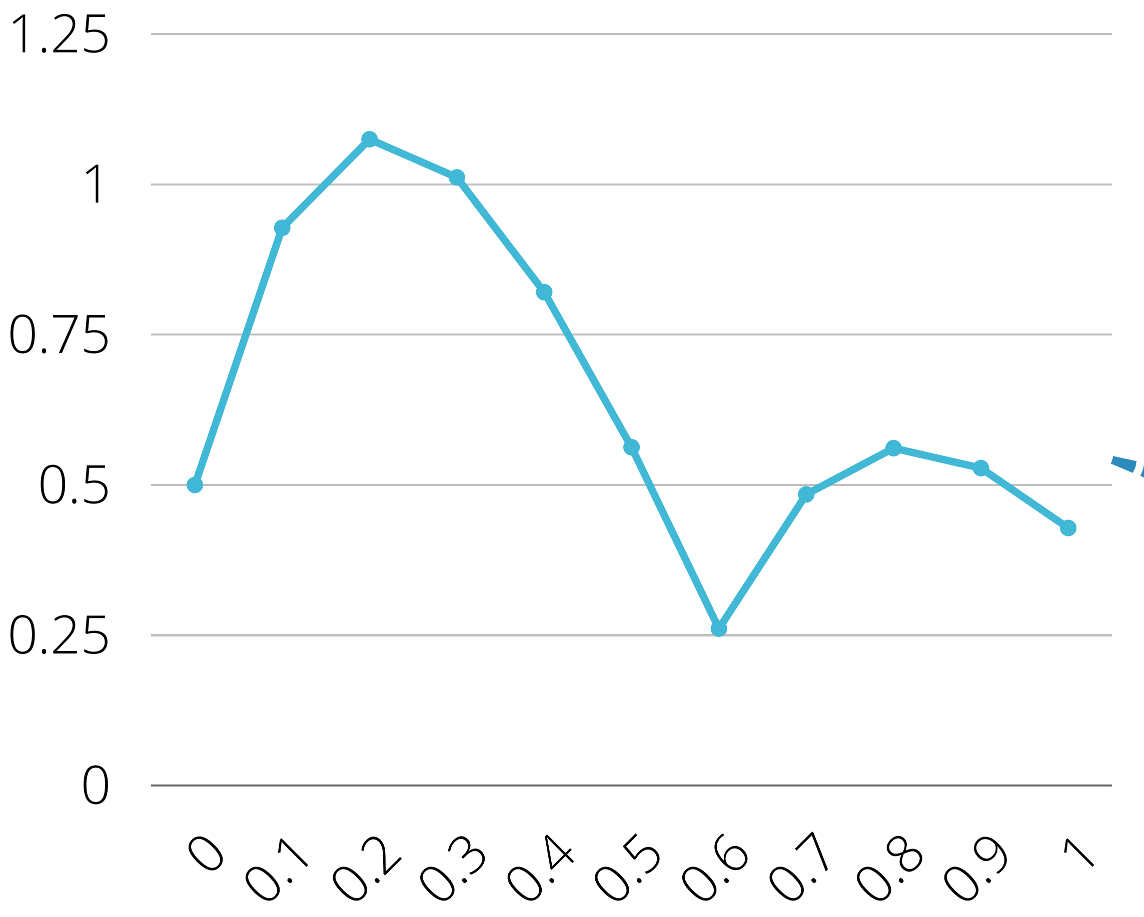
$$f(t) = \lim_{\Delta t \to 0} \frac{P(t < T \leq t + \Delta t)}{\Delta t}$$



$f(\mathbf{s}) = f(t_1) f(t_2|t_1) f(t_3|t_2, t_1) \ldots$ по цепному правилу

Время возврата как случайная величина

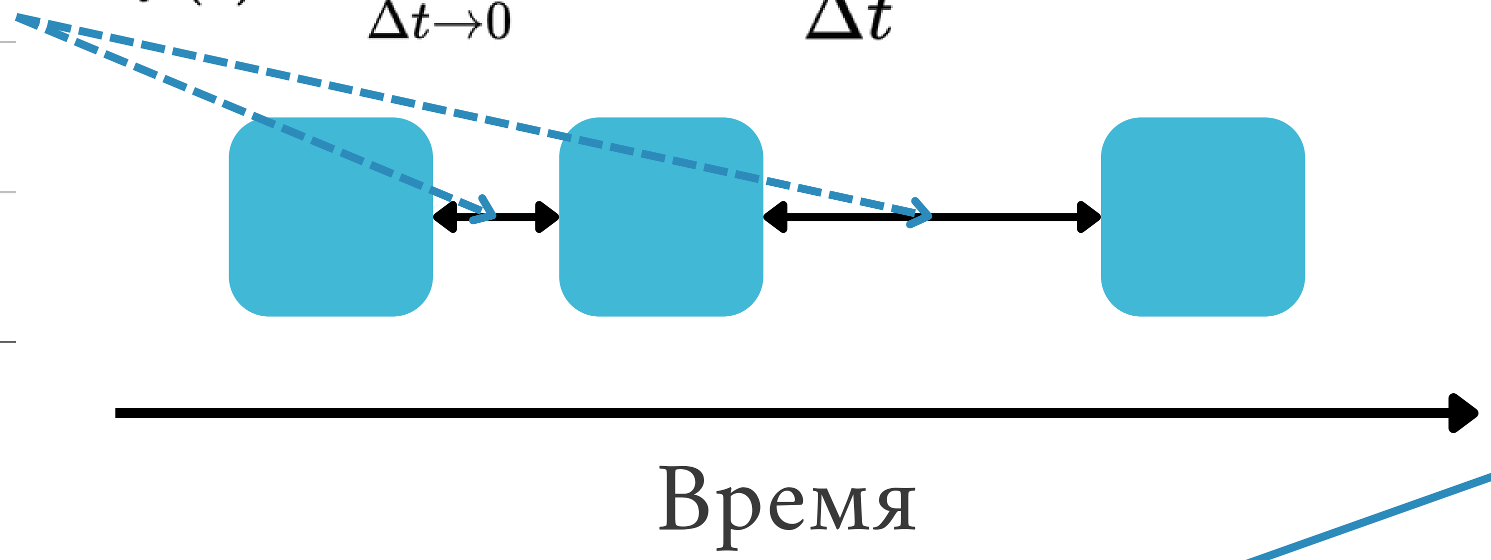
- Пусть T случайная величина, отвечающая за время возврата
- Случайные величины могут быть описаны с помощью плотности вероятности



PDF

Плотность вероятности характеризует вероятность события в бесконечно малом интервале

