Quantitative Logics

David N. Jansen

Probability theory revisited

David N. Jansen

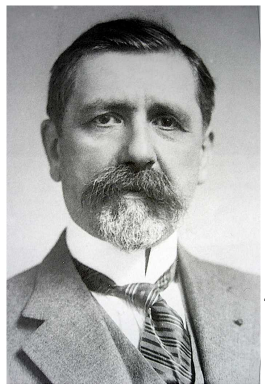
A better definition

- assign probabilities to subsets of Ω in a systematic way
- A σ-algebra $\mathcal{A}$ is a set of subsets:
  - $\Omega \in \mathcal{A}$
  - $A \in \mathcal{A} \rightarrow \Omega \backslash A \in \mathcal{A}$
  - $A_i \in \mathcal{A}$ for all $i = 1,2,\ldots \rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$
- Generally: It is sensible to assign a probability to each set in the σ-algebra.

Example: Borel-σ-algebra

- $\Omega = \mathbb{R}$
- $\mathcal{B}$ = the smallest σ-algebra
  that contains all intervals $[r,s)$, for $r,s \in \mathbb{R}$

- standard σ-algebra for the real numbers
- Émile Borel, French mathematician,
  1871–1956, wrote *Le Hasard*



neurotree.org

Measure space

- A measure is a function $\mu:\mathcal{A} \rightarrow [0,\infty]$ with the properties:
  - $\mu(\emptyset) = 0$
  - σ-additivity: if $A_i \in \mathcal{A}$ for all i=1,2,... are pairwise disjoint sets, then
    $\mu(\cup A_i) = \Sigma\, \mu(A_i)$
- A measure space is a triple $(\Omega,\mathcal{A},\mu)$ where $\mathcal{A}$ is a σ-algebra over Ω and $\mu:\mathcal{A} \rightarrow [0,\infty]$ is a measure.

Finally: Probability space

- A probability measure is a measure μ with: $\mu(\Omega) = 1$

often written as *P*

- A probability space is a measure space $(\Omega,\mathcal{A},P)$ where P is a probability measure.

Example

- A process takes between 2 and 5 minutes ...
- Outcomes: $\Omega = [2,5]$
- σ-algebra: $\mathcal{B}$ restrained to Ω
- measure: $\mu([r,s]) = \mu([r,s)) = (s-r)/3$
- For example:

  $\mu([2,5]) = 1$ $\mu([2,3)) = \frac{1}{3}$

  $\mu([2,3) \cup [4,5)) = \frac{2}{3}$

A little measure theory

- A function defined on a suitable subset of $\mathcal{A}$ can be extended uniquely to a measure on $\mathcal{A}$.

- needed to define probability space of Markov chains

Ring

- A ring $\mathcal{R}$ is a set of subsets of Ω:
  - $\varnothing \in \mathcal{R}$
  - $A_1, A_2 \in \mathcal{R} \rightarrow A_1 \setminus A_2 \in \mathcal{R}$
  - $A_1, A_2 \in \mathcal{R} \rightarrow A_1 \cup A_2 \in \mathcal{R}$

Premeasure

- A premeasure is a function $\mu:\mathcal{R} \rightarrow [0,\infty]$ with the properties:
  - $\mu(\emptyset) = 0$
  - σ-additivity: if $A_i \in \mathcal{R}$ for all i=1,2,... are pairwise disjoint sets, then

    $\mu(\cup A_i) = \Sigma\, \mu(A_i)$

Outer measure

- For every premeasure μ on a ring $\mathcal{R}$, there is at least one measure on $\sigma(\mathcal{R})$ that extends μ.
- Prove using *outer measure* μ^* (Carathéodory):
  - $\mathcal{U}(Q)$:= all sequences of sets $\in \mathcal{R}$ that cover $Q \subseteq \Omega$
  - $\mu^*(Q) := \inf \{ \Sigma\, \mu(A_i) \mid (A_i)_{i\in\mathbb{N}} \in \mathcal{U}(Q) \}$
  - $\mathcal{A}^* := \{ A \subseteq \Omega \mid \mu^*(Q) \geq \mu^*(Q \cap A) + \mu^*(Q \setminus \mathrm{A})$ for all $Q \subseteq \Omega \}$

    is a σ-algebra $\supseteq \mathcal{R}$.
  - μ^* is a measure on $\mathcal{A}^*$.

