Probability Foundations

Reminder: Probability Theory

- **Goal**: Make statements and/or predictions about results of physical processes.
- Even processes that seem to be simple at first sight may reveal considerable difficulties when trying to predict.
- Describing real-world physical processes always calls for a simplifying mathematical model.
- Although everybody will have some intuitive notion about probability, we have to formally define the underlying mathematical structure.
- Randomness or chance enters as the incapability of precisely modelling a process or the inability of measuring the initial conditions.
    - *Example*: Predicting the trajectory of a billard ball over more than 9 banks requires more detailed measurement of the initial conditions (ball location, applied momentum etc.) than physically possible according to Heisenberg's uncertainty principle.

Reality vs. Model

- Producing a result of a physical process is referred to as an **observed outcome**.
- Assessing or predicting the probability of every possible outcome is not straightforward but often implicitly assumed to be clear.
- We will study this “non-straightforwardness” with three real-world examples:
  - Rolling a die.
  - Arrivals of inquiries at a call center.
  - The weight of a bread roll purchased from a bakery.
    (Inspired by a broadcast of Quarks & Co. from WDR.)
- Obviously, all examples differ in the nature of the space of possible observable outcomes.

Example 1: Rolling a Die

- **Physical Process**
  Shaking a six-sided die in a dice cup.
  Then cast it and read off the number of pips.

- **Possible Outcomes**

- **Sources of Randomness**
  - Inaccurate knowledge about locations, momenta.
  - Inellastic collisions inside the dice cup.
  - Inhomogeneous material distribution of the die.
  - Uneven table surface.
  - Unknown frictions, airflow etc.

- **Model**
  Outcomes have equal probability.

Example 2: Phone Calls at a Call Center

- **Physical Process**
  Counting the number of phone calls that arrive
  at a call center within a predefined time window.

- **Possible Outcomes**
  The events (if any) happening in time and space.

- **Sources of Randomness**
  - Calls are initiated by human beings: no predictability.
  - Misdialed calls.
  - Technical problems resulting in lost calls.

- **Model**
  Poisson distribution of number of calls.

Example 3: Bread Rolls at a Bakery

- **Physical Process**
  Baking a bread roll from a piece of dough.
  Measuring its weight (with arbitrary precision).

- **Possible Outcomes**
  Bread rolls.

- **Sources of Randomness**
  - Amount of dough put on the baking sheet.
  - Baking process (ingredients, temperature, time).

- **Model**
  Gaussian distribution of the weight.

Formal Approach on the Model Side

- We conduct an experiment that has a set Ω of possible outcomes.
  E. g.:
  - Rolling a die ($\Omega = \{1, 2, 3, 4, 5, 6\}$)
  - Arrivals of phone calls ($\Omega = \mathbb{N}_0$)
  - Bread roll weights ($\Omega = \mathbb{R}_+$)
- Such an outcome is called an **elementary event**.
- All possible elementary events are called the **frame of discernment** Ω (or sometimes **universe of discourse**).
- The set representation stresses the following facts:
  - All possible outcomes are covered by the elements of Ω. (**collectively exhaustive**).
  - Every possible outcome is represented by exactly one element of Ω. (**mutual disjoint**).

Events

- Often, we are interested in *higher-level* events (e. g. casting an odd number, arrival of at least 5 phone calls or purchasing a bread roll heavier than 80 grams)

- Any subset $A \subseteq \Omega$ is called an **event** which **occurs**, if the outcome $\omega_0 \in \Omega$ of the random experiment lies in A:

$$\text{Event } A \subseteq \Omega \text{ occurs} \quad \Leftrightarrow \quad \bigvee_{\omega \in A} (\omega = \omega_0) = \mathsf{true} \quad \Leftrightarrow \quad \omega_0 \in A$$

- Since events are sets, we can define for two events A and B:
  - $A \cup B$ occurs if A or B occurs; $A \cap B$ occurs if A and B occurs.
  - $\overline{A}$ occurs if A does not occur (i. e., if $\Omega \backslash A$ occurs).
  - A and B are *mutually exclusive*, iff $A \cap B = \emptyset$.