$$f(t) = \lim_{\Delta t \to 0} \frac{P(t < T \leq t + \Delta t)}{\Delta t}$$

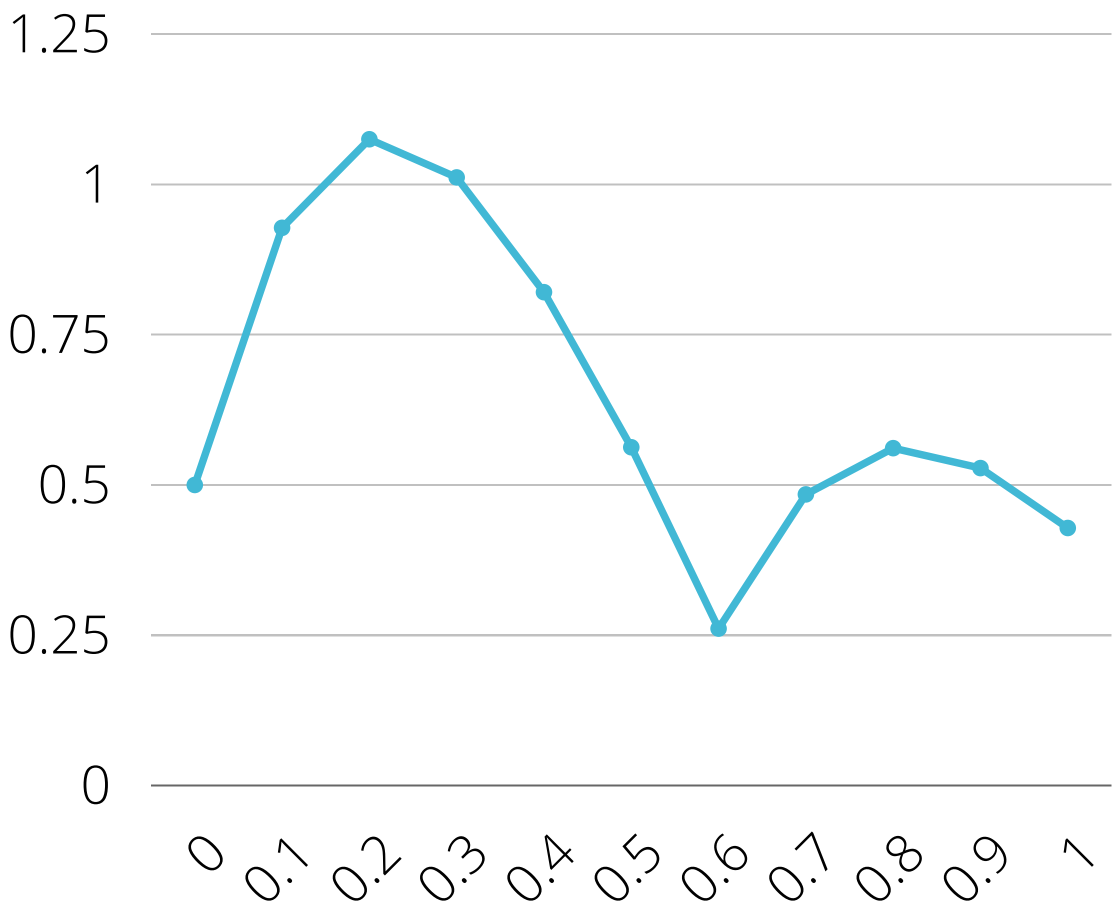


Трудно моделировать

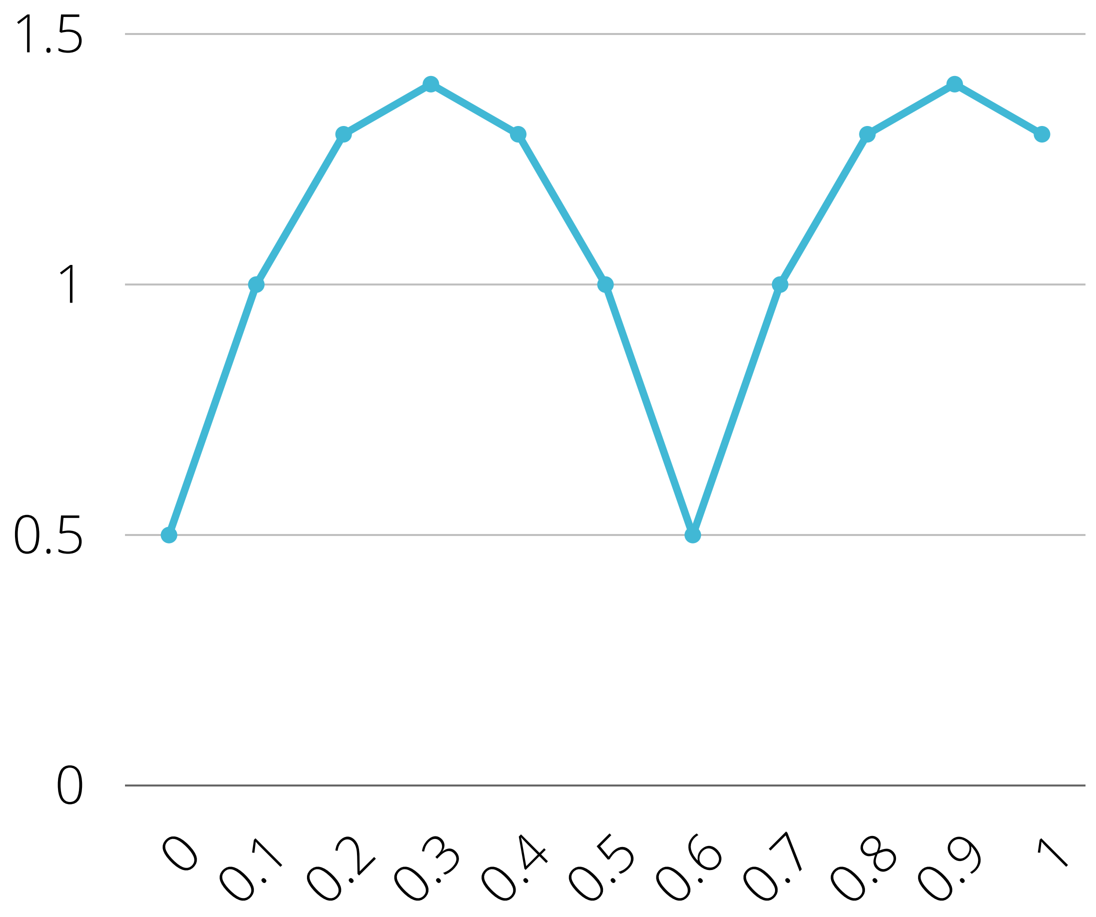
$f(\mathbf{s}) = f(t_1) f(t_2|t_1) f(t_3|t_2, t_1) \ldots$ по цепному правилу

Время возврата как случайная величина

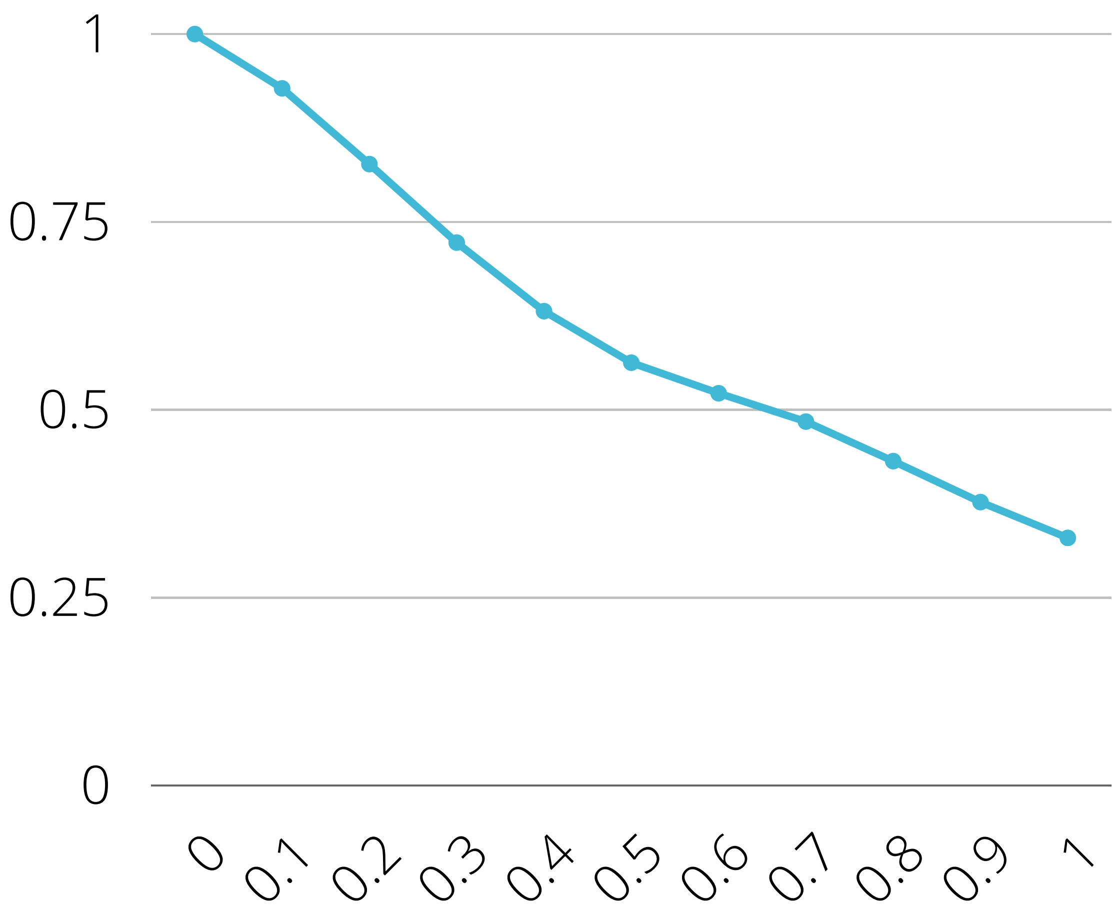
- Для последовательностей событий полезными становятся также другие функции, характеризующие распределение