Uniqueness of extension

- If $\mathcal{E}$ generates a σ-algebra $\mathcal{A}$ and
  - $\mathcal{E}$ is ∩-stable (i. e. $E_1, E_2 \in \mathcal{E} \rightarrow E_1 \cap E_2 \in \mathcal{E}$)
  - there exists a sequence $(E_i)_{i \in \mathbb{N}}$ in $\mathcal{E}$ with $\cup E_i = \Omega$
- then any two measures that coincide on $\mathcal{E}$ and are finite on $(E_i)_{i \in \mathbb{N}}$ are equal.

Recapitulation

- Ω is a set of possible outcomes
  of a random experiment.
- A subset $A \subseteq \Omega$ has probability $P(A)$, i. e.,
  the probability that the random experiment
  results in some outcome $\in A$ is $P(A)$.
- The σ-algebra $\mathcal{A} \subseteq \mathbf{P}(\Omega)$ describes
  to which sets such a probability can be assigned.
- Because P is a measure,
  the probabilities assigned to different subsets of Ω
  are consistent with each other.

Further questions I will not answer

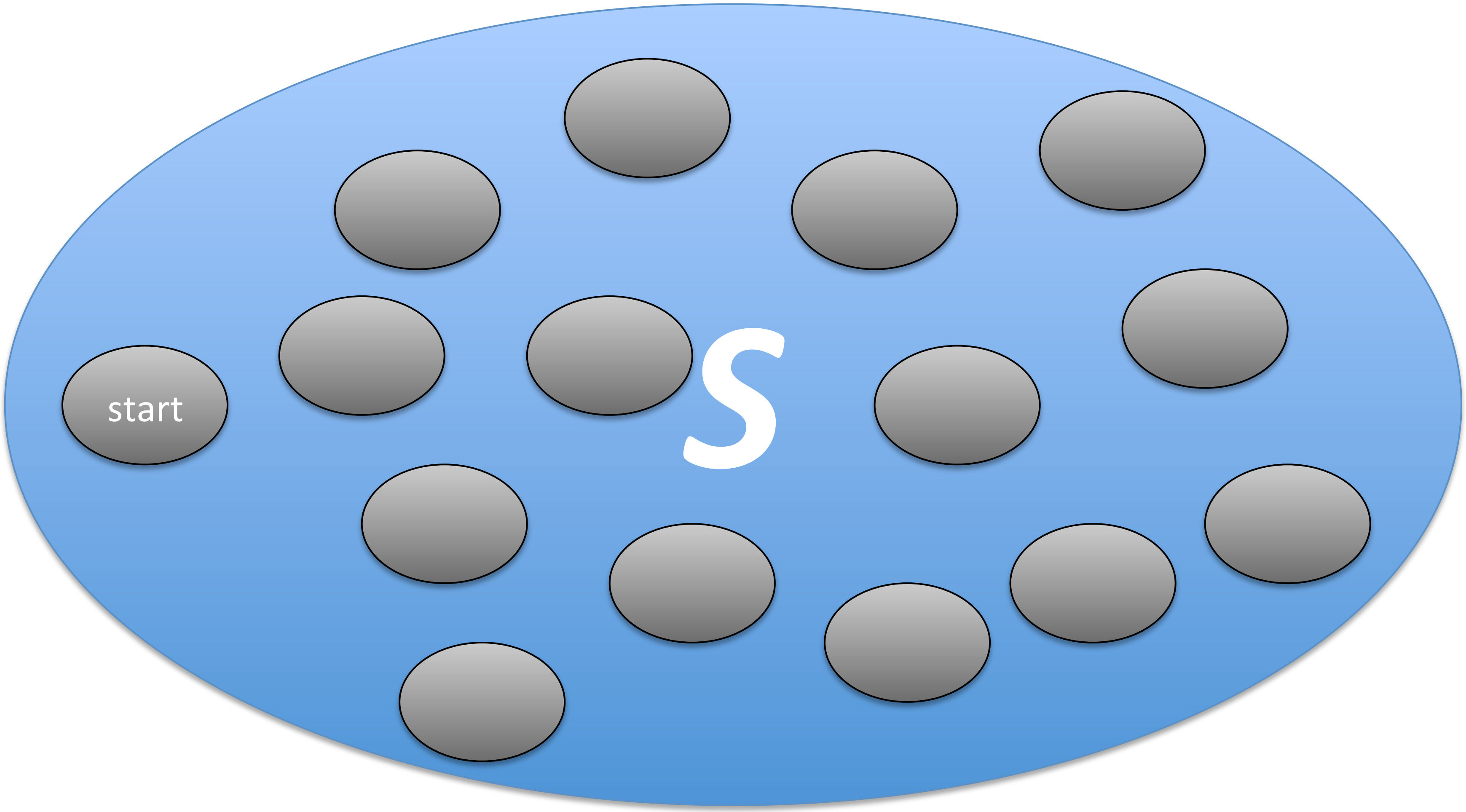
- How do you find the right probability values?
  - I assume that the values are given;
    we do the calculations based on the givens
- What is a probability?
  - frequentist model? model of bets?
- Why should probability spaces be additive?
  - other choices will lead
    to a surely-winning scheme of bets
    http://www.numdam.org/item?id=AIHP_1937__7_1_1_0, page 7
    http://info.phys.unm.edu/~caves/reports/dutchbook.pdf