Event Algebra

- A family of sets $\mathcal{E} = \{E_1, \ldots, E_n\}$ is called an **event algebra**, if the following conditions hold:
  - The **certain event** Ω lies in $\mathcal{E}$.
  - If $E \in \mathcal{E}$, then $\overline{E} = \Omega \backslash E \in \mathcal{E}$.
  - If E_1 and E_2 lie in $\mathcal{E}$, then $E_1 \cup E_2 \in \mathcal{E}$ and $E_1 \cap E_2 \in \mathcal{E}$.
- If Ω is uncountable, we require the additional property:

  For a series of events $E_i \in \mathcal{E}, i \in \mathbb{N},$ the events $\bigcup_{i=1}^{\infty} E_i$ and $\bigcap_{i=1}^{\infty} E_i$ are also in $\mathcal{E}$.

  $\mathcal{E}$ is then called a **σ-algebra**.

Side remarks:

- Smallest event algebra: $\mathcal{E} = \{\emptyset, \Omega\}$
- Largest event algebra (for finite or countable Ω): $\mathcal{E} = 2^\Omega = \{A \subseteq \Omega \mid \mathsf{true}\}$

Probability Function

- Given an event algebra $\mathcal{E}$, we would like to assign every event $E \in \mathcal{E}$ its probability with a **probability function** $P: \ \mathcal{E} \to [0, 1]$.

- We require P to satisfy the so-called **Kolmogorov Axioms**:

  - $\forall E \in \mathcal{E}: \ 0 \ \leq \ P(E) \ \leq \ 1$
  - $P(\Omega) \ = \ 1$
  - For pairwise disjoint events $E_1, E_2, \ldots \in \mathcal{E}$ holds:

$$P(\bigcup_{i=1}^{\infty} E_i) \ = \ \sum_{i=1}^{\infty} P(E_i)$$

  Note that for $|\Omega| < \infty$ the union and sum are finite also.

- From these axioms one can conclude the following (incomplete) list of properties:

  - $\forall E \in \mathcal{E}: \ P(\overline{E}) \ = \ 1 - P(E)$
  - $P(\emptyset) \ = \ 0$
  - If $E_1, E_2 \in \mathcal{E}$ are mutually exclusive, then $P(E_1 \cup E_2) = P(E_1) + P(E_2)$.

Elementary Probabilities and Densities

Question 1: How to calculate P?

Question 2: Are there "default" event algebras?

- Idea for question 1: We have to find a way of distributing (thus the notion *distribution*) the unit mass of probability over all elements $\omega \in \Omega$.
  - If Ω is finite or countable a **probability mass function** p is used:

$$p: \ \Omega \to [0,1] \quad \text{and} \quad \sum_{\omega \in \Omega} p(\omega) = 1$$

  - If Ω is uncountable (i. e., continuous) a **probability density function** f is used:

$$f: \ \Omega \to \mathbb{R} \quad \text{and} \quad \int_{\Omega} f(\omega) \, \mathrm{d}\omega = 1$$

“Default” Event Algebras

- Idea for question 2 (“default” event algebras) we have to distinguish again between the cardinalities of Ω:
  - Ω finite or countable: $\mathcal{E} = 2^{\Omega}$
  - Ω uncountable, e. g. $\Omega = \mathbb{R}$: $\mathcal{E} = \mathcal{B}(\mathbb{R})$
- $\mathcal{B}(\mathbb{R})$ is the **Borel Algebra**, i. e., the smallest σ-algebra that contains all closed intervals $[a, b] \subset \mathbb{R}$ with $a < b$.
- $\mathcal{B}(\mathbb{R})$ also contains all open intervals and single-item sets.
- It is sufficient to note here, that all intervals are contained

$$\{[a, b]\,,]a, b]\,,]a, b[\,, [a, b[\, \subset \mathbb{R} \mid a < b\} \subset \mathcal{B}(\mathbb{R})$$

  because the event of a bread roll having a weight between 80 g and 90 g is represented by the interval $[80, 90]$.