PDF



Intensity Function



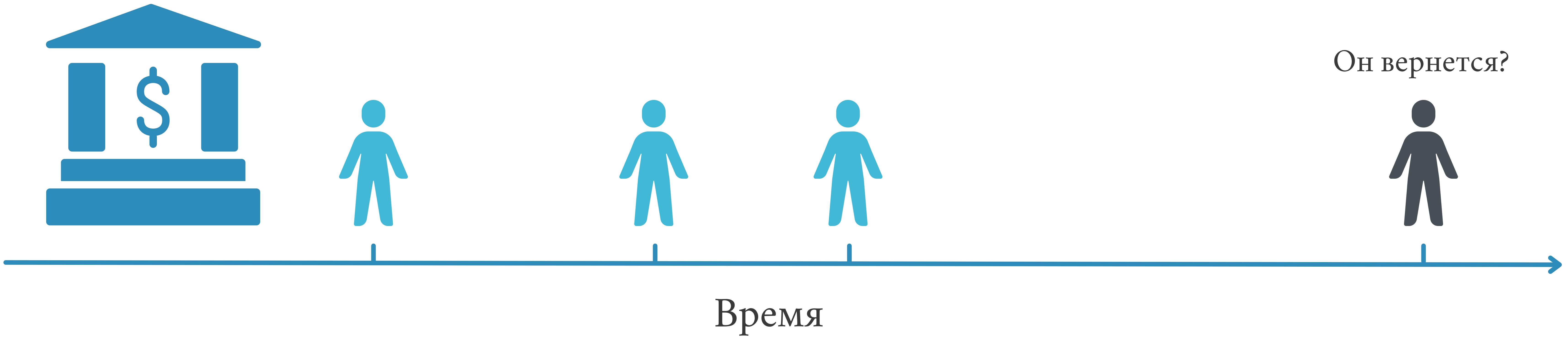
Survival Function

Функция Выживания

- Показывает вероятность того, что событие не произошло до момента времени t

$$S(t) = P(T \geq t) = \int_t^{+\infty} f(z)dz$$

- Данная функция полезна в анализе выживания
- Функция выживания равна 1 в начальный момент времени и монотонно убывает



Функция интенсивности

- Показывает частоту события
- Пусть $N(t) = \#\{t_i | t_i < t\}$
- Функция интенсивности

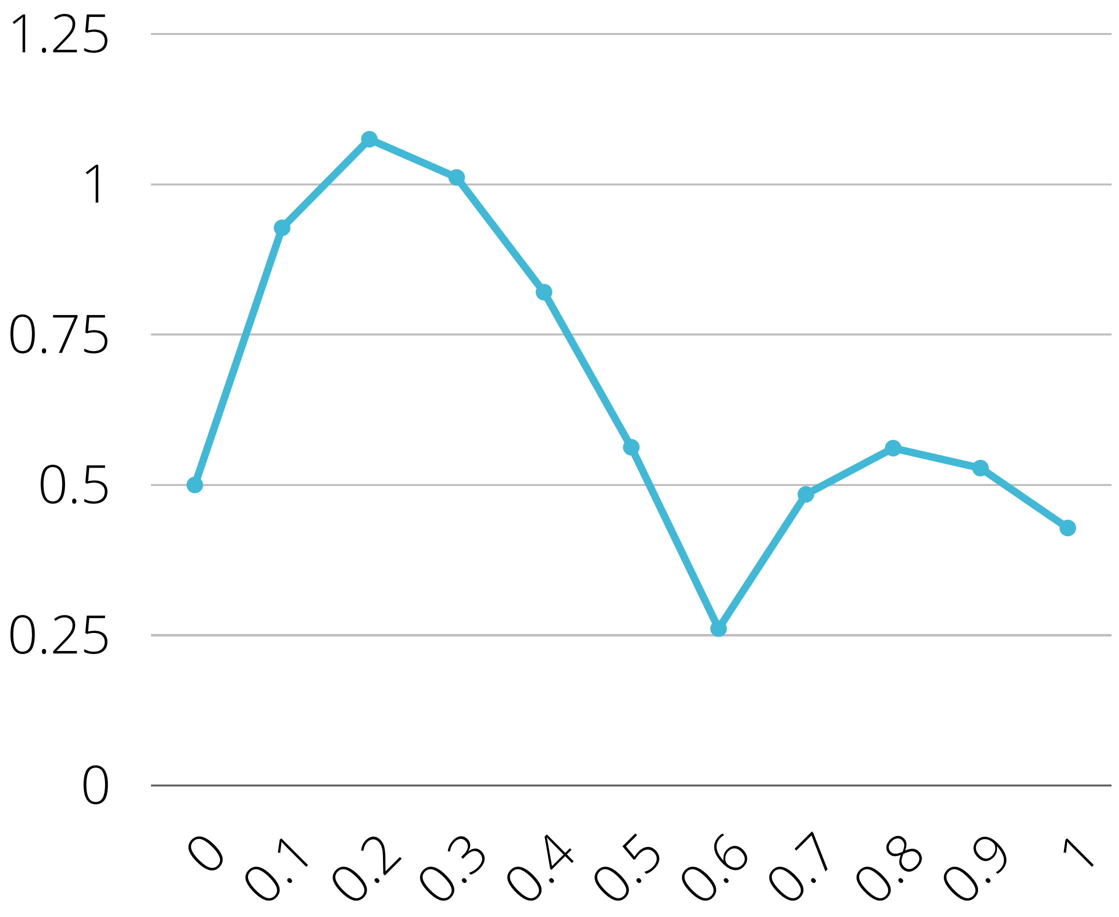
$$\lambda(t) = \mathbb{E}\frac{dN(t)}{dt} = \frac{f(t)}{S(t)} = \lim_{\Delta t \to 0} \frac{P(t < t \leq t + dt | T > t)}{\Delta t}$$

- Преимущества интенсивности:
  - Требуется только неотрицательность
  - Для двух независимых последовательностей, совместная интенсивность - это сумма
- Связь с плотностью вероятности:

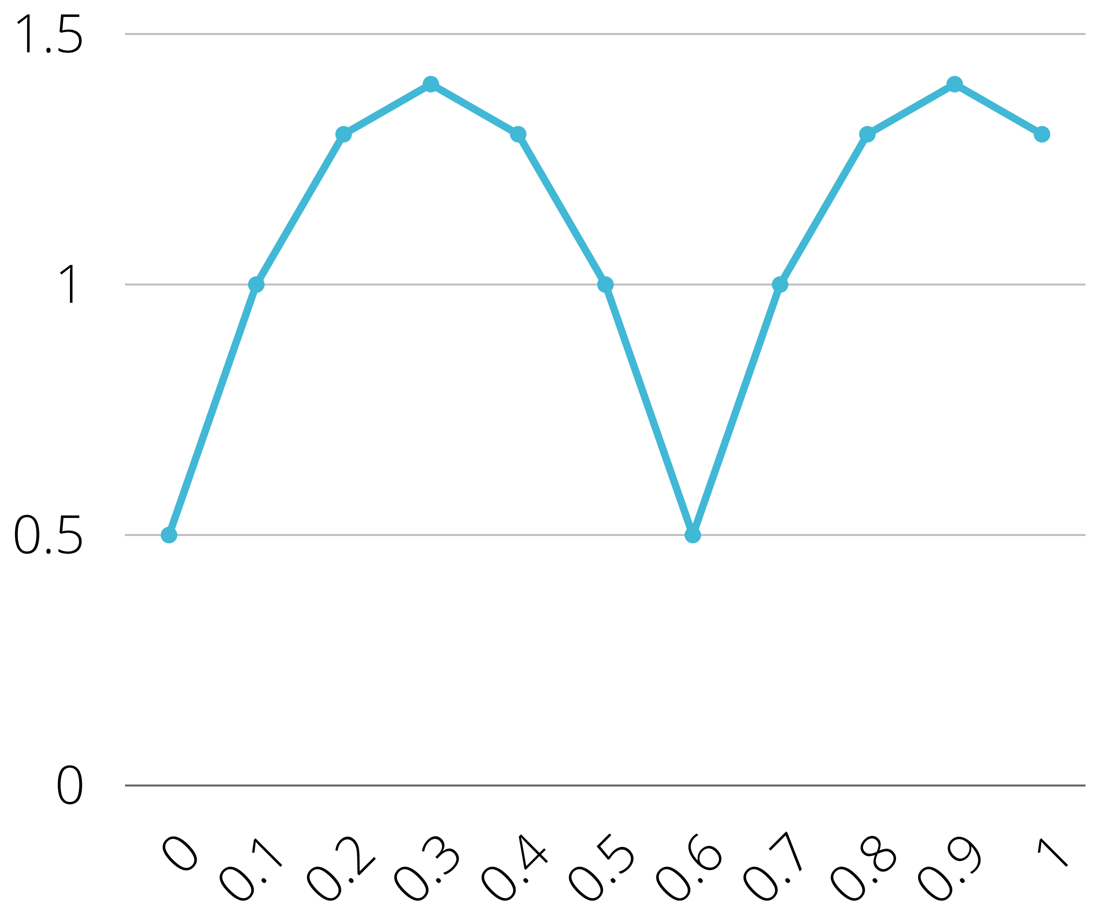
$$f(\mathbf{s}) = \prod_i \lambda(t_i) \exp\left(- \int_0^T \lambda(t) dt \right)$$

