Algorithms for NP-Hard Problems

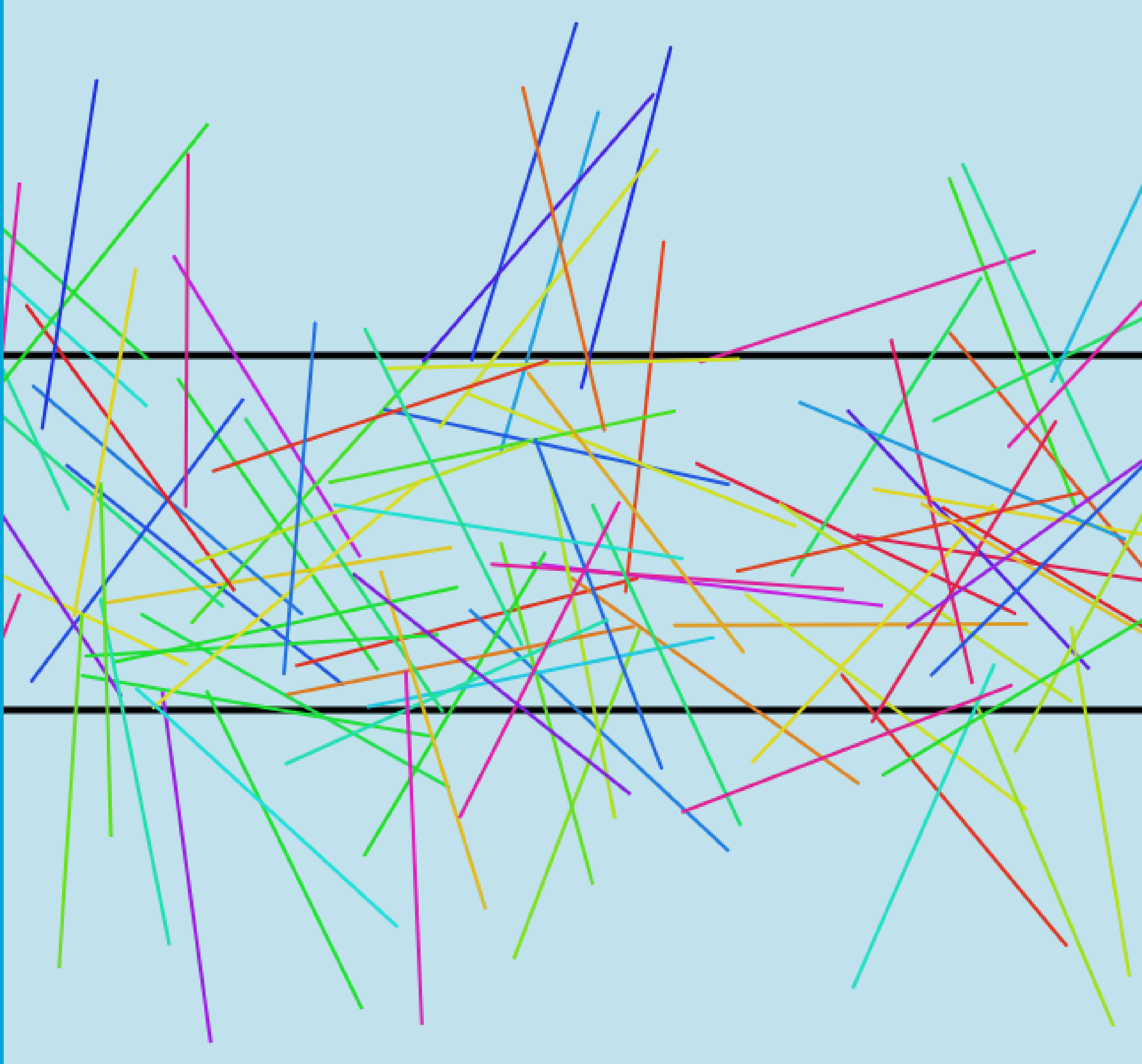
Part III: Simulation-Based Approximation

Lecture 1: Estimating Uncertainty

Instructor: Dr. Anna Lukina

a.lukina@tudelft.nl

TU Delft



We’re done with

Part I: Heuristic Search

Part II: Modeling and Exhaustive Search

Now we’ll do

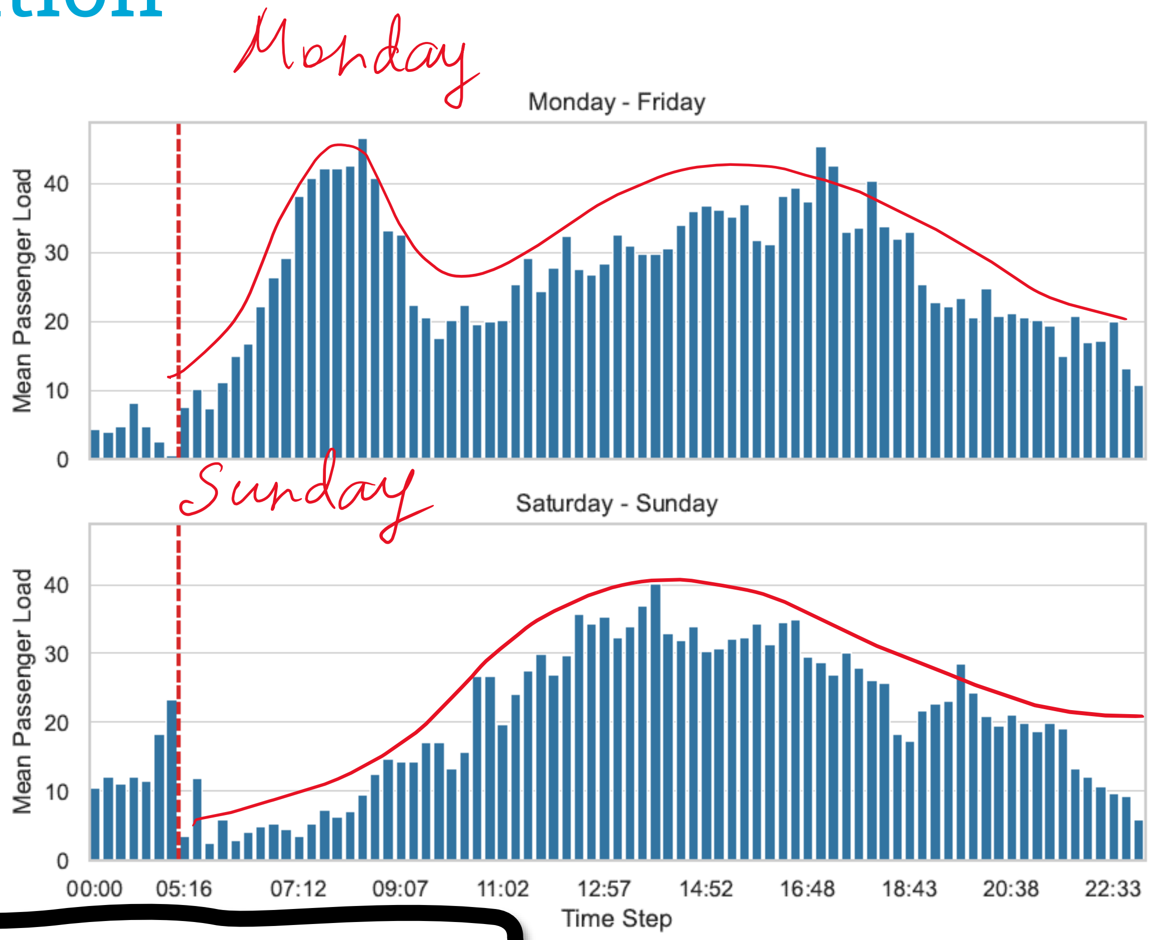
Part III: Simulation-Based Methods:

- Estimating uncertainty

WHERE IS UNCERTAINTY?