Probability Spaces

- For a sample space A, an event algebra B (over A) and a probability function C, we call the triple

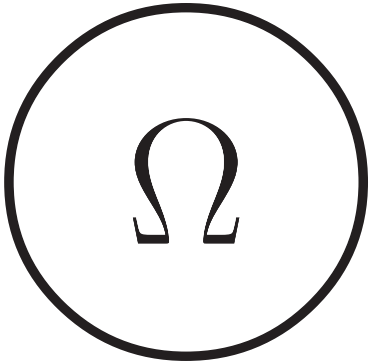
$$(A, B, C)$$

a **probability space**.

Real World

$(\Xi, \mathcal{X}, Q)$

Model



$(\Omega, \mathcal{E}, P)$

Reminder: Preimage of a Function

- Let $f: D \to M$ be a function that assigns to every value of D a value in M.
- For every value of $y \in M$ we can ask which values of $x \in D$ are mapped to y:

$$D \supseteq \{x \in D \mid f(x) = y\} \stackrel{\text{Def}}{=} f^{-1}(y)$$

- $f^{-1}(y)$ is called the **preimage** of y under f, denoted also as $\{f = y\}$.
- The notion can be generalized from $y \in M$ to sets $B \subseteq M$:

$$D \supseteq \{x \in D \mid f(x) \in B\} \stackrel{\text{Def}}{=} f^{-1}(B)$$

- If f is bijective then $\forall y \in M : \left|f^{-1}(y)\right| = 1$.
- Examples:
  - $\sin^{-1}(0) = \{k \cdot \pi \mid k \in \mathbb{Z}\}$
  - $\exp^{-1}(1) = \{0\}$
  - $\operatorname{sgn}^{-1}(1) = (0, +\infty) \subset \mathbb{R}$

Random Variable

We still need a means of mapping real-world outcomes in Ξ to our space Ω.

- A function $X: D \to M$ is called a **random variable** iff the preimage of any value of M is an event (in some probability space).

- If X maps Ξ onto Ω, we define

$$P_X(X \in A) \;=\; Q(\{\xi \in \Xi \mid X(\xi) \in A\}).$$

- X may also map from Ω to another domain: $X: \Omega \to \text{dom}(X)$. We then define:

$$P_X(X \in A) \;=\; P(\{\omega \in \Omega \mid X(\omega) \in A\}).$$

- If X is numeric, we call $F(x)$ with

$$F(x) \;=\; P(X \leq x)$$

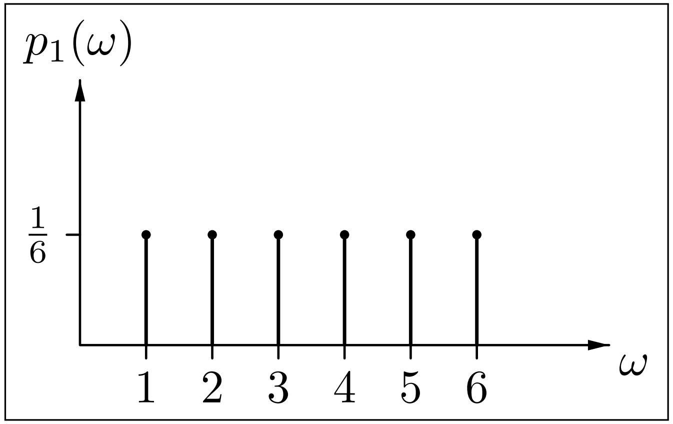
the **distribution function** of X.

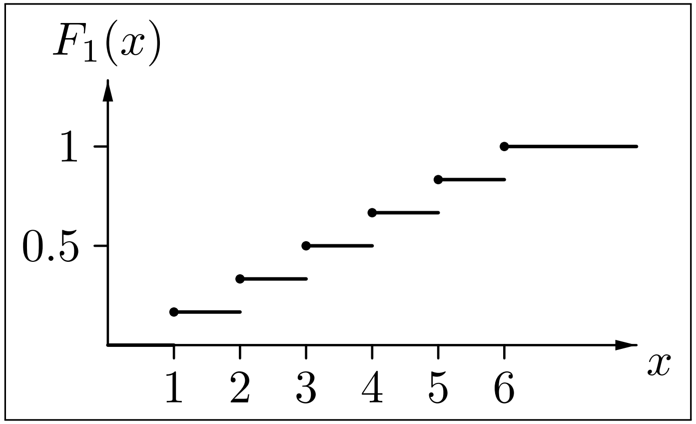
Example: Rolling a Die

$\Omega = \{1, 2, 3, 4, 5, 6\} \quad X = \text{id}$

$p_1(\omega) = \frac{1}{6}$

$F_1(x) = P(X \leq x)$





$$\sum_{\omega \in \Omega} p_1(\omega) = \sum_{i=1}^{6} p_1(\omega_i) = \sum_{i=1}^{6} \frac{1}{6} = 1$$

