Modeling and Verification of Probabilistic Systems

Joost-Pieter Katoen

Lehrstuhl für Informatik 2
Software Modeling and Verification Group

http://moves.rwth-aachen.de/teaching/ws-1819/movep18/

October 9, 2018

Overview

Geometric distribution

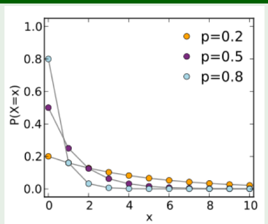
Geometric distribution

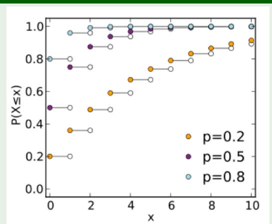
Let X be a discrete random variable, natural $k > 0$ and $0 < p \leqslant 1$. The mass function of a *geometric distribution* is given by:

$$Pr\{ X = k \} = (1-p)^{k-1} \cdot p$$

We have $E[X] = \frac{1}{p}$ and $Var[X] = \frac{1-p}{p^2}$ and cdf $Pr\{ X \leqslant k \} = 1 - (1-p)^k$.

Geometric distributions and their cdf's





Memoryless property

Theorem

1. For any random variable X with a geometric distribution:

$$Pr\{X = k + m \mid X > m\} \; = \; Pr\{X = k\} \quad \text{for any} \quad m \in T, k \geqslant 1$$

This is called the memoryless property, and X is a memoryless r.v..

2. Any discrete random variable which is memoryless is geometrically distributed.

Proof:

Exercise.

Andrei Andrejewitsch Markow

Markov property

The conditional probability distribution of future states of a Markov process only depends on the current state and not on its further history.

Markov process

A discrete-time stochastic process $\{ X(t) \mid t \in T \}$ over state space $\{ d_0, d_1, \ldots \}$ is a *Markov process* if for any $t_0 < t_1 < \ldots < t_n < t_{n+1}$:

$$Pr\{ X(t_{n+1}) = d_{n+1} \mid X(t_0) = d_0, X(t_1) = d_1, \ldots, X(t_n) = d_n \}$$
$$=$$
$$Pr\{ X(t_{n+1}) = d_{n+1} \mid X(t_n) = d_n \}$$

The distribution of $X(t_{n+1})$, given the values $X(t_0)$ through $X(t_n)$, only depends on the current state $X(t_n)$.

Invariance to time-shifts

Time homogeneity

Markov process $\{ X(t) \mid t \in T \}$ is *time-homogeneous* iff for any $t' < t$:

$$Pr\{ X(t) = d \mid X(t') = d' \} \;=\; Pr\{ X(t - t') = d \mid X(0) = d' \}.$$

A time-homogeneous stochastic process is invariant to time shifts.

Discrete-time Markov chain

A *discrete-time Markov chain* (DTMC) is a time-homogeneous Markov process with discrete parameter T and discrete state space.

Discrete-time Markov chain

Discrete-time Markov chain

A *discrete-time Markov chain* (DTMC) is a time-homogeneous Markov process with discrete parameter T and discrete state space S.

Transition probabilities

The *(one-step) transition probability* from $s \in S$ to $s' \in S$ at epoch $n \in \mathbb{N}$ is given by:

$$p^{(n)}(s, s') \;=\; Pr\{ X_{n+1} = s' \mid X_n = s \} \;=\; Pr\{ X_1 = s' \mid X_0 = s \}$$

where the last equality is due to time-homogeneity.

Since $p^{(n)}(\cdot) = p^{(k)}(\cdot)$, the superscript (n) is omitted, and we write $p(\cdot)$.

Transition probability matrix

Discrete-time Markov chain

A *discrete-time Markov chain* (DTMC) is a time-homogeneous Markov process with discrete parameter T and discrete state space S.

Transition probability matrix

Let $\mathbf{P}$ be a function with $\mathbf{P}(s_i, s_j) = p(s_i, s_j)$. For finite state space S, function $\mathbf{P}$ is called the *transition probability matrix* of the DTMC with state space S.

Properties

1. $\mathbf{P}$ is a (right) *stochastic* matrix, i.e., it is a square matrix, all its elements are in $[0, 1]$, and each row sum equals one.
2. $\mathbf{P}$ has an eigenvalue of one, and all its eigenvalues are at most one.
3. For all $n \in \mathbb{N}$, $\mathbf{P}^n$ is a stochastic matrix.