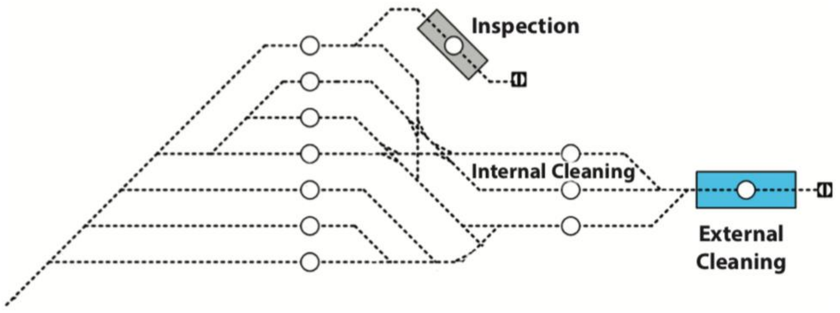
Public Transport Optimization

Pictures taken from MSc thesis of D. van Gelder, TU Delft 2021



UNKNOWN PASSENGER LOAD!

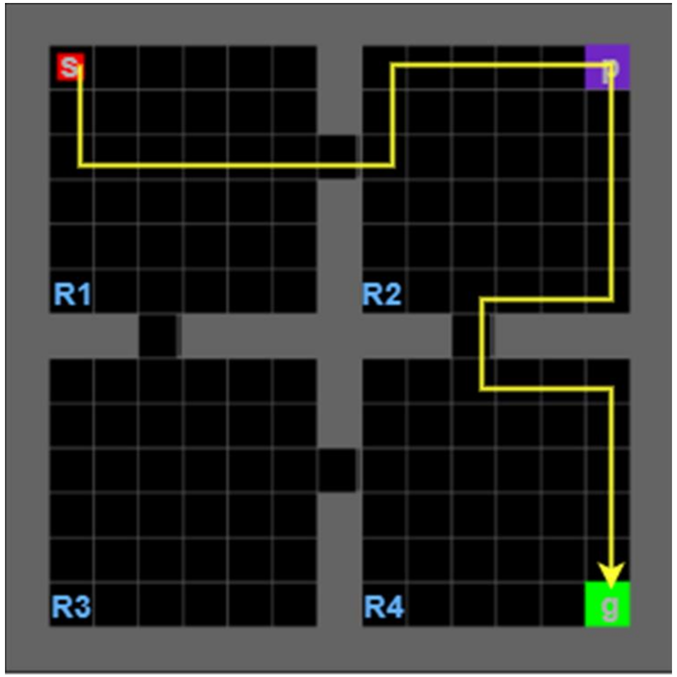
Train Shunting

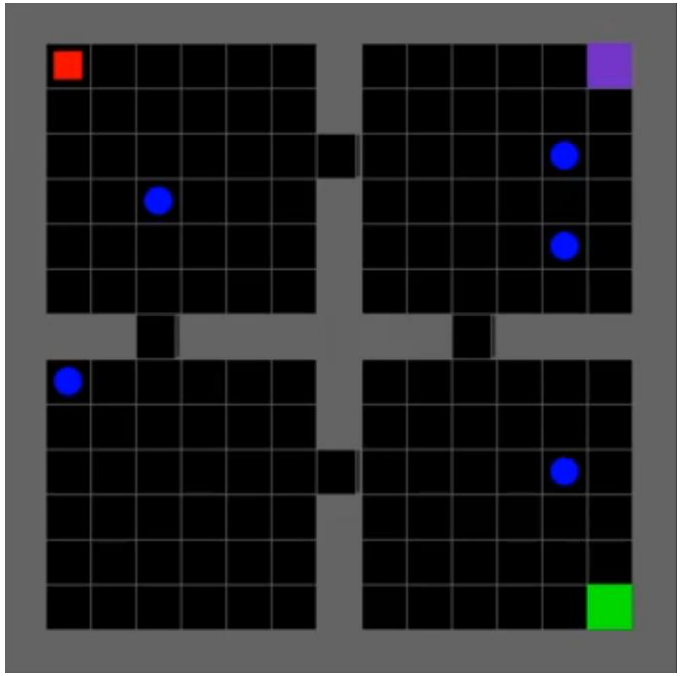


Picture taken from the Dutch Railways and ProRail

UNKNOWN DELAY TIME!

Robot Planning





Pictures taken from MSc thesis by C. van Rijn, TU Delft 2023

UNKNOWN OBSTACLES!

Part III: Simulation-Based Approximation

Lecture 1: Estimating Uncertainty

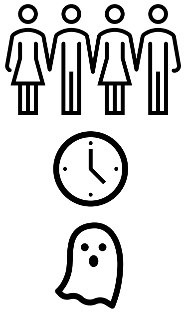
Part III: Simulation-Based Approximation

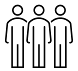
Lecture 1: Estimating Uncertainty

1. What and why?
   i. How to model uncertainty
   ii. Long before Monte-Carlo method
   iii. Essential theory
2. Monte-Carlo method
   i. General method and complexity
   ii. Central Limit Theorem

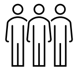
Random Variables

- A random variable is some aspect of the world about which we (may) have uncertainty
  - C = How crowded is it on the train?
  - T = How much is the train delayed?
  - O = Where are the moving obstacles?



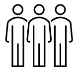
- Random variables have domains
  - C in domain {  ,  }
  - T in domain $[0, \infty)$
  - O in domain $\{(0,0), (0,1), (1,0), \ldots\}$

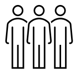
Probability Model

- Associate probability with each possible value of the random variable
  - Crowdedness: C in domain $\{$  ,  $\}$

$\xi = \xi(\omega)$

$$X \begin{cases} x = 50 \\ x = 500 \end{cases}$$

C	$\mathrm{P}(C)$
	0.2
	0.8

$\omega \in \Omega = \{$  ,  $\}$

The set of numbers $\{\mathrm{P}_\xi(x)\}$ is called the (discrete) probability distribution of the random variable ξ, where $\mathrm{P}_\xi(x) = \mathrm{P}\{\,\xi = x, \;\; x \in X\}$.

Must hold: $\forall x \in X \;\; \mathrm{P}(\xi = x) \geq 0$ and $\sum_{x \in X} \mathrm{P}(\xi = x) = 1$

For now, we assume a discrete, countable sample space Ω.

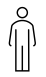
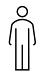
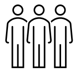
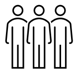
Joint Probability Distribution

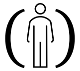
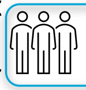
- Specify a real number for each assignment (outcome) over a set of random variables $\xi_1, \xi_2, \dots, \xi_n$

Must hold:

$$P(\xi_1 = x_1, \xi_2 = x_2, \dots, \xi_n = x_n) \geq 0$$

$$\sum_{(x_1, x_2, \dots, x_n)} P(x_1, x_2, \dots, x_3) = 1$$

C	T	$P(C, T)$
	$x_2 < 10\ min$	0.1
	$10\ min \leq x_2$	0.1
	$x_2 < 10\ min$	0.4
	$10\ min \leq x_2$	0.4

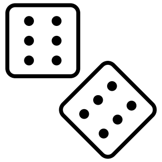
- Marginalization
  - P()?
  - $P(10\ min \leq x_2)$?
  - P( , $x_2 < 10\ min$)?