$$P(X \leq x) = \sum_{x' \leq x} P(X = x')$$

$$P(a < X \leq b) = F_1(b) - F_1(a)$$

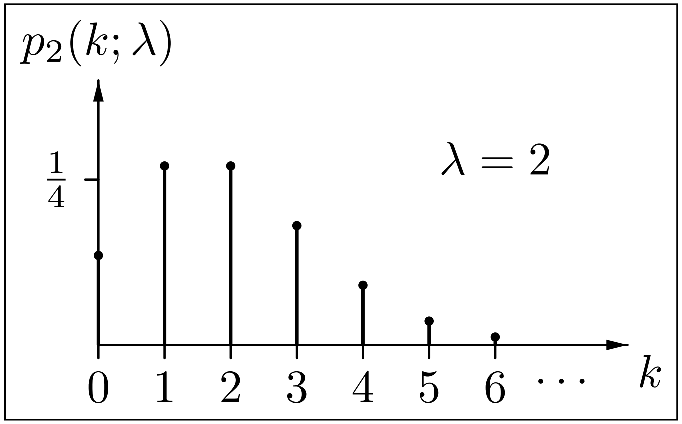
$$P(X = x) = P(\{X = x\}) = P(X^{-1}(x)) = P(\{\omega \in \Omega \mid X(\omega) = x\})$$

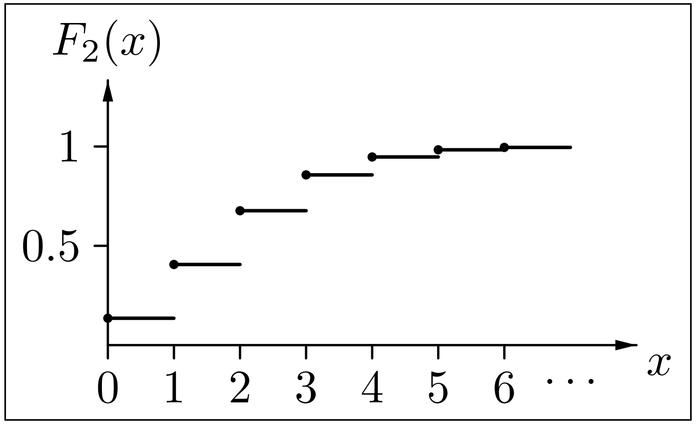
Example: Arriving Phone Calls

$\Omega = \mathbb{N}_0 \quad X = \mathrm{id}$

$$p_2(k;\lambda) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$

$$F_2(k;\lambda) = \sum_{i=0}^{k} e^{-\lambda} \cdot \frac{\lambda^i}{i!}$$





$$\sum_{k\in\mathbb{N}_0} p_2(k;\lambda) = \sum_{k=0}^{\infty} e^{-\lambda} \cdot \frac{\lambda^k}{k!} = e^{-\lambda} \cdot \underbrace{\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}}_{=e^{\lambda}} = e^{-\lambda} \cdot e^{\lambda} = 1$$