Временной точечный процесс

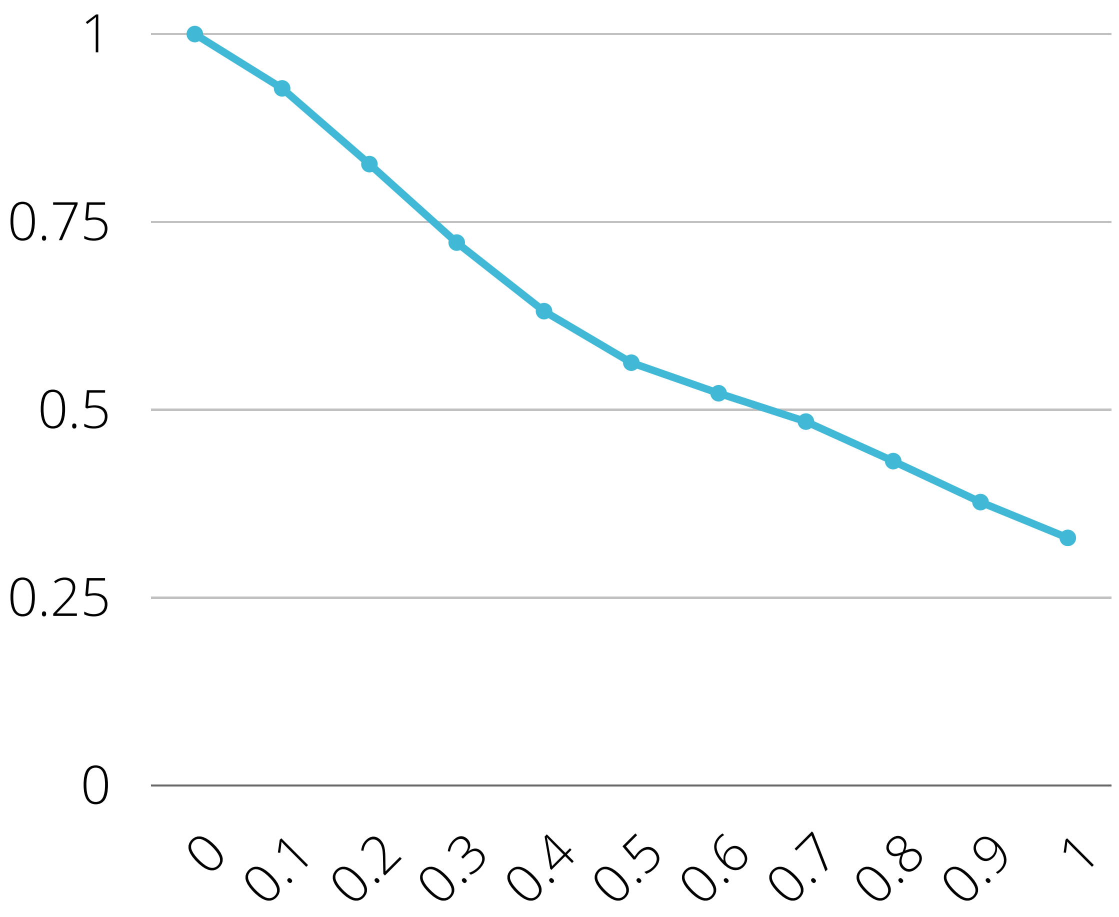
- Предложенный подход часто используется для описания временных точечных процессов. Их реализацией является последовательность событий



PDF



Intensity Function



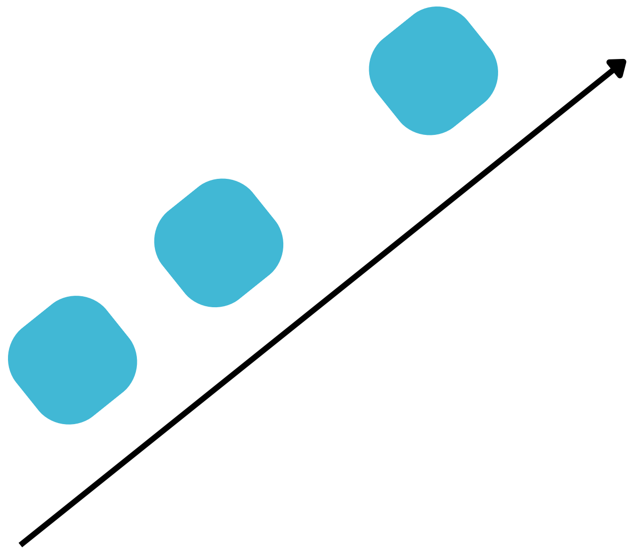
Survival Function

Когда интенсивности недостаточно

- Интенсивность - очень полезная величина
- Интенсивность доказала свою пользу для множества реальных задач
- Не все временные точечные процессы могут быть описаны с помощью функции интенсивности
- Например, интенсивность сама может быть случайной величиной
- В данной работе рассматривался случай измеримой интенсивности, поэтому углубляться в данную тему не будем

Часть 2

- Последовательности событий
- Маркированные последовательности событий
- Классические модели
- Нейросетевые модели

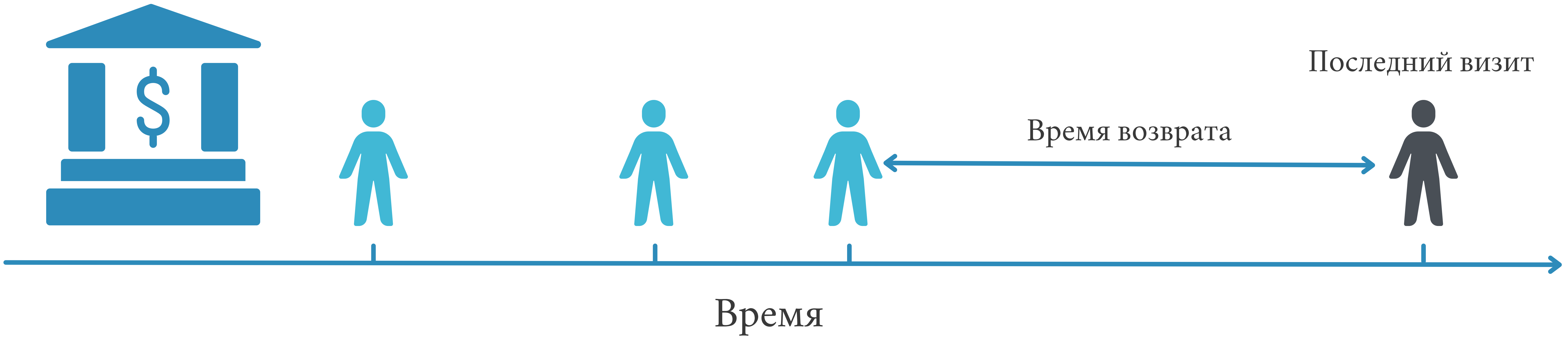


Время возврата

- Когда пользователь в следующий раз вернется в банк?
- Мы можем еще предсказывать тип события
- Время возврата - это время между событий

$$\Delta t_{i+1} = t_{i+1} - t_i$$

Что если в событии есть еще информация?