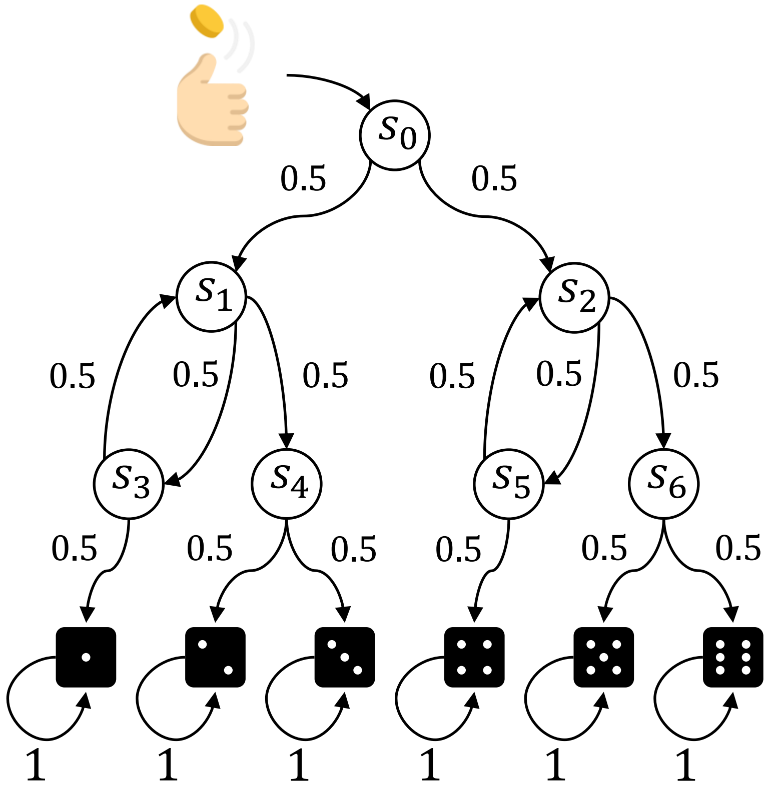
An event A

“Probability provides a way of summarizing the uncertainty that comes from our laziness and ignorance”

Russell & Norvig: Artificial Intelligence. A Modern Approach, 4th Global Edition, 2022.

Simulating a Dice (with only fair coins)





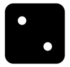
Joost-Pieter Katoen. 2016. The Probabilistic Model Checking Landscape. In Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science (LICS '16).

[Knuth & Yao, 1976]

Heads = "go left"; tails = "go right".

Does this model a six-sided dice?

Reachability Probabilities (Knuth-Yao's Die):

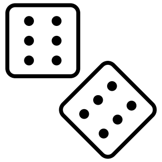
- Consider the event $A =$ 
- We model: $\xi = I_A(\omega)$,

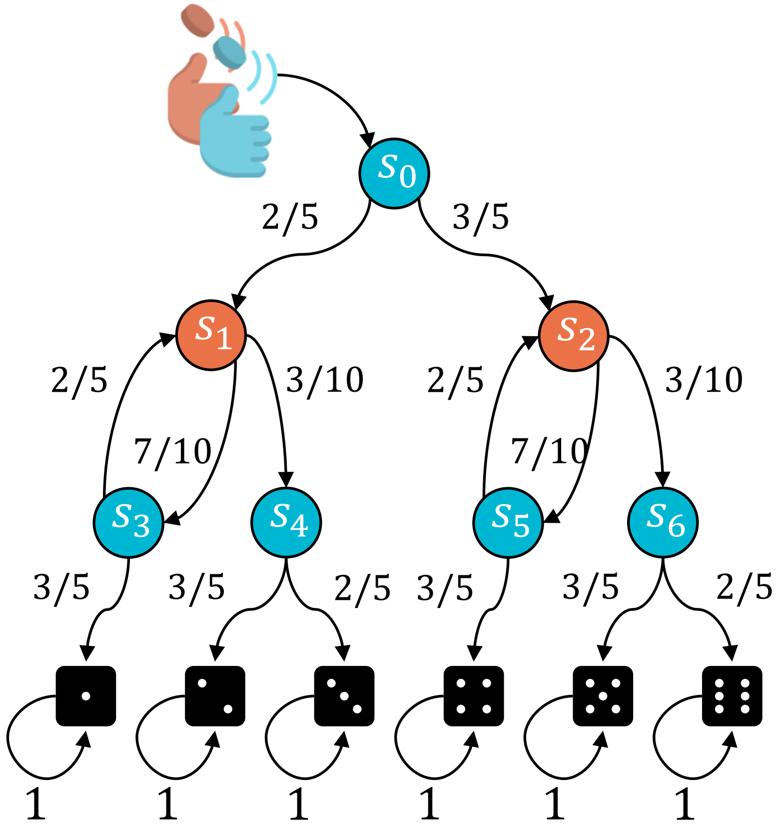
where $I_A(\omega) = \begin{cases} 1, \omega \in A, \\ 0, \omega \notin A. \end{cases}$

What is the probability of reaching A from s_0?

$$p_{s_4} = \frac{1}{2}, p_{s_1} = \frac{1}{3}, p_{s_3} = \frac{1}{6}, \text{and } p_{s_0} = \frac{1}{6}.$$

Simulating a Dice (with two unfair coins)



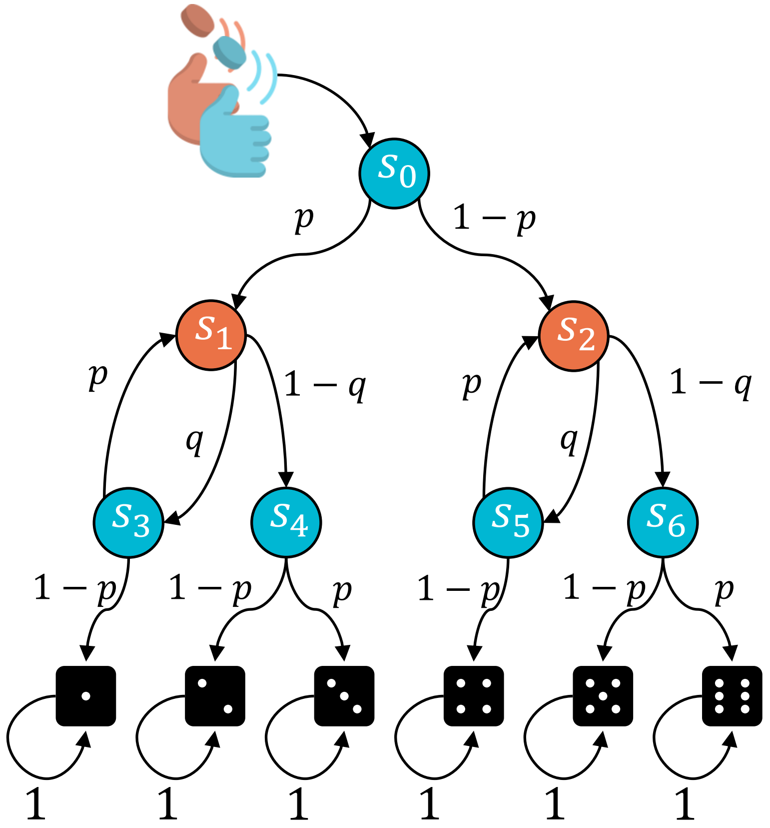


Is the probability of reaching ⚁ $> \frac{3}{20}$?

No! It is $\frac{1}{10}$.

Sebastian Junges, Joost-Pieter Katoen, Guillermo A. Pérez, and Tobias Winkler, "The complexity of reachability in parametric Markov decision processes," *J. Comput. Syst. Sci.*, 2021.

Parametric Reachability Probability Problem



What should p and q be, so that the probability of reaching is $> \frac{3}{20}$?

This problem is NP-complete!