$$P(X \le x) = \sum_{x' \le x} P(X = x')$$

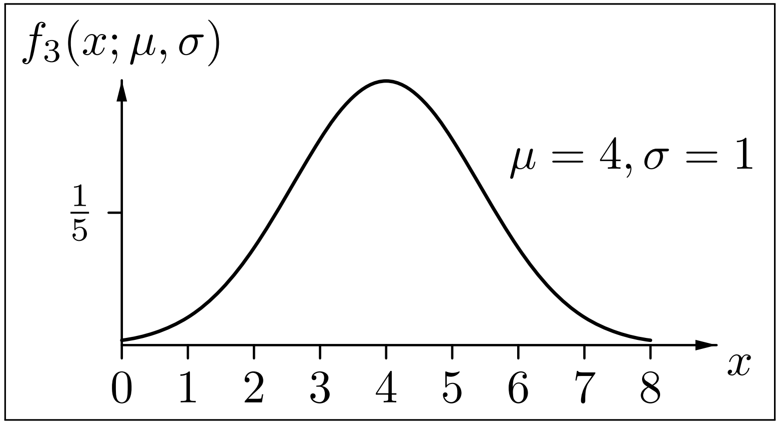
$$P(a < X \le b) = F_2(b) - F_2(a)$$

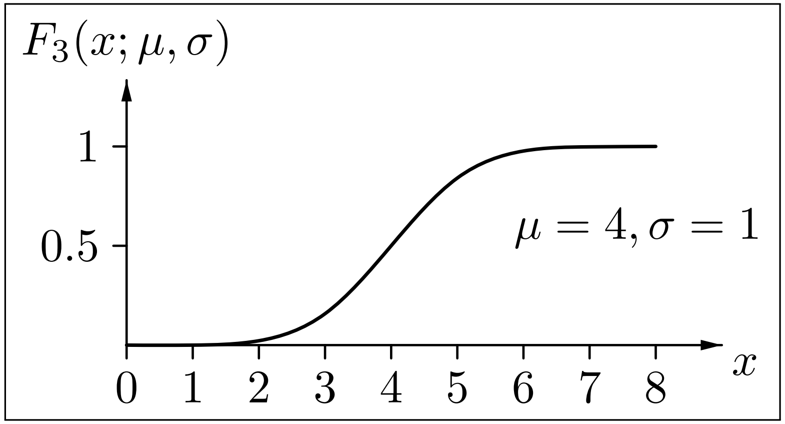
Example: Weight of a Bread Roll

$\Omega = \mathbb{R} \quad X = \text{id}$

$$f_3(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$F_3(x) = \int_{-\infty}^{x} f_3(x)\,\mathrm{d}x$$