Discrete-time Markov chains

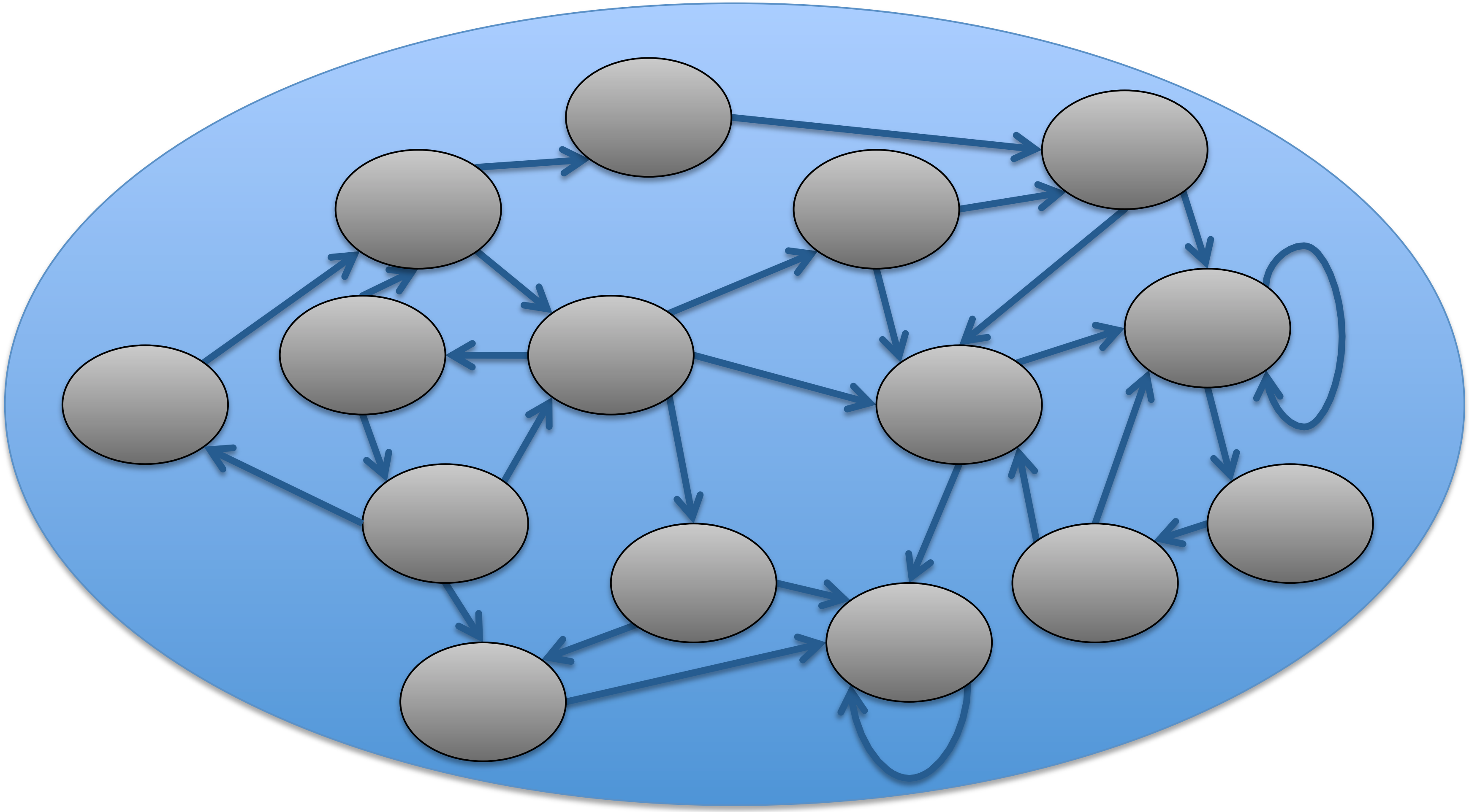
David N. Jansen

Let’s play a game

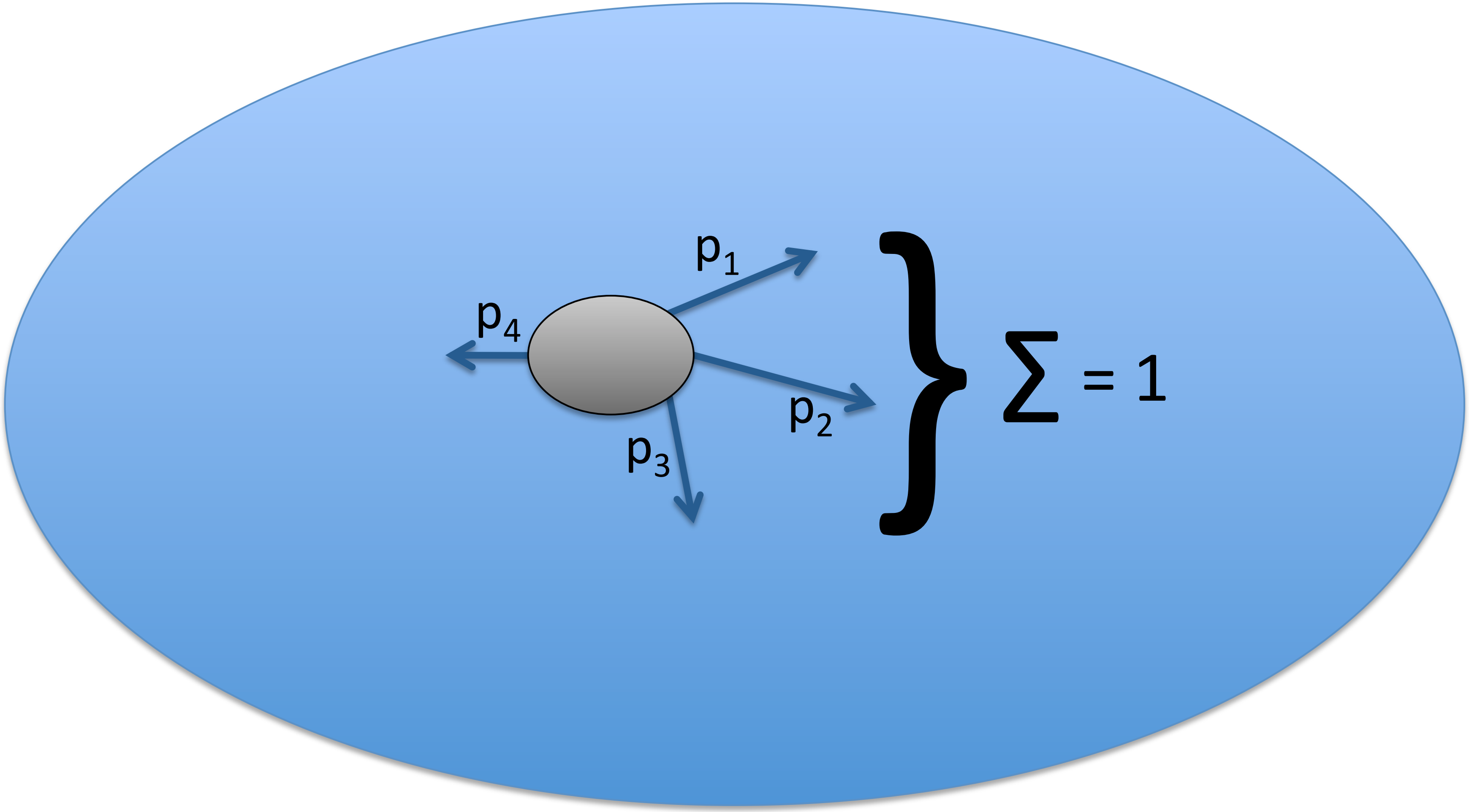
States



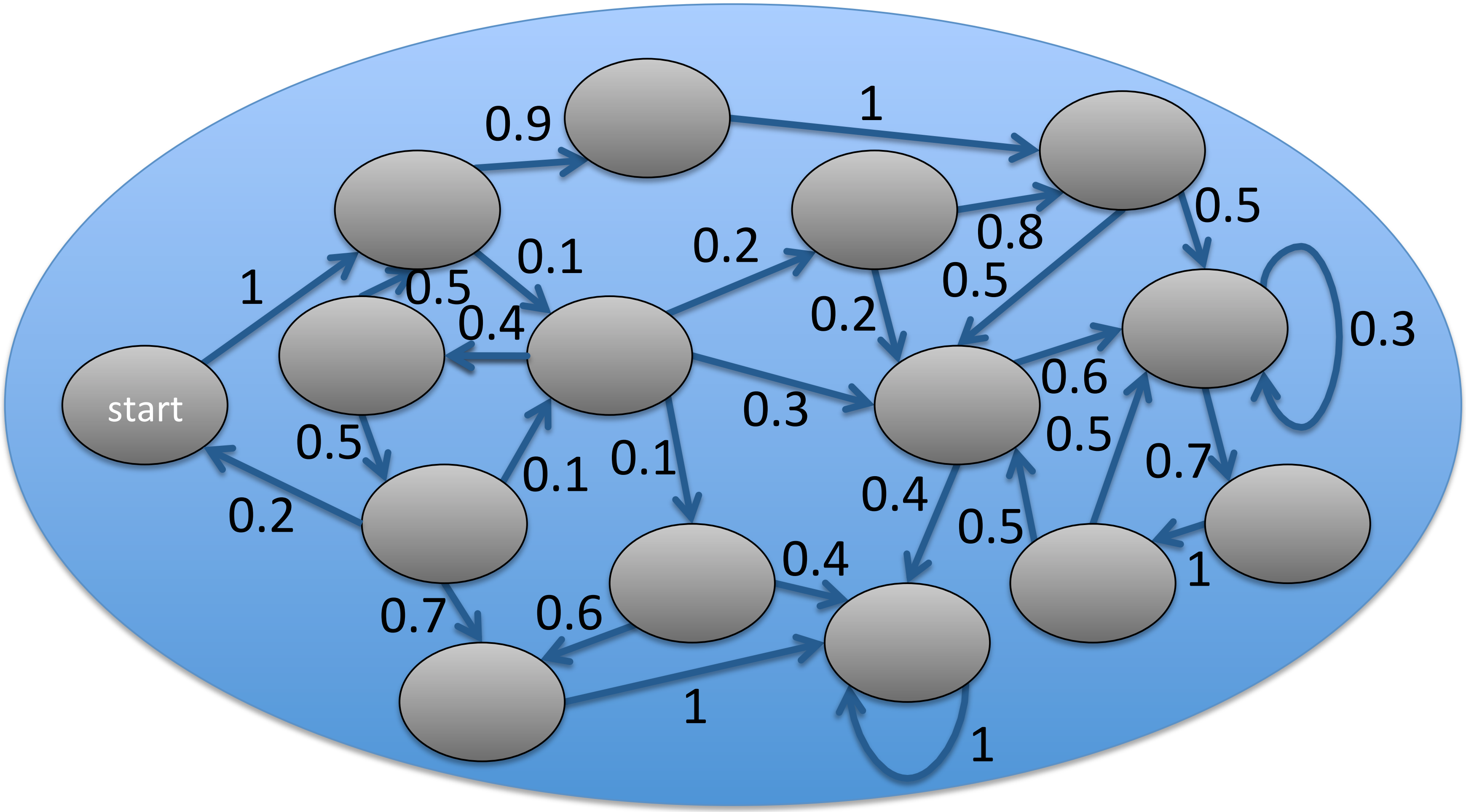
Transitions



Transition Probabilities



Markov Chain



Formal definition

- A Markov chain consists of:
  - S — finite set of states (often $S = \{1, 2, \ldots n\}$)
  - $\mathbf{P}: S \times S \rightarrow [0,1]$ — transition probability matrix (with row sums = 1)
  - $\pi_0: S \rightarrow [0,1]$ — initial state distribution (sometimes)

Semantics of a Markov Chain

- similar to Kripke structure:
  - system starts in one of the initial states, chosen according to π_0
  - system is always in a state
  - from time to time, a transition is taken
  - when the system leaves state *i*, the next state is *j* with probability $\mathbf{P}(i,j)$

Semantics of a Markov Chain

- What is “from time to time”?
  Two interpretations:
  - we do not measure / we consider uninteresting how long the time between transitions is
  - all times between transitions are equal