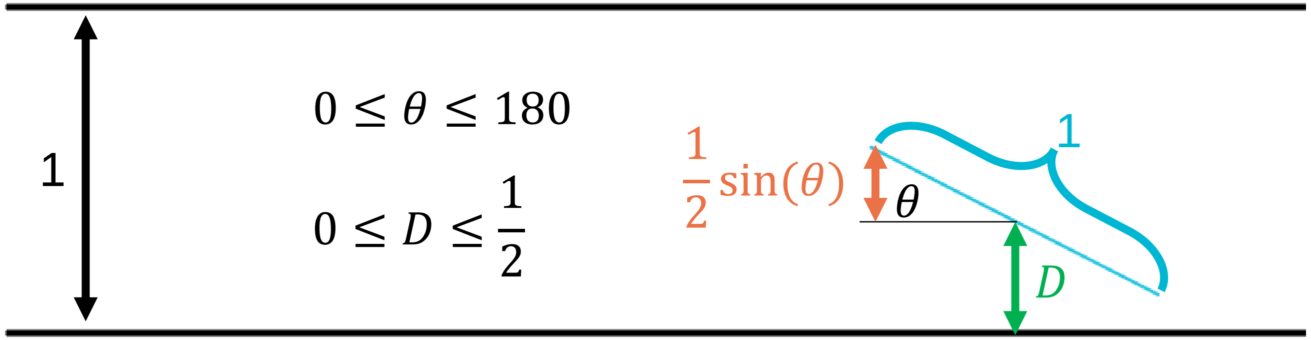
Sebastian Junges, Joost-Pieter Katoen, Guillermo A. Pérez, and Tobias Winkler, “The complexity of reachability in parametric Markov decision processes,” *J. Comput. Syst. Sci.*, 2021.

Nils Jansen, Sebastian Junges, and Joost-Pieter Katoen, “Parameter Synthesis in Markov Models: A Gentle Survey,” in *Principles of Systems Design*, 2022.

“Monte Carlo is the art of approximating an expectation by the sample mean of a function of simulated random variables.”

Yu A Shreider. The Monte Carlo method: the method of statistical trials, volume 87. Elsevier, 2014.

Long Before Monte-Carlo Method



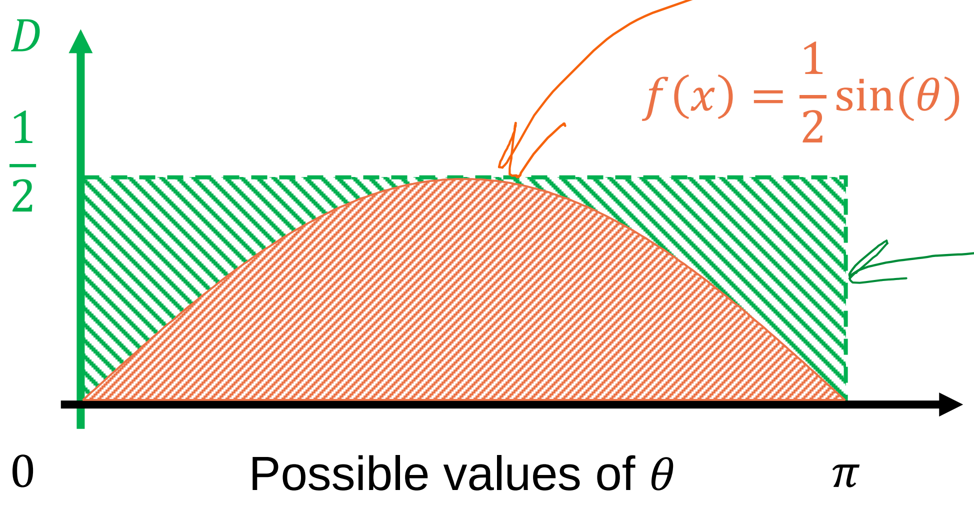
Buffon's needle (1777):

A needle of a unit length falls on two lines a unit distance apart.

- The needle misses the line.
- The needle will hit the line if $D \leq \frac{1}{2}\sin(\theta)$.
- How often will this occur?