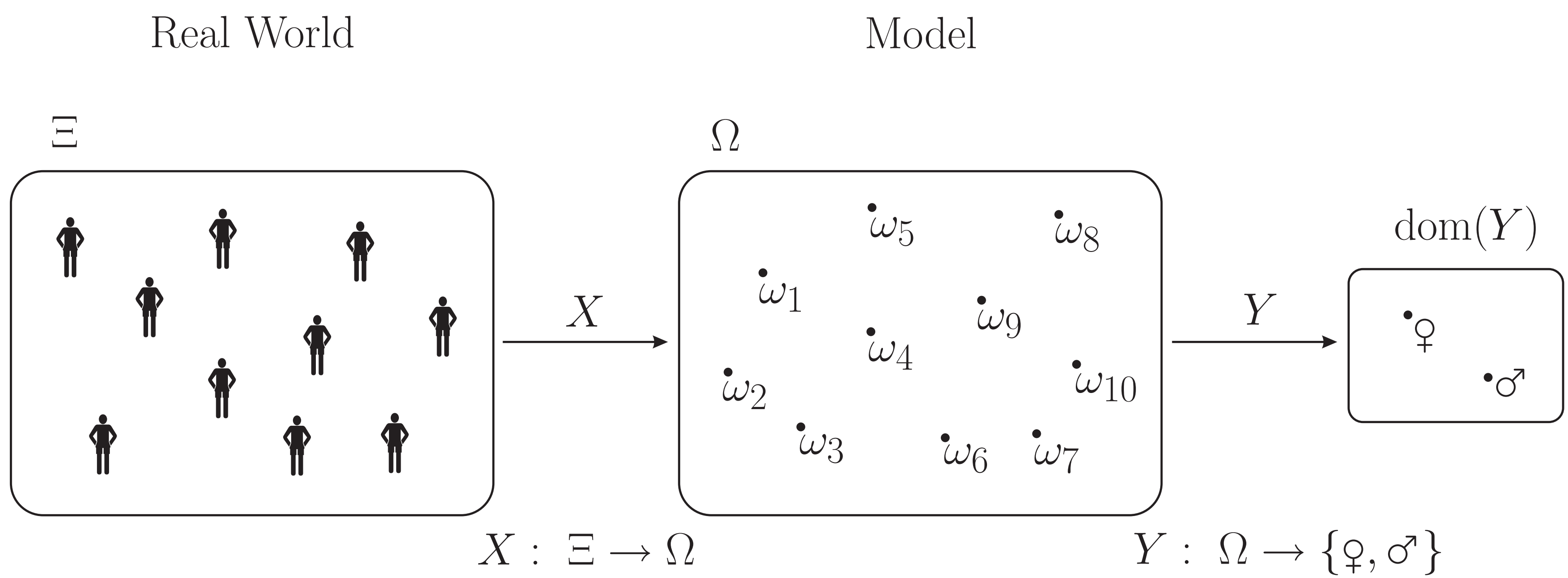


$$\int_{-\infty}^{+\infty} f_3(x)\,\mathrm{d}x = 1$$

$$\begin{aligned} P(X \le x) &= P(]-\infty, x]) \\ &= \int_{-\infty}^{x} f_3(x)\mathrm{d}x \end{aligned}$$

$$\begin{aligned} P(a < X \le b) &= P(]a, b]) \\ &= \int_{a}^{b} f_3(x)\mathrm{d}x \\ &= F_3(b) - F_3(a) \end{aligned}$$

The Big Picture



$$Q\Big(\{\xi \in \Xi \mid X(\xi) \in Y^{-1}(♀)\}\Big) \quad = \quad P\Big(\{\omega \in \Omega \mid Y(\omega) = ♀\}\Big) \quad = \quad P\Big(Y = ♀\Big) \quad = \quad P\Big(♀\Big)$$

Poisson Distribution

- Limit case of the Binomial distribution:

$$\lim_{n\to\infty} b_X(k; n, p) \;=\; \lim_{n\to\infty} \binom{n}{k} p^k (1-p)^{n-k} \;=\; e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$

  with $k = 0, 1, 2, \ldots$ and $\lambda = n \cdot p$.

- Expected Value: $E(X) = \lambda$
- Variance: $V(X) = \lambda$
- Models, e. g.
  - Number of cars that pass a gate.
  - Number of customers at a register.
  - Number of calls at a call center.
- λ is the rate parameter (i. e., occurrences per unit time)

Exponential Distribution

- A continuous random variable with density function

$$f_X(x;\lambda) = \begin{cases} \lambda \cdot e^{-\lambda x} & \text{if } x \geq 0, \lambda > 0 \\ 0 & \text{otherwise} \end{cases}$$

  is **exponentially distributed**.

- Expected Value: $E(X) = \frac{1}{\lambda}$ $\qquad F_X(x;\lambda) = \begin{cases} 1 - e^{-\lambda x} & \text{if } x \geq 0, \lambda > 0 \\ 0 & \text{otherwise} \end{cases}$

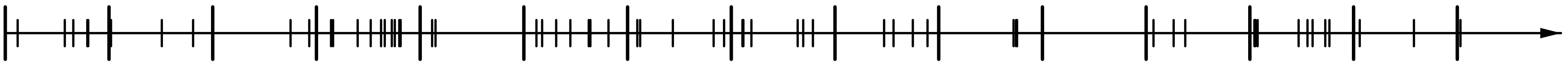
- Variance: $V(X) = \frac{1}{\lambda^2}$

- Models, e. g.
  - Lifetime of electrical devices.
  - Waiting times in a queue.
  - Time between failures of a system.

Relation between Poisson and Exponential Distributions

- Assume an arrival process with λ arrivals (per unit time, say 1h)
- The random variable that describes the **number of arrivals** within the next unit time interval is **Poisson distributed** with parameter λ.
- The random variable that describes the probability of the **waiting times between two arrivals** is **exponentially distributed** with (the same!) λ.

Example:



- Small ticks denote arrivals, large ticks mark unit time windows.
- 60 arrivals, 15 unit time windows.
- Poisson sample $\vec{x}_P = (4, 3, 2, 10, 2, 7, 5, 6, 4, 3, 0, 3, 8, 2, 1)$
- Exponential sample $\vec{x}_E = (0.1192, 0.4544, 0.0821, 0.1352, \ldots)$
- $\lambda = 4$