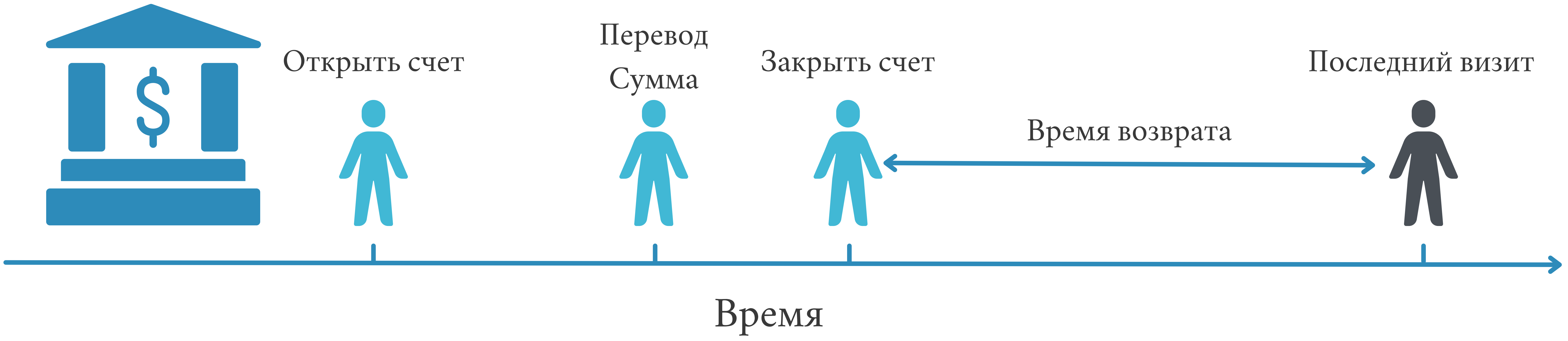


Время возврата

- Когда пользователь в следующий раз вернется в банк?
- Мы можем еще предсказывать тип события
- Время возврата - это время между событий

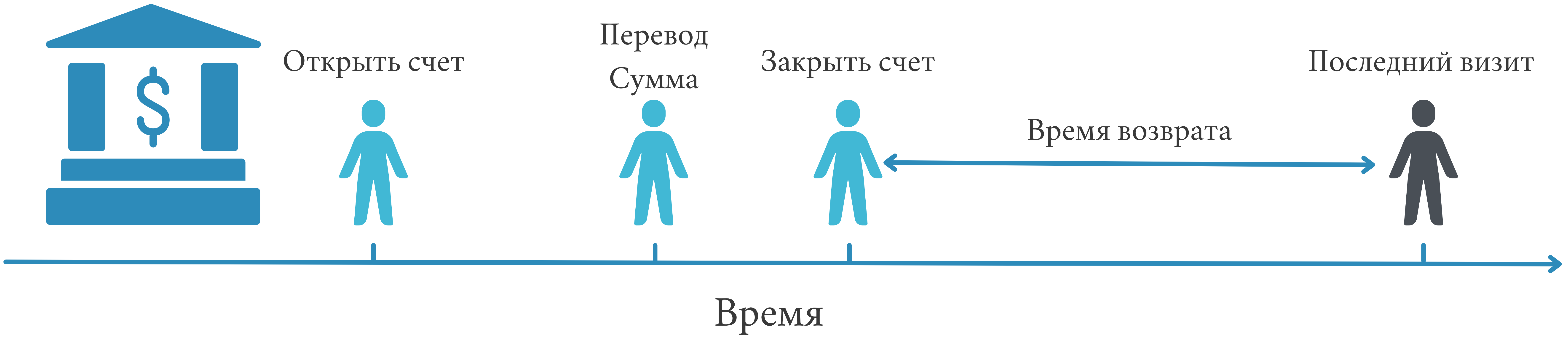
$$\Delta t_{i+1} = t_{i+1} - t_i$$

Что если в событии есть еще информация?



Время возврата

- Рассмотрим маркированные события, где марка - это тип события из множества $\{1, \ldots, C\}$.
- Теперь событие определяется парой время-тип события



Функция интенсивности с марками

- Показывает частоту <u>типа</u> события
- Пусть $N_c(t) = \#\{t_i < t | c_i = c\}$
- Функция интенсивности

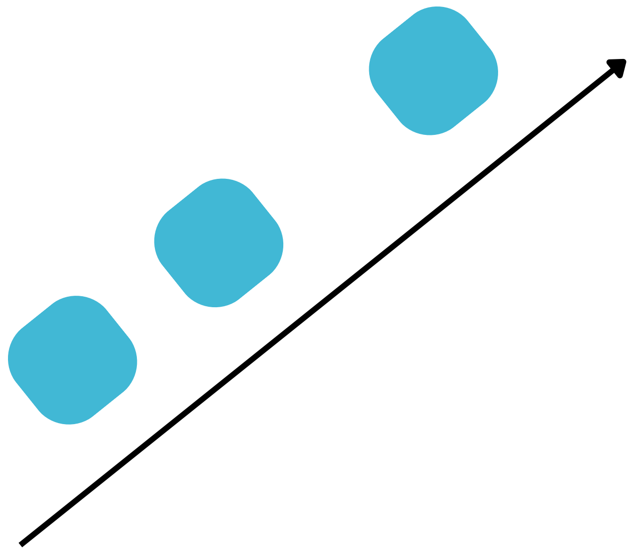
$$\lambda_c(t) = \mathbb{E}\frac{dN_c}{dt}$$

- Общая интенсивность - сумма интенсивностей
- Связь с плотностью вероятности:

$$f(\boldsymbol{s}) = \prod_i \lambda_{c_i}(t_i) \exp\left(-\int_0^T \sum_c \lambda_c(\tau) d\tau\right)$$