https://mste.illinois.edu/activity/buffon/

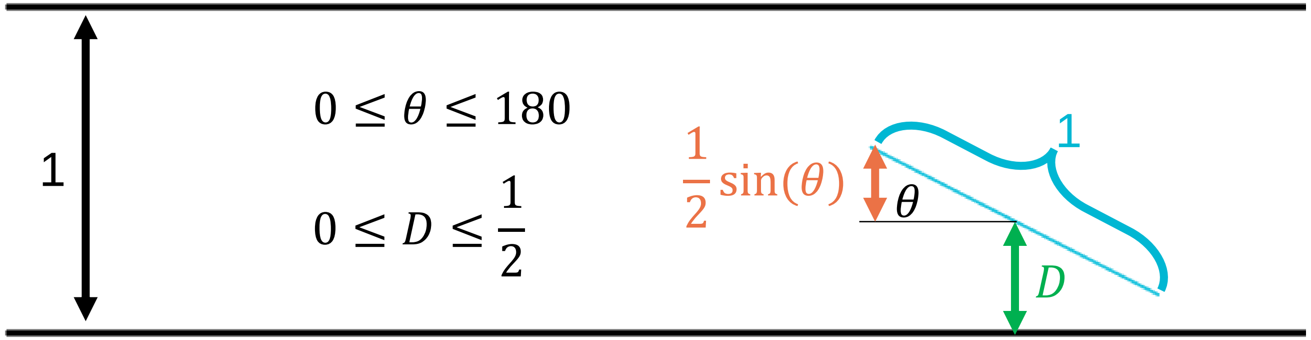
Long Before Monte-Carlo Method



$$\int_0^{\pi} f(x)dx = 1$$

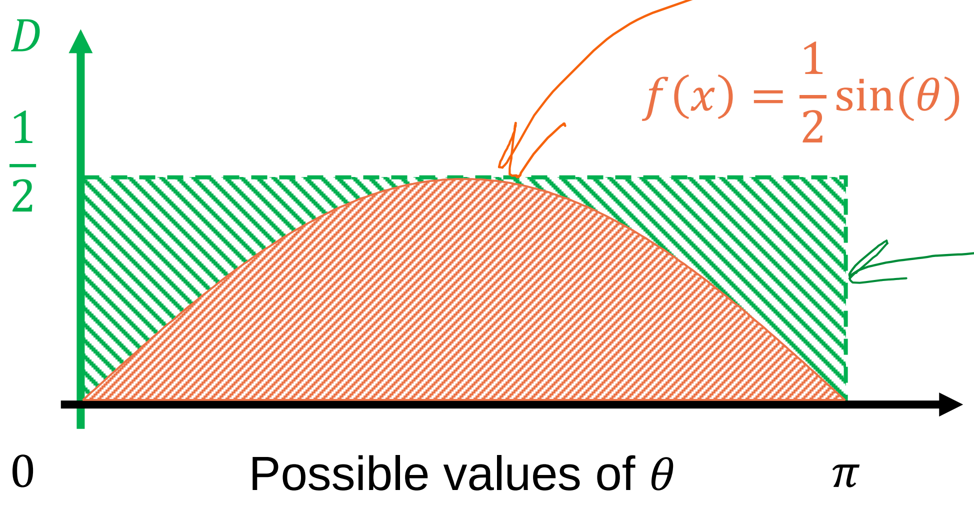
$$\frac{\pi}{2}$$

$$\mathrm{P}(hit) = \frac{2}{\pi}$$



https://mste.illinois.edu/activity/buffon/

Long Before Monte-Carlo Method

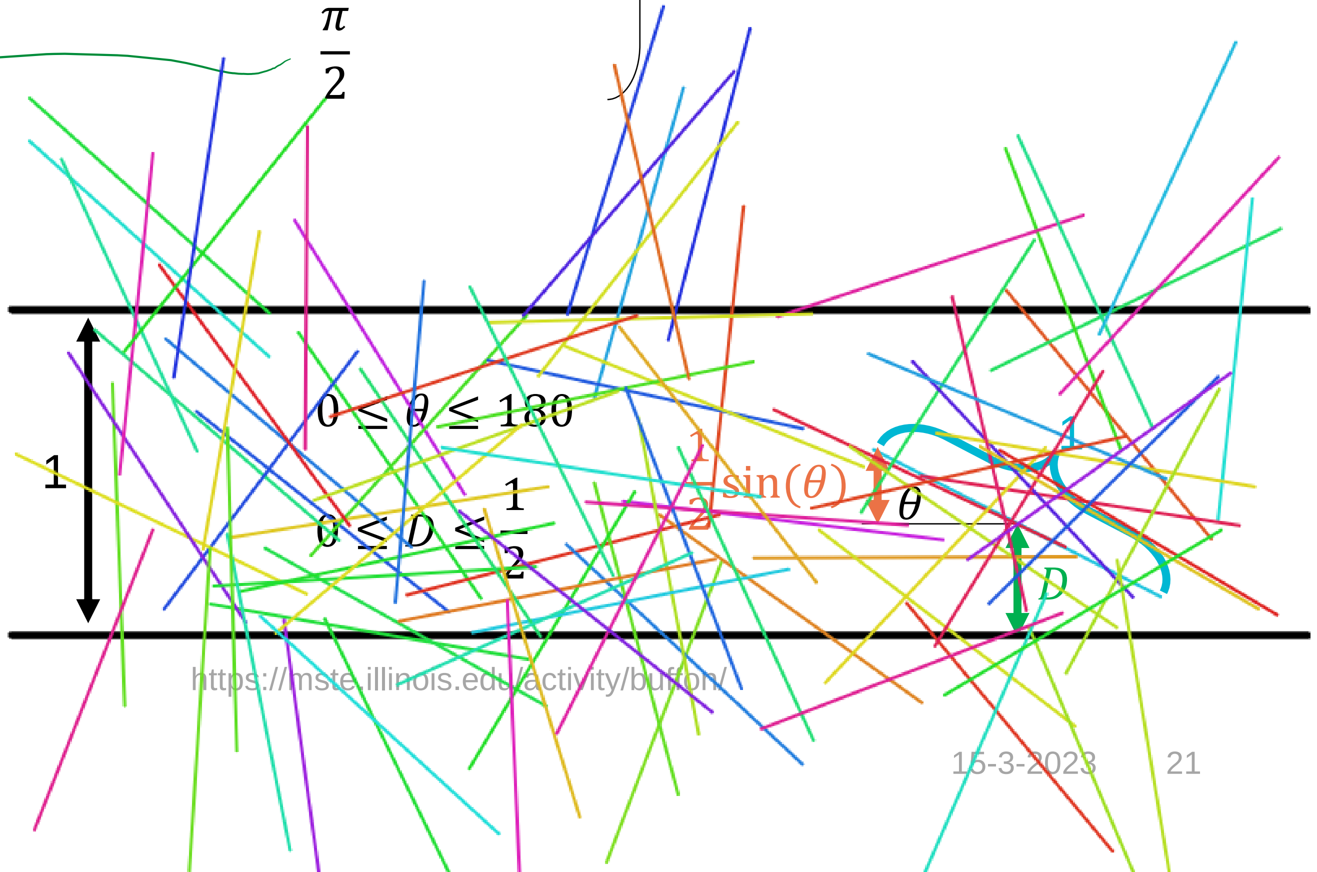


By the Law of Large Numbers

$$\int_0^\pi f(x)dx = 1$$

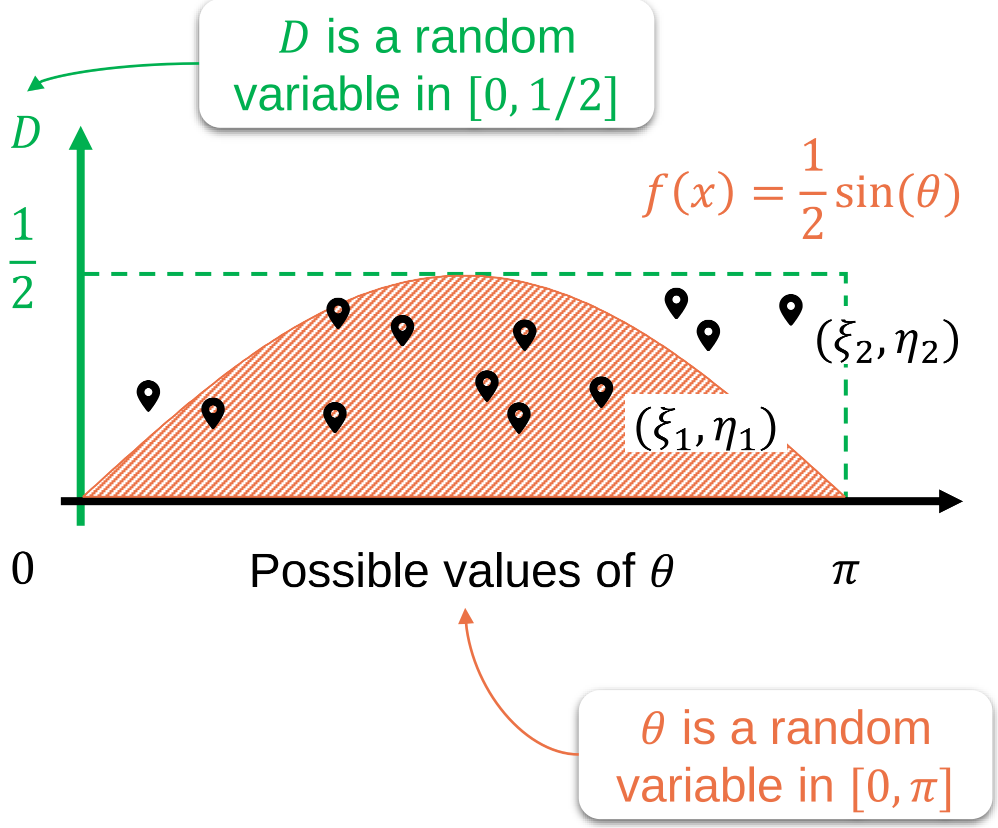
$$\frac{\pi}{2}$$

$$P(hit) = \frac{2}{\pi} \Rightarrow \pi \approx 2\frac{drops}{hits}$$



https://mste.illinois.edu/activity/buffon/

Monte-Carlo Method for Integration



Find the area bounded by the curve $f(x)$

- Let a point fall at random in $[0, \pi] \times [0, 1/2]$.
- What is the probability p that the point falls in the area under the curve?
- We model two independent random variables ξ uniformly distributed on $[0, \pi]$ and η uniformly distributed on $[0, 1/2]$.
- Did event $f(\xi_1) < \eta_1$ happen? Yes. 🏆
- $\forall (\xi_i, \eta_i), i = 2, \ldots, N$ check $f(\xi_i) < \eta_i$.
- $\frac{\sum 🏆}{N} \approx p = \frac{\int_0^{\pi} f(x)dx}{\pi/2}$.