DTMCs — A transition system perspective

Discrete-time Markov chain

A DTMC $\mathcal{D}$ is a tuple $(S, \mathbf{P}, \iota_{\text{init}}, AP, L)$ with:

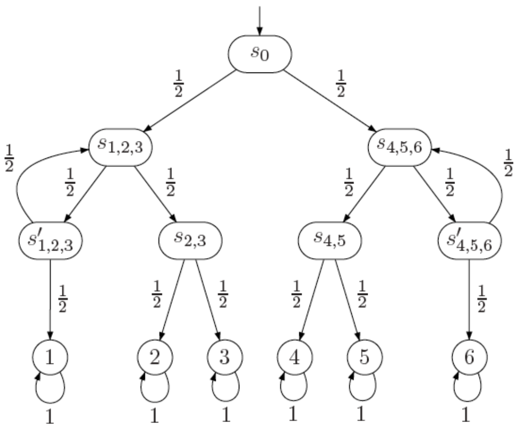
- S is a countable nonempty set of states
- $\mathbf{P} : S \times S \to [0, 1]$, transition probability function s.t. $\sum_{s'} \mathbf{P}(s, s') = 1$
- $\iota_{\text{init}} : S \to [0, 1]$, the initial distribution with $\sum_{s \in S} \iota_{\text{init}}(s) = 1$
- AP is a set of atomic propositions.
- $L : S \to 2^{AP}$, the labeling function, assigning to state s, the set $L(s)$ of atomic propositions that are valid in s.

Initial states

- $\iota_{\text{init}}(s)$ is the probability that DTMC $\mathcal{D}$ starts in state s
- the set $\{ s \in S \mid \iota_{\text{init}}(s) > 0 \}$ are the possible initial states.

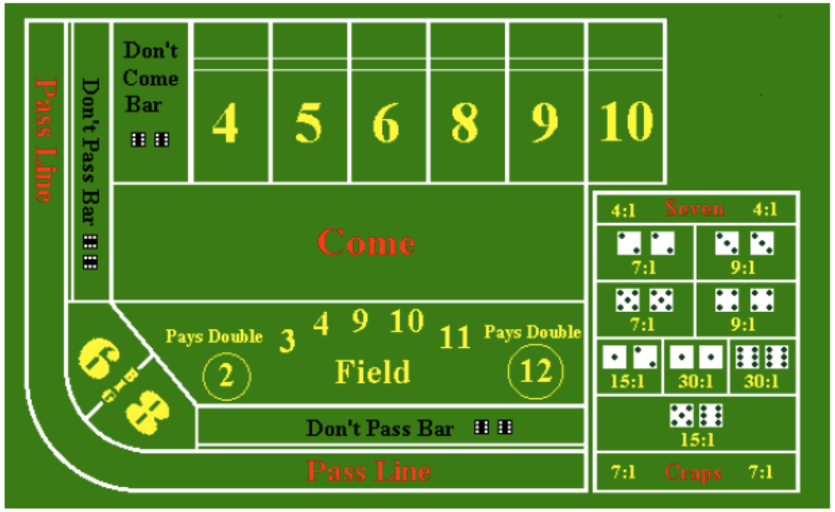
Example: roulette in Monte Carlo, 1913

Simulating a die by a fair coin [Knuth & Yao]



Heads = "go left"; tails = "go right". Does this DTMC adequately model a fair six-sided die?

Craps

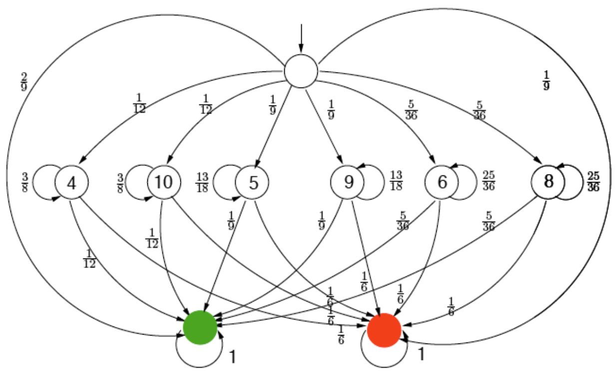


Craps

- Roll two dice and bet
- Come-out roll (“pass line” wager):
  - outcome 7 or 11: win
  - outcome 2, 3, or 12: lose (“craps”)
  - any other outcome: roll again (outcome is “point”)
- Repeat until 7 or the “point” is thrown:
  - outcome 7: lose (“seven-out”)
  - outcome the point: win
  - any other outcome: roll again

A DTMC model of Craps

- Come-out roll:
  - 7 or 11: win
  - 2, 3, or 12: lose
  - else: roll again
- Next roll(s):
  - 7: lose
  - point: win
  - else: roll again



Overview

State residence time distribution

Let T_s be the number of epochs of DTMC $\mathcal{D}$ to stay in state s:

$$\begin{aligned} Pr\{\, T_s = 1 \,\} &= 1 - \mathbf{P}(s,s) \\ Pr\{\, T_s = 2 \,\} &= \mathbf{P}(s,s) \cdot (1 - \mathbf{P}(s,s)) \\ \ldots\ldots & \ldots \ \ldots\ldots \\ Pr\{\, T_s = n \,\} &= \mathbf{P}(s,s)^{n-1} \cdot (1 - \mathbf{P}(s,s)) \end{aligned}$$

So, the state residence times in a DTMC obey a *geometric* distribution.