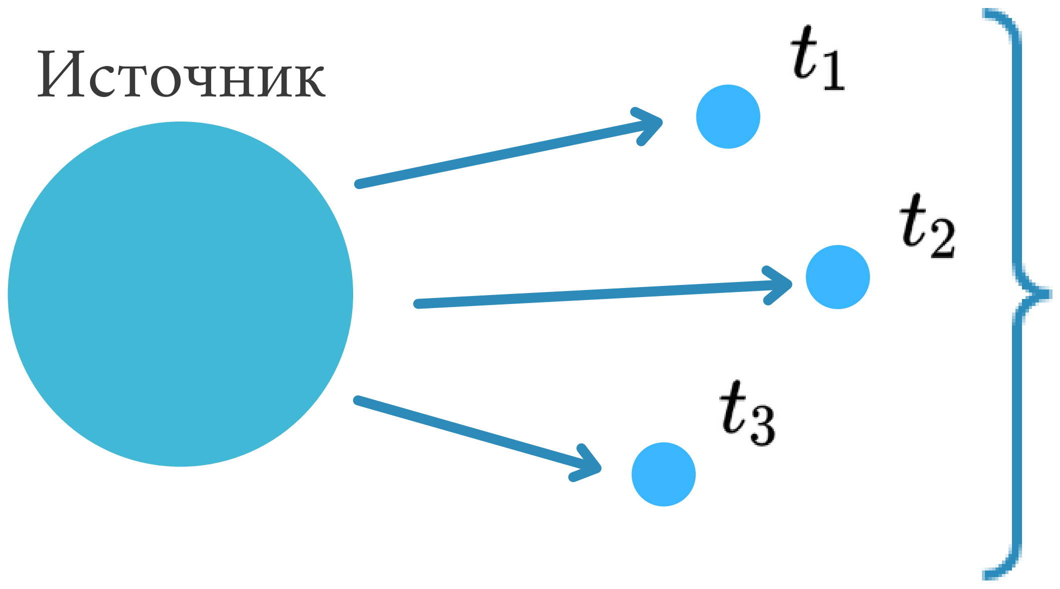
Часть 3

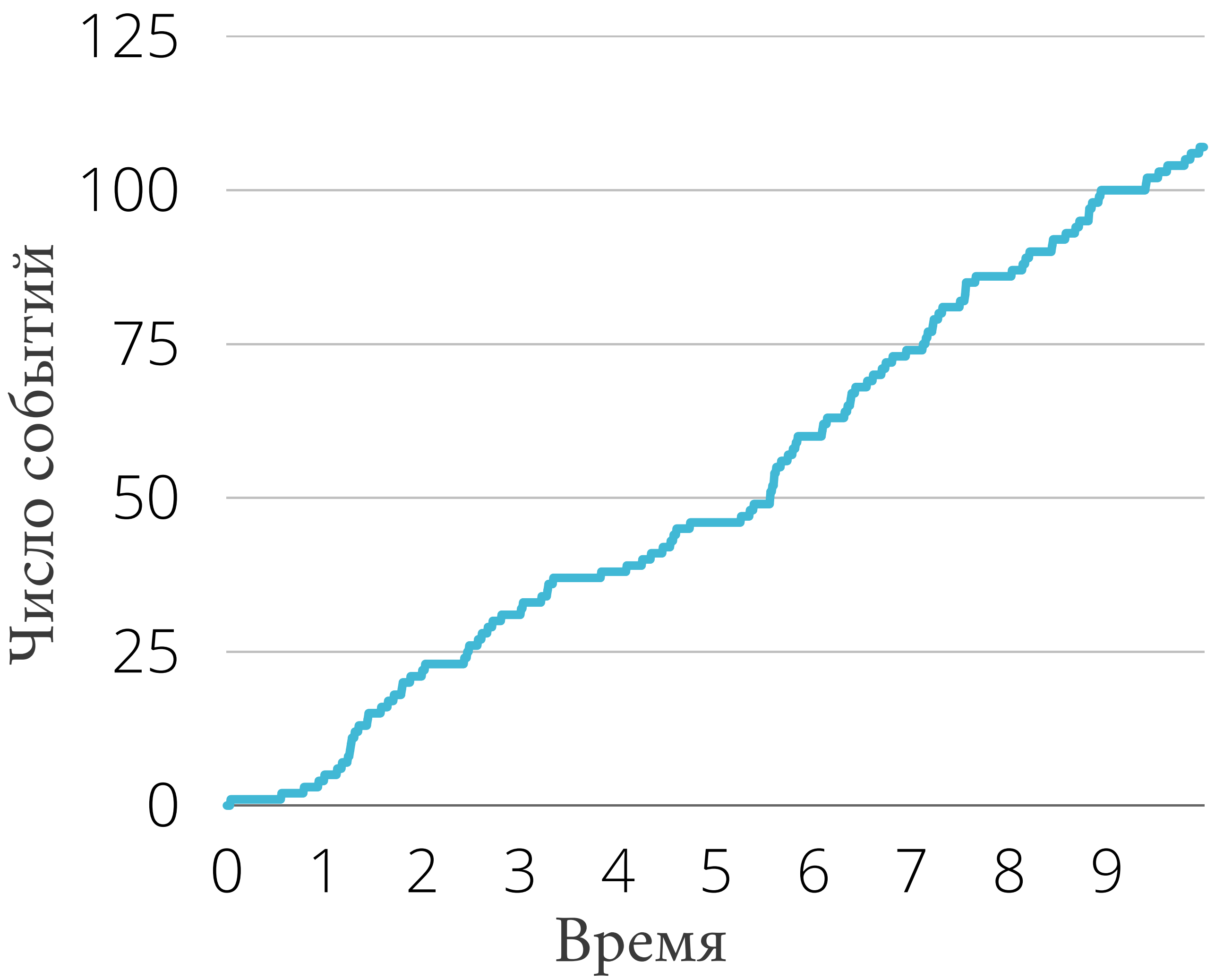
- Последовательности событий
- Маркированные последовательности событий
- Классические модели
- Нейросетевые модели



Процесс Пуассона

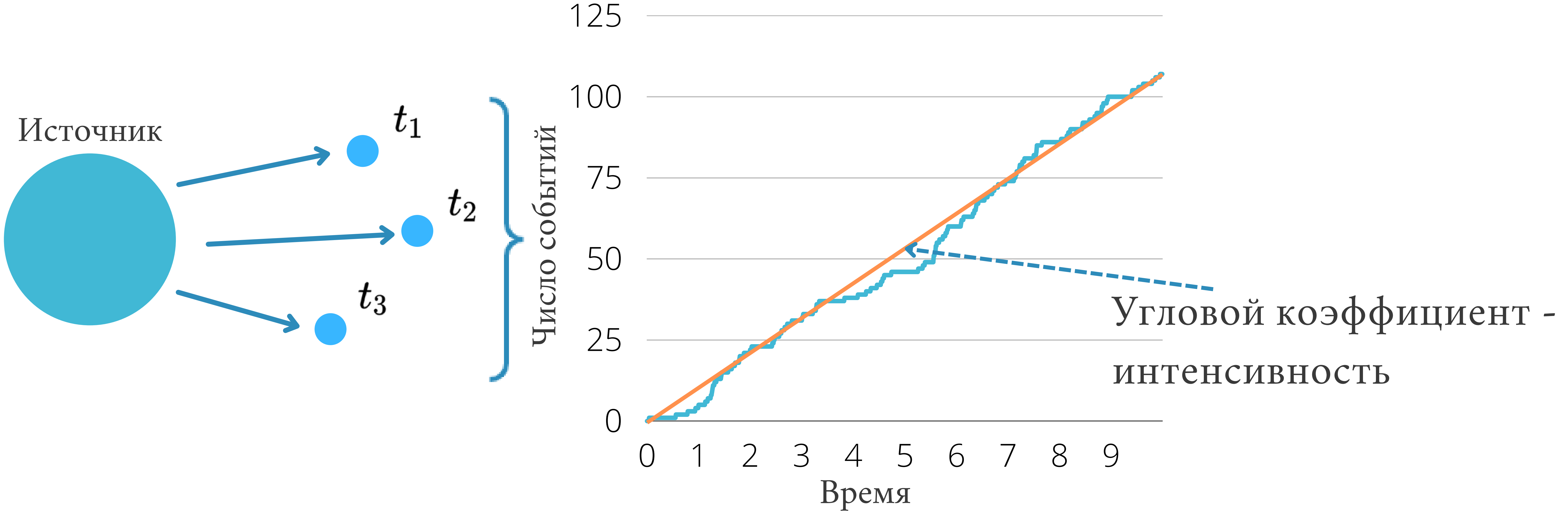
- Простейшая модель для последовательностей событий
- $\lambda = \text{const}$
- Источник с фиксированной интенсивностью, например поток космических частиц
- Если интенсивность зависит от времени, то это неоднородный процесс
- Редко встречается в реальном мире





Процесс Пуассона

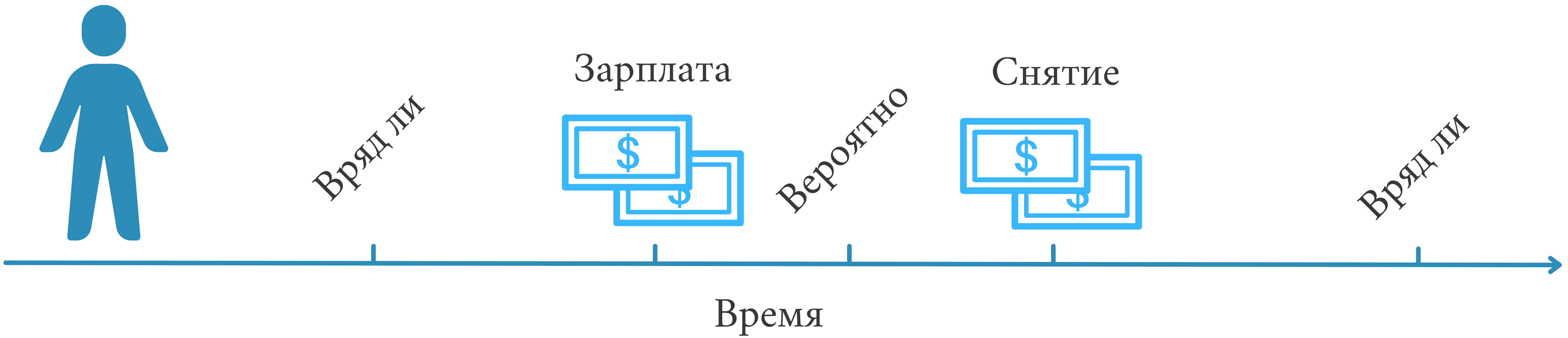
- Простейшая модель для последовательностей событий
- $\lambda = \text{const}$
- Источник с фиксированной интенсивностью, например поток космических частиц
- Если интенсивность зависит от времени, то это неоднородный процесс
- Редко встречается в реальном мире