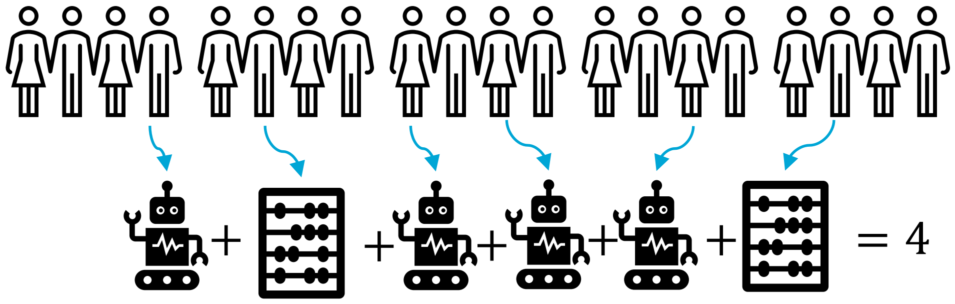
Error $O\left(\sqrt{\frac{D}{N}}\right), D = const$

https://mste.illinois.edu/activity/buffon/

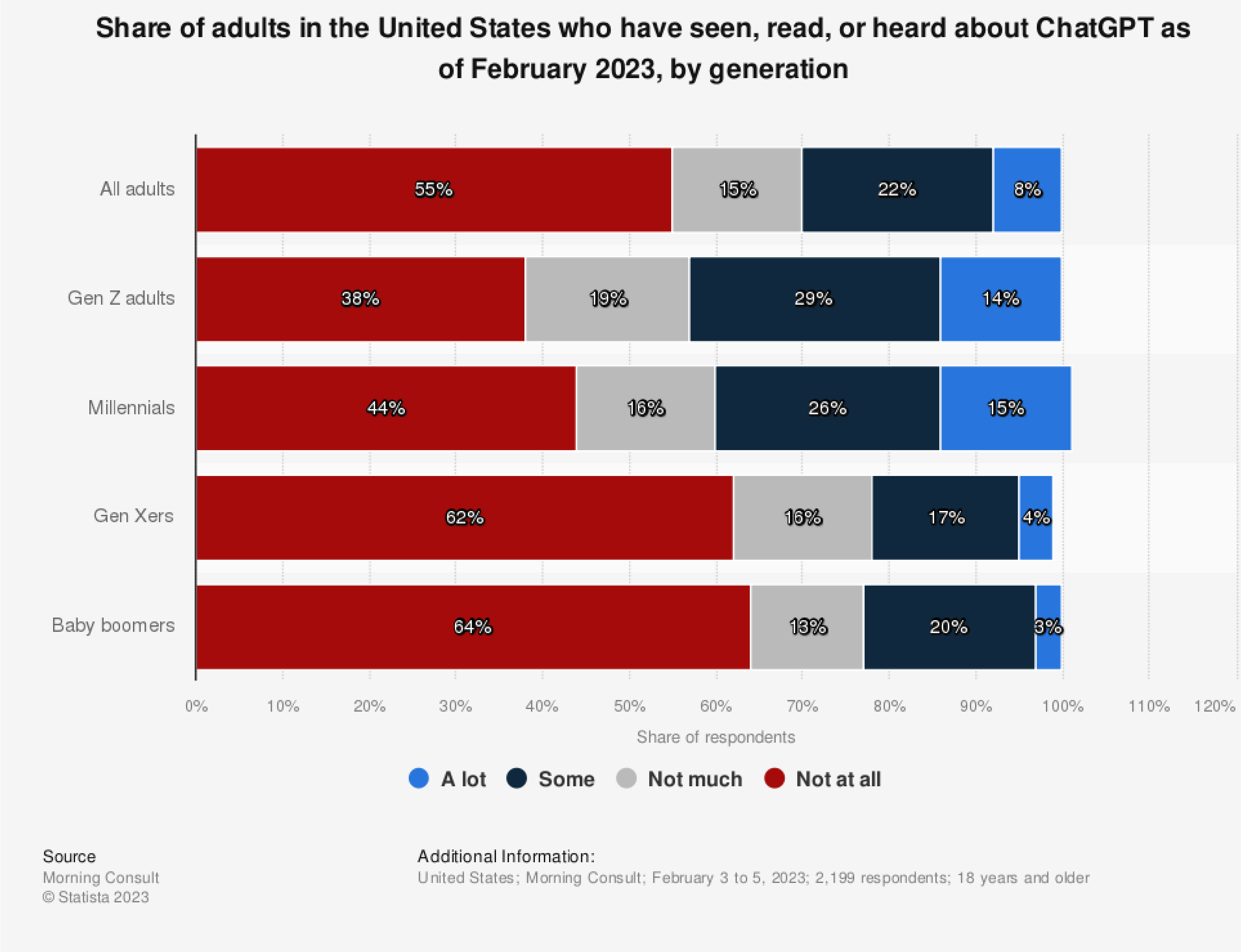
General Method

Bernoulli scheme

- Let random variable $\xi = \begin{cases} 1, \text{if aware}, \\ 0, \text{otherwise}. \end{cases}$
- Let random variable $S_N = \xi_1 + \cdots + \xi_N$.



$$\text{Sample mean} = \frac{2}{3}$$

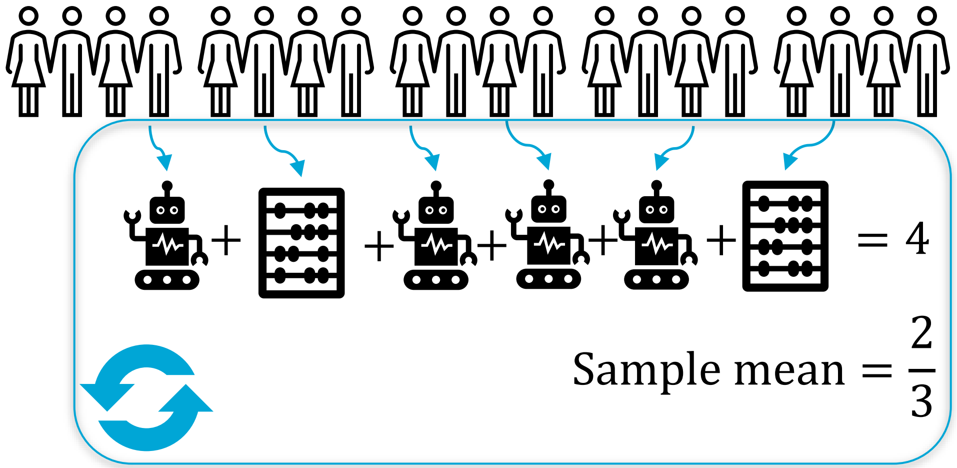


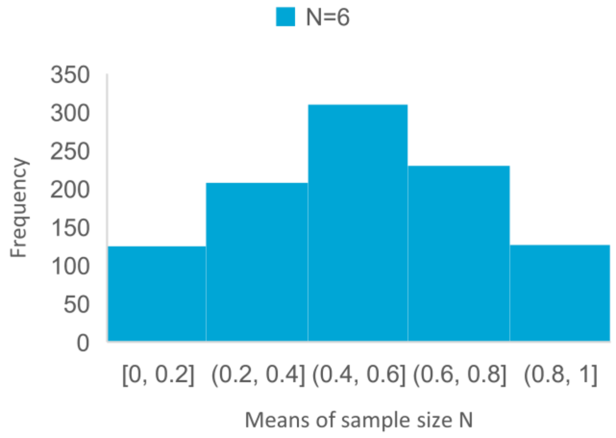
~~UNKNOWN~~ ESTIMATED MEAN!