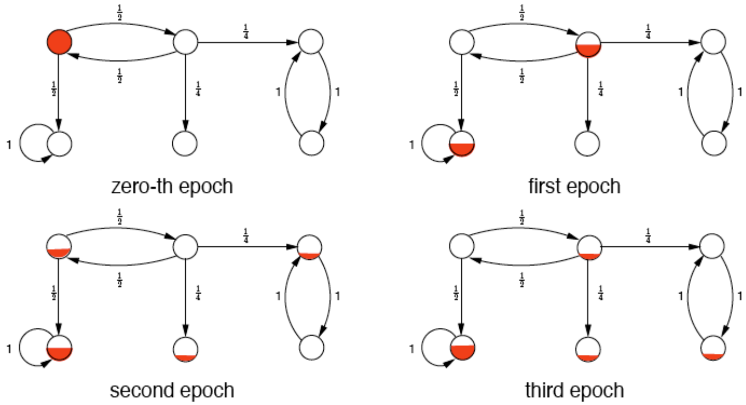
The expected number of time steps to stay in state s equals $E[T_s] = \frac{1}{1-\mathbf{P}(s,s)}$.

The variance of the residence time distribution is $Var[T_s] = \frac{\mathbf{P}(s,s)}{(1-\mathbf{P}(s,s))^2}$.

Recall: the geometric distribution is the only discrete probability distribution that is memoryless.

Overview

Evolution of an example DTMC



We want to determine $p_{s,s'}(n) = Pr\{\, X(n) = s' \mid X(0) = s \,\}$ for $n \in \mathbb{N}$.

Determining n-step transition probabilities

n-step transition probabilities

The probability to move from s to s' in $n \in \mathbb{N}$ steps is inductively defined:

$$p_{s,s'}(0) = 1 \quad \text{if } s = s', \quad \text{and 0 otherwise,}$$

$p_{s,s'}(1) = \mathbf{P}(s, s')$, and for $n > 1$ by the Chapman-Kolmogorov equation:

$$p_{s,s'}(n) = \sum_{s''} p_{s,s''}(l) \cdot p_{s'',s'}(n-l) \qquad \text{for some } 0 < l < n$$

Proof: see black board.

For $l = 1$ and $n > 0$ we obtain: $p_{s,s'}(n) = \sum_{s''} p_{s,s''}(1) \cdot p_{s'',s'}(n-1)$

$\mathbf{P}^{(n)} = \mathbf{P}^{(1)} \cdot \mathbf{P}^{(n-1)} = \mathbf{P} \cdot \mathbf{P}^{(n-1)}$ is the n-step transition probability matrix

Repeating this scheme: $\mathbf{P}^{(n)} = \mathbf{P} \cdot \mathbf{P}^{(n-1)} = \ldots = \mathbf{P}^{n-1} \cdot \mathbf{P}^{(1)} = \mathbf{P}^n$.

Transient probability distribution

Transient distribution

$\mathbf{P}^n(s,t)$ equals the probability of being in state t after n steps given that the computation starts in s.

The probability of DTMC $\mathcal{D}$ being in state t after exactly n transitions is:

$$\Theta_n^{\mathcal{D}}(t) \;=\; \sum_{s \in S} \iota_{\mathrm{init}}(s) \cdot \mathbf{P}^n(s,t)$$

$\Theta_n^{\mathcal{D}}(t)$ is called the *transient state probability* at epoch n for state t. The function $\Theta_n^{\mathcal{D}}$ is the *transient state distribution* at epoch n of DTMC $\mathcal{D}$.

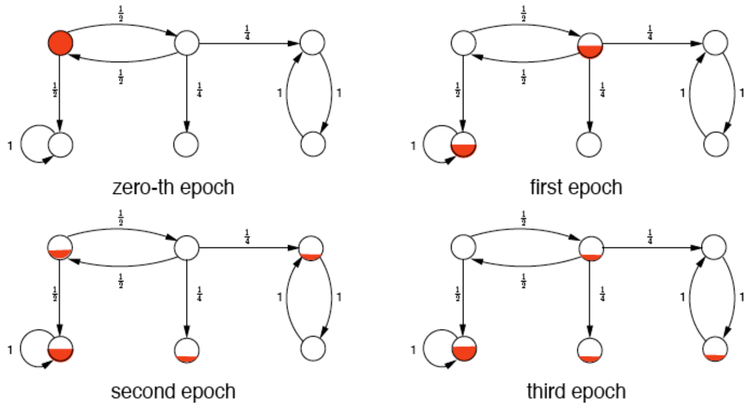
When considering $\Theta_n^{\mathcal{D}}$ as vector $(\Theta_n^{\mathcal{D}})_{t \in S}$ we have:

$$\Theta_n^{\mathcal{D}} \;=\; \iota_{\mathrm{init}} \cdot \underbrace{\mathbf{P} \cdot \mathbf{P} \cdot \ldots \cdot \mathbf{P}}_{n \text{ times}} \;=\; \iota_{\mathrm{init}} \cdot \mathbf{P}^n.$$