Some general properties of Markov chains

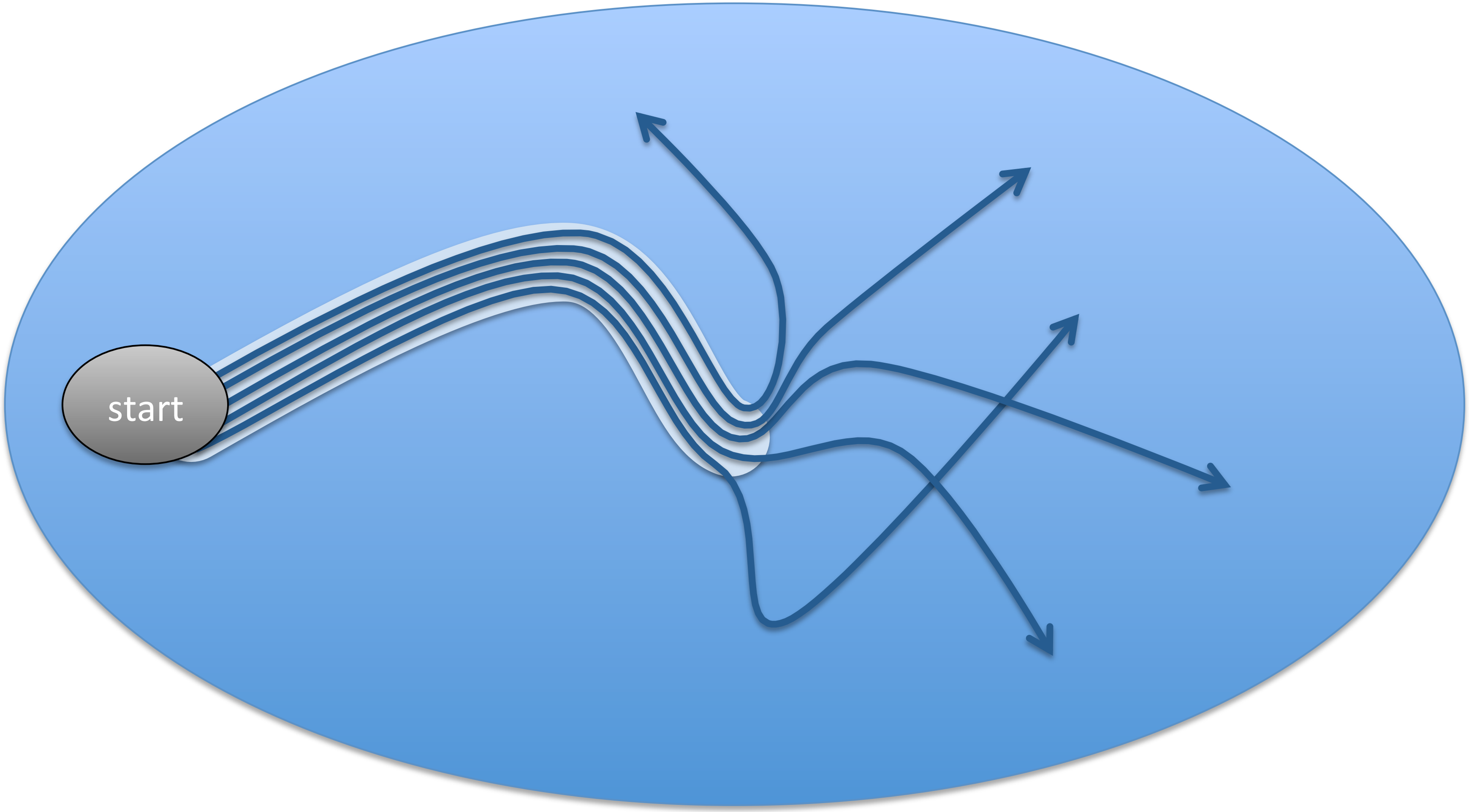
- "Markov" property
- Chapman–Kolmogorov equations

Example: Gambler's ruin

- a gambler plays a game repeatedly
- each time,
  he either wins €1 (with probability p)
  or he loses €1 (with probability $1–p$)
- gambler plays until he is bankrupt

- Draw this Markov chain!

Cylinder Set of a Markov Chain



Probability Space of a Markov Chain

- Ω all paths
- $\mathcal{F}$ σ-algebra generated by cylinder sets
  - $Cyl(s_0, s_1, \ldots, s_n)$:= paths starting with $s_0, s_1, \ldots, s_n$
  - complements and unions of cylinder sets
- μ unique extension of

  $\mu(Cyl(s_0, s_1, \ldots, s_n))$

  $= \pi_0(s_0) \cdot \mathbf{P}(s_0, s_1) \cdot \mathbf{P}(s_1, s_2) \cdots \mathbf{P}(s_{n-1}, s_n)$

Analysis of a Markov chain

- Interesting measures:
  - transient state distribution:
    What is the probability to be in state i
    after t transitions?
    Notation: $p_i(t)$ and $\pi_t = (p_1(t), p_2(t), ..., p_n(t))$
  - steady-state distribution:
    What is the probability to be in state i
    in equilibrium / after a long time ($t \rightarrow \infty$)?
    Notation: p_s and $\pi = (p_1, p_2, ..., p_n)$

transient state distribution

- Given:
  - initial distribution π_0
  - transition probabilities **P** = **P(1)**
- Requested:
  - transient probabilities $\pi_t = (p_1(t), p_2(t), \ldots, p_n(t))$