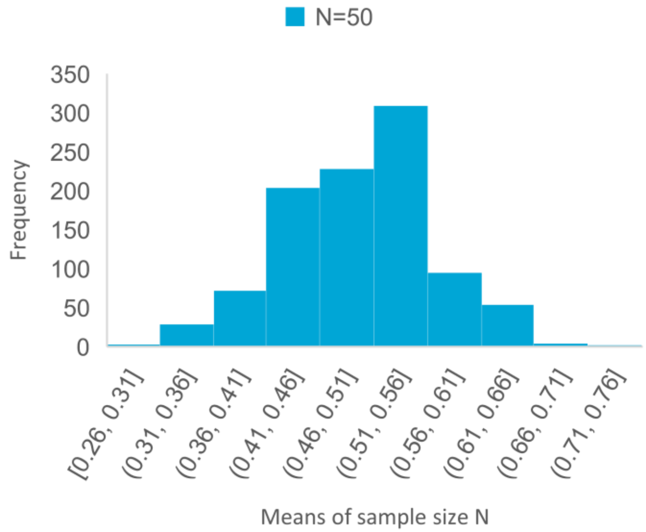
General Method

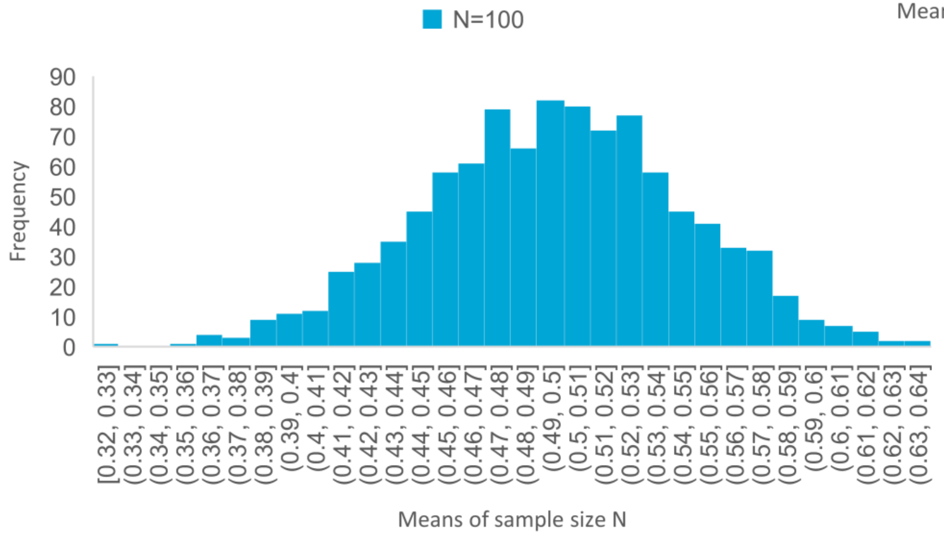
Bernoulli scheme

- Let random variable $\xi = \begin{cases} 1, \text{if aware,} \\ 0, \text{otherwise.} \end{cases}$
- Let random variable $S_N = \xi_1 + \cdots + \xi_N$.









Formally Central Limit Theorem

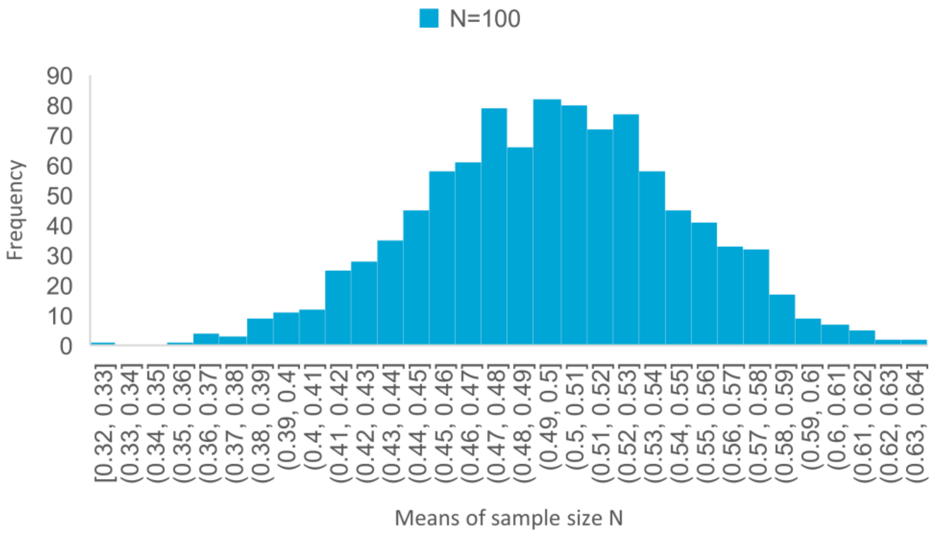
Let $\xi 1, \xi 2, \ldots$ be a sequence of independent identically distributed (nondegenerate) random variables with $\mathrm{E}\xi_1^2 < \infty$ and $S_N = \xi_1 + \cdots + \xi_N$. Then

$$\mathrm{P}\left\{\frac{(S_N - \mathrm{E}S_N)}{\sqrt{\mathrm{Var}S_N}} \le x\right\} \xrightarrow[n\to\infty]{} \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du\,.$$

For large sample size N holds:

$$S_N \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{N}}\right),$$

where μ and σ are true values under estimation



Part III: Simulation-Based Approximation

Lecture 1: Estimating Uncertainty

1. What and why?
   1. How to model uncertainty
   2. Long before Monte-Carlo method
   3. Essential theory
2. Monte-Carlo method
   1. General method and complexity
   2. Central Limit Theorem