Transient probability distribution: example

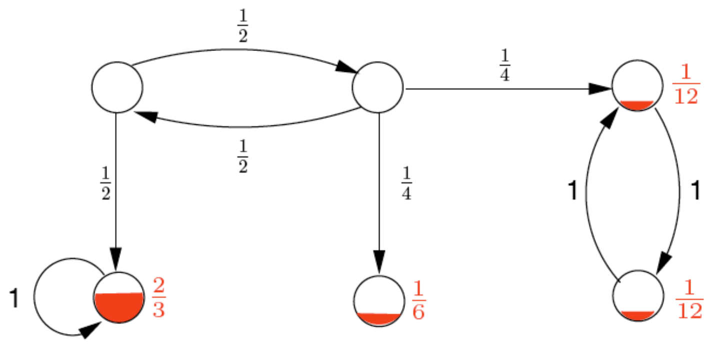
Overview

Evolution of an example DTMC



We want to determine the probability to be in a state on the long run.

On the long run



The probability mass on the long run is only left in bottom SCCs.

Limiting distribution

Ergodic stochastic matrix

Stochastic matrix $\mathbf{P}$ is called *ergodic* if:

$$\mathbf{P}^{\infty} \;=\; \lim_{n\to\infty} \mathbf{P}^{n} \quad \text{exists and has identical rows}$$

Ergodicity theorem

If the transition probability matrix $\mathbf{P}$ of a DTMC is ergodic, then:

1. $\underline{p}(n)$ converges to a limiting distribution $\underline{v}$ independent from $\underline{p}(0)$
2. each row of $\mathbf{P}^{\infty}$ equals the limiting distribution

Proof.

$$\lim_{n\to\infty} \underline{p}(0)\cdot\mathbf{P}^{n} = \underline{p}(0)\cdot \underbrace{\lim_{n\to\infty}\mathbf{P}^{n}}_{\mathbf{P}^{\infty}} = \underline{p}(0)\cdot \begin{pmatrix} v_{s_0} & \dots & v_{s_n} \\ \dots & \dots & \dots \\ v_{s_0} & \dots & v_{s_n} \end{pmatrix} = \underline{v} \qquad \square$$

Limiting distribution

- We also have:

$$\underline{v} = \lim_{n\to\infty} \underline{p}(n{+}1) = \lim_{n\to\infty} \underline{p}(0)\cdot\mathbf{P}^{n+1} = \left(\lim_{n\to\infty} \underline{p}(0)\cdot\mathbf{P}^{n}\right)\cdot\mathbf{P} = \underline{v}\cdot\mathbf{P}$$

- Thus, limiting probabilities can be obtained by solving the (homogeneous) system of linear equations:

$$\boxed{\underline{v} = \underline{v}\cdot\mathbf{P} \quad \text{or} \quad \underline{v}\cdot(\mathbf{I}-\mathbf{P}) = \underline{0} \quad \text{under } \textstyle\sum_i \underline{v}(i) = 1}$$

  - vector $\underline{v}$ is the left Eigenvector of $\mathbf{P}$ with Eigenvalue 1
- $\underline{v}$ is called the *limiting* state-probability vector

Two interpretations of $\underline{v}(s)$:

- the long-run proportion of time that the DTMC “spends” in state s
- the probability the DTMC is in s when making a snapshot after a very long time

Examples

Summary

What are Markov chains?

- A discrete-time Markov chain (DTMC) is a time-homogeneous Markov process with discrete parameter T and discrete state space S.
- State residence times are geometrically distributed.
- Alternative: a DTMC $\mathcal{D}$ is a tuple $(S, \mathbf{P}, \iota_{\mathrm{init}}, AP, L)$

What are transient probabilities?

- $\Theta_n^{\mathcal{D}}(s)$ is the probability to be in state s after n steps.
- These transient probabilities satisfy: $\Theta_n^{\mathcal{D}} \; = \; \iota_{\mathrm{init}} \cdot \mathbf{P}^n$.

What are long-run probabilities?

- $\underline{v}(s)$ is the probability to be in state s after infinitely many steps.
- long-run probabilities satisfy: $\underline{v} \cdot (\mathbf{I} - \mathbf{P}) = \underline{0}$ under $\sum_i \underline{v}(i) = 1$.