Не Пуассон

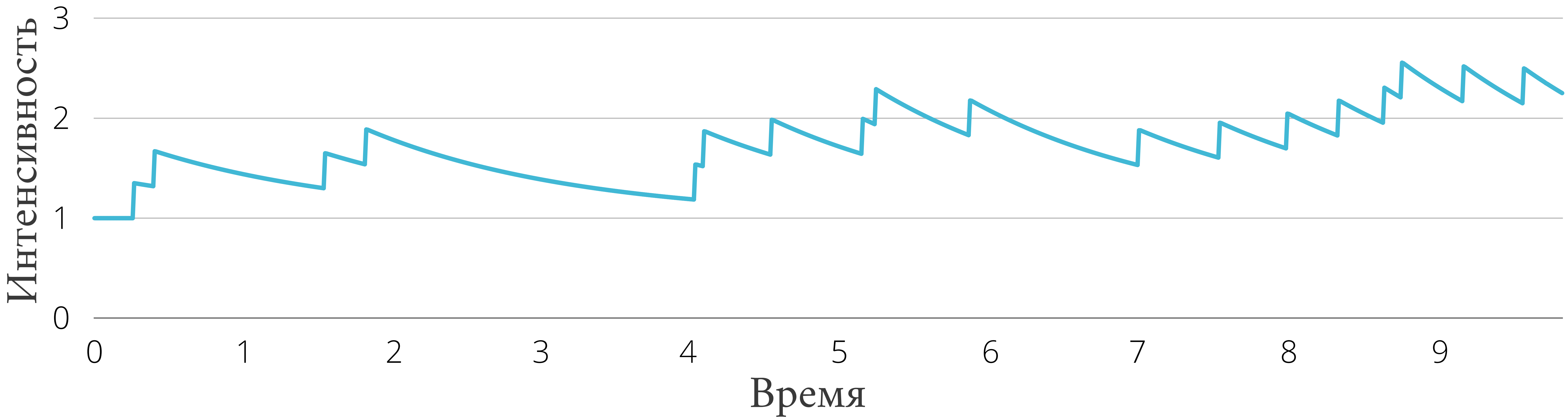
- В пуассоновском процессе интенсивность не зависит от истории
- В реальности это часто не так
- Например:
  - Зарплата - внешний параметр, может быть включен с помощью неоднородного пуассона
  - Снятие наличных - зависит от зарплаты и предыдущих снятий

Когда пользователь снимет наличные



Процесс Хоукса

- Одна из простейших моделей, учитывающих историю - процесс Хоукса
- Был предложен Аланом Хоуксом в 1971
- Предположения:
  - Каждое событие влияет на интенсивность
  - Влияния аддитивны
  - Влияния неотрицательны



Процесс Хоукса

- Интенсивность может быть записана как

$$\lambda(t|\mathcal{H}) = \mu(t) + \sum_{i:t_i<t} \varphi(t - t_i)$$

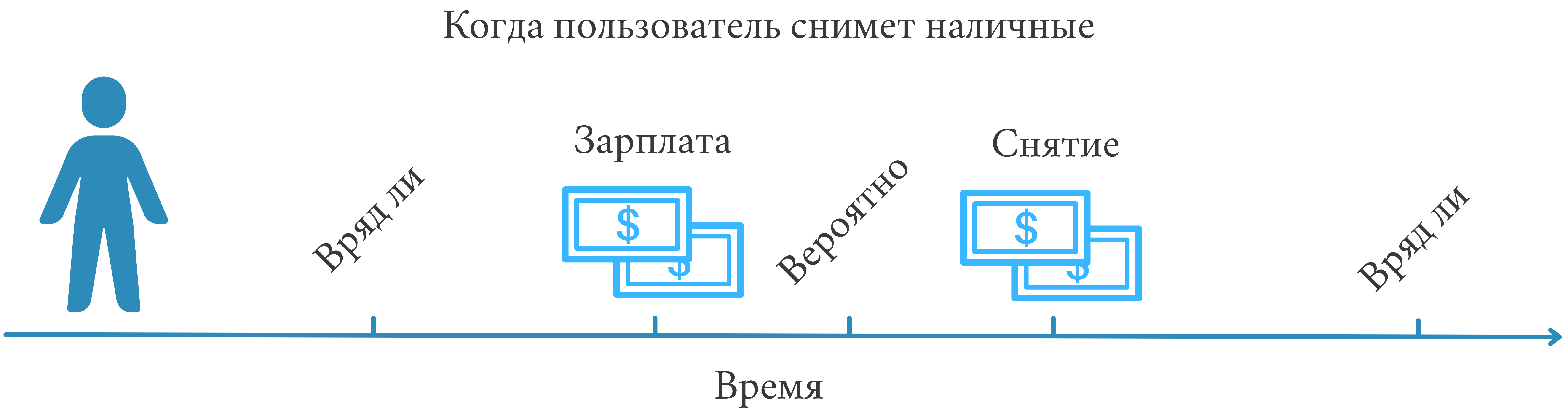
- Таким образом мы вводим условную интенсивность
- Она связана с условной вероятностью
- Данная интенсивность все еще может быть использована

$$f(\mathbf{s}) = \prod_i \lambda(t_i|\mathcal{H}_{t_i}) \exp\left(-\int_0^T \lambda(t|\mathcal{H})dt\right)$$

- Это можно доказать используя правило цепи
- Вопрос:
  - Какие ограничения у процесса Хоукса?

Подавление и отсутствие аддитивности

- Не все процессы самовозбуждающиеся с аддитивным вкладом
- Чтобы добавить подавления и неаддитивность можно либо модифицировать Хоукс (добавить ReLU), либо использовать другие модели (Нейросети)

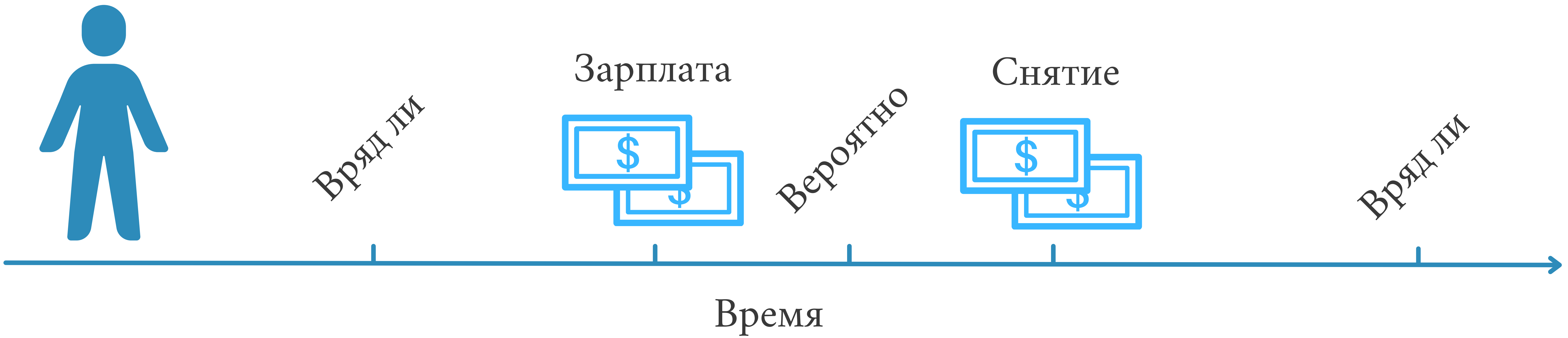


Несколько типов событий

- В последовательности событий события могут быть разных типов
- Мы все еще можем использовать те же модели
- Интенсивность в качестве вектора

$$\lambda_c(t|\mathcal{H}) = \mu_c(t) + \sum_{i:t_i<t} \varphi_{c,c_i}(t - t_i)$$

Когда пользователь снимет наличные



Тренировка моделей

- Основная задача - максимизация правдоподобия

$$f(\mathbf{s}) = \prod_i \lambda(t_i|\mathcal{H}_{t_i}) \exp\left(-\int_0^T \lambda(t|\mathcal{H})dt\right)$$

- Необходимо выбрать модель и затем максимизировать правдоподобие относительно параметров модели
- Вычисление интеграла часто является наиболее трудоемкой задачей

Часть 4

- Последовательности событий
- Маркированные последовательности событий
- Классические модели
- **Нейросетевые модели**

Нейросетевые модели

Классические временные ряды

Последовательности событий

Нейросетевые модели

Классические временные ряды

- Рекуррентные нейронные сети (RNN, LSTM, GRU)
- CNN для временных рядов
- Трансформеры

Последовательности событий

Нейросетевые модели

Классические временные ряды

- Рекуррентные нейронные сети (RNN, LSTM, GRU)
- CNN для временных рядов
- Трансформеры

Временной лаг используется неявно, все точки равноудаленные, явного времени в модели нет

Последовательности событий

Нейросетевые модели

Классические временные ряды

- Рекуррентные нейронные сети (RNN, LSTM, GRU)
- CNN для временных рядов
- Трансформеры

Временной лаг используется неявно, все точки равноудаленные, явного времени в модели нет

Последовательности событий

- Рекуррентные нейронные сети (RNN, LSTM, GRU)?
- CNN для временных рядов?
- Трансформеры?