- Calculate: $\pi_t = \pi_0 \cdot \mathbf{P}(t) = \pi_0 \cdot \mathbf{P}^t$

Example: Weather on the Island of Hope

Each day, the weather is classified as sunny, cloudy or rainy.
Tomorrow's weather forecast is, depending on today:

- If the present day is sunny, then the next day will be
  - sunny with probability 0.7
  - cloudy with probability 0.1
  - rainy with probability 0.2
- If the present day is cloudy, the values are 0.5, 0.25 and 0.25.
- If the present day is rainy, the values are 0.4, 0.3 and 0.3.

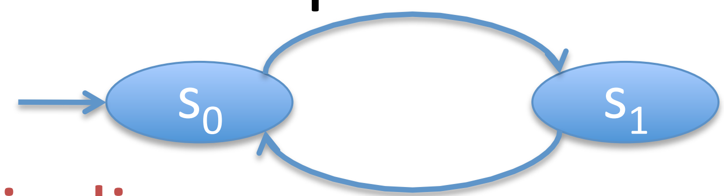
Can you give a multi-day weather forecast?

steady-state distribution

- Given:
  - initial distribution π_0
  - transition probabilities $\mathbf{P} = \mathbf{P}(1)$
- Requested:
  - steady-state distribution π

some properties of Markov chains

- A Markov chain is irreducible if every state is reachable from every other state.
- A state *s* of a Markov chain is periodic with period *k* if any return to *s* occurs in some multiple of *k* steps.



- A Markov chain is aperiodic if all its states have period 1.

steady-state distribution

Theorem: If a finite Markov chain is irreducible and aperiodic, it has a steady-state distribution that does not depend on the initial distribution.

- What if MC is reducible?
  → need to find irreducible parts
  and probability to reach them
- What if MC is periodic?
  → no steady-state distribution at all

Proof of Theorem

- There exists an invariant distribution:
  - The row sums of **P** are all 1.
  - The column vector $(1, ..., 1)^T$ is a right eigenvector for eigenvalue 1: $\mathbf{P}\cdot(1, ..., 1)^T = (1, ..., 1)^T$
  - Therefore there exists a left eigenvector π for eigenvalue 1, i. e. $\pi\cdot\mathbf{P} = \pi$
  - Now choose a π that sums to 1;
    this is an invariant distribution.

Proof of Theorem

- It is independent from initial state:
  [Norris: Markov chains. Thm. 1.8.3]
  - Assume there are two Markov chains X, Y
    with the same transition matrix but different initial states
  - combine the chains: W := X x Y
  - Pick a fixed state b.
    W will visit (b,b) in finite time with probability 1 .
  - Z := the behaviour of X until W first visits (b,b) and the behaviour of Y afterwards
  - Z has same transition matrix + initial state as X
    and same steady-state distribution as Y

steady-state distribution

- Given:
  - MC is irreducible and aperiodic
  - initial distribution π_0
  - transition probabilities $\mathbf{P} = \mathbf{P}(1)$
- Requested:
  - steady-state distribution π
- Calculate: $\pi = \pi \cdot \mathbf{P}$

$$\sum_{s \in S} \pi(s) = 1$$

Long-term weather means on the Island of Hope

- Weather model on the Island of Hope is finite, aperiodic and irreducible.
  So steady-state distribution exists!
- This distribution shows
  the long-term weather means.

- Calculate it!

Example: Google PageRank

Model the WWW as a Markov chain as follows:

- Each webpage is a state of the MC.
- Each hyperlink is a transition.
  In each state, the hyperlinks from that state have the same probability.

The steady-state probability of a state corresponds to its PageRank.

Recapitulation

- Markov chains describe
  the behaviour of discrete-state systems.
- Discrete-time Markov chains
  count the number of steps,
  but not how long a step takes.
- Transient state and steady-state analysis
  serve to calculate state probabilities.