Нейросетевые модели

Классические временные ряды

- Рекуррентные нейронные сети (RNN, LSTM, GRU)
- CNN для временных рядов
- Трансформеры

Временной лаг используется неявно, все точки равноудаленные, явного времени в модели нет

Последовательности событий

- Рекуррентные нейронные сети (RNN, LSTM, GRU)?
- CNN для временных рядов?
- Трансформеры?

Ожидаем что-то подобное.
Вопрос: а что со временем события?

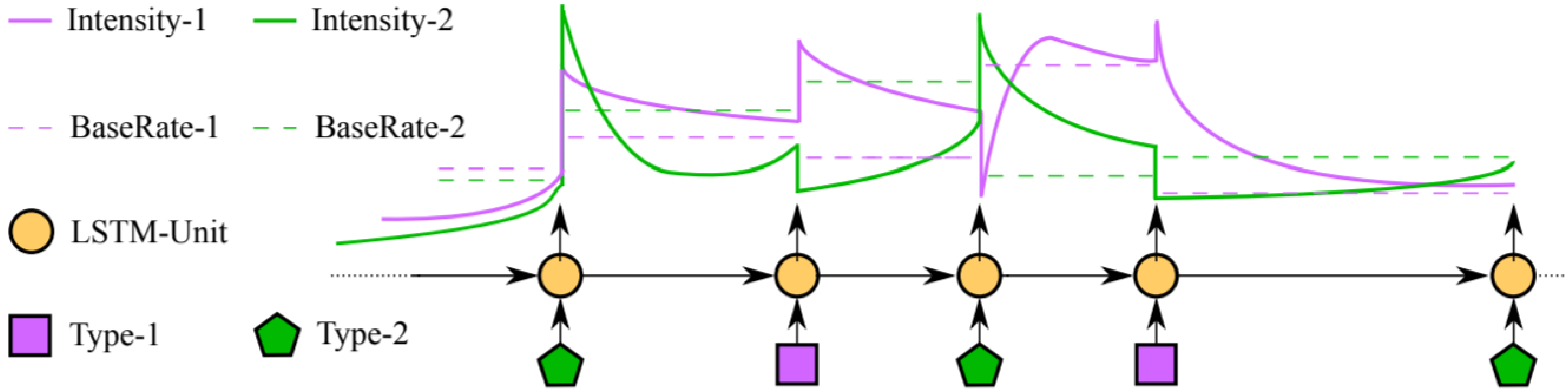
Neural Hawkes

- LSTM с дополнительной эволюцией состояния между событиями

$$\mathbf{h}(t) = \mathbf{o}_i \odot (2\sigma(2\mathbf{c}(t)) - 1)$$

$$\mathbf{c}(t) = \bar{\mathbf{c}}_{i+1} + (\mathbf{c}_{i+1} - \bar{\mathbf{c}}_{i+1}) \exp(-\boldsymbol{\delta}_{i+1}(t - t_i))$$

$$\lambda_k(t) = f_k(\mathbf{w}_k^\top \mathbf{h}(t)), \; f_k(u) = s_k \log(1 + \exp(u/s_k))$$

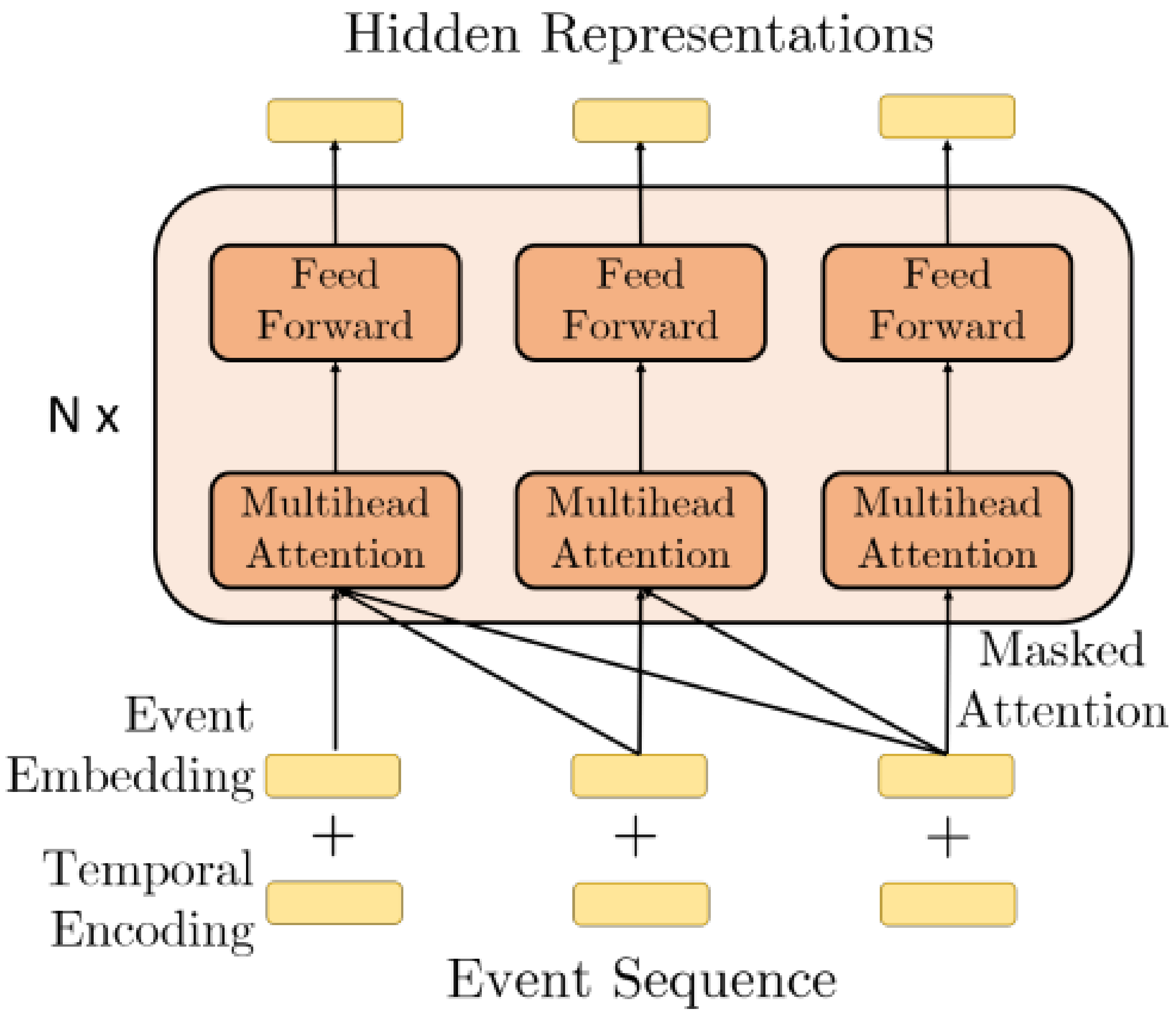


Transformer Hawkes

- Трансформер с кодированием времени и интерполяцией интенсивности

$$\lambda_k(t|\mathcal{H}_t) = f_k\left(\alpha_k \frac{t - t_j}{t_j} + \mathbf{w}_k^\top \mathbf{h}(t_j) + b_k\right),\ t \in [t_j, t_{j+1})$$

$$f_k(u) = s_k \log(1 + \exp(u/s_k))$$



Тренировка моделей

- Основная задача - максимизация правдоподобия

$$f(\mathbf{s}) = \prod_i \lambda(t_i|\mathcal{H}_{t_i}) \exp\left(-\int_0^T \lambda(t|\mathcal{H})dt\right)$$

- Необходимо выбрать модель и затем максимизировать правдоподобие относительно параметров модели
- Вычисление интеграла часто является наиболее трудоемкой задачей, используем метод Монте-Карло

$$\hat{\Lambda}_{MC} = \sum_{j=2}^{L}(t_j - t_{j-1})\left(\frac{1}{N}\sum_{i=1}^{N}\lambda(u_i)\right),\ u_i \sim Unif(t_{j-1}, t_j)$$