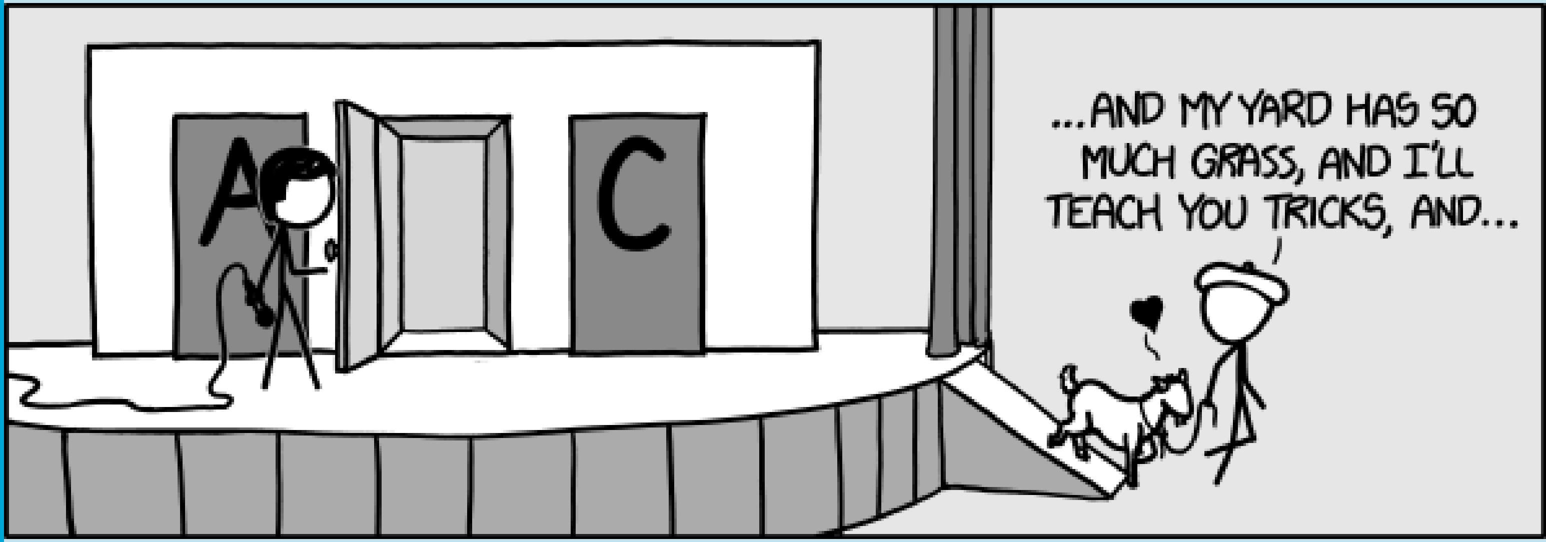
Algorithms for NP-Hard Problems

Part III: Simulation-Based Approximation

Lecture 2: Advanced Sampling Methods

Instructor: Dr. Anna Lukina

a.lukina@tudelft.nl

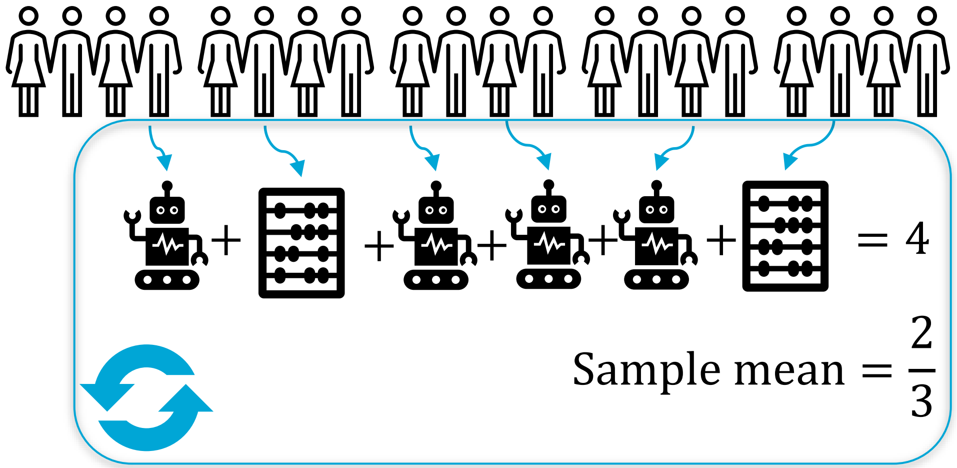


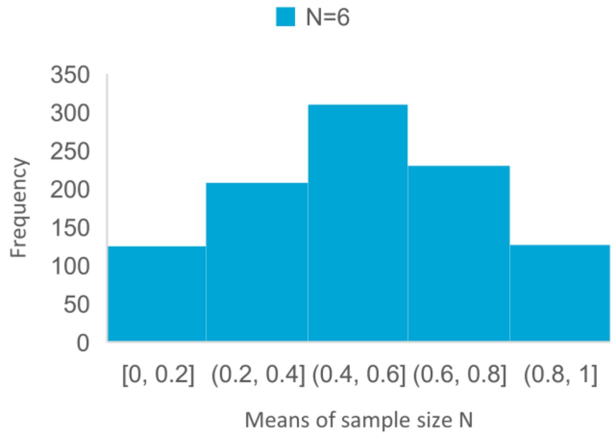
TU Delft

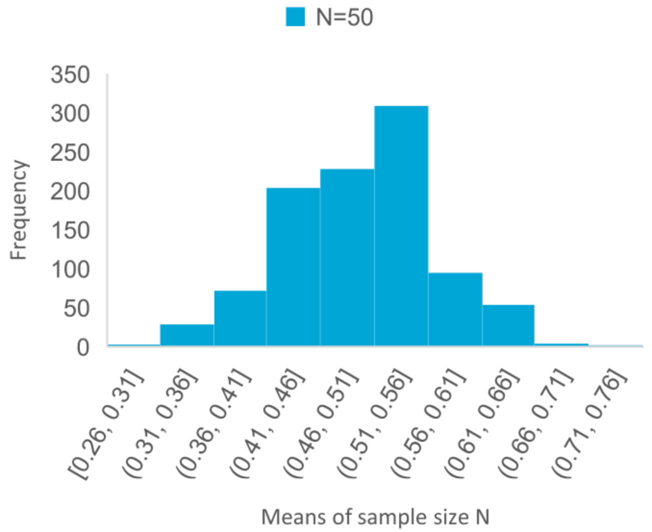
General Monte-Carlo Method

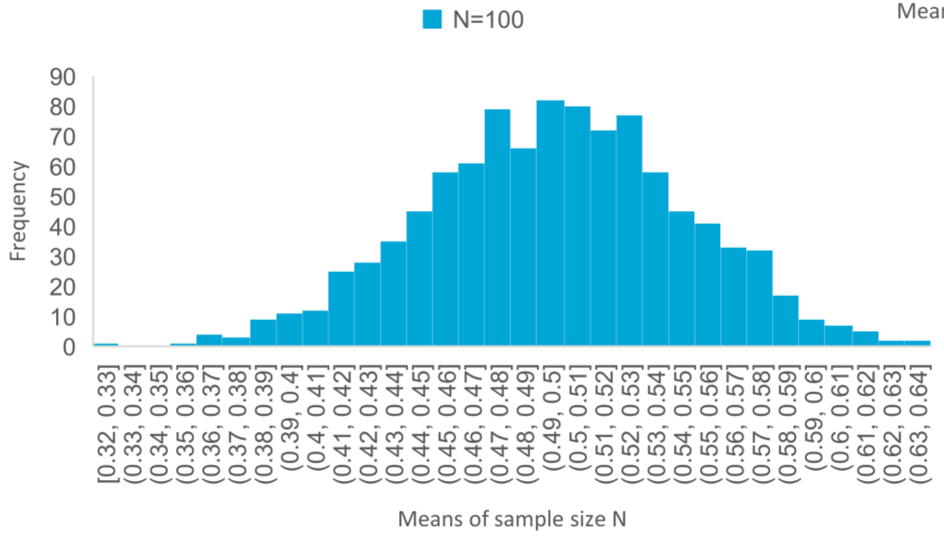
Bernoulli scheme

- Let random variable $\xi = \begin{cases} 1, \text{if aware}, \\ 0, \text{otherwise}. \end{cases}$
- Let random variable $S_N = \xi_1 + \cdots + \xi_N$.









General Central Limit Theorem

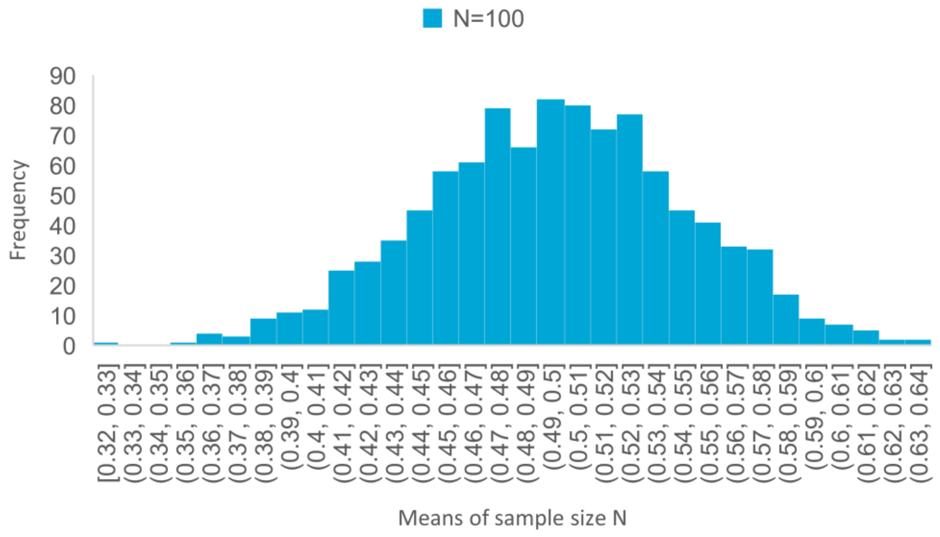
Let $\xi_1, \xi_2, \ldots$ be a sequence of independent identically distributed (nondegenerate) random variables with $\mathrm{E}\xi_1^2 < \infty$ and $S_N = \xi_1 + \cdots + \xi_N$. Then

$$\mathrm{P}\left\{\frac{(S_N - \mathrm{E}S_N)}{\sqrt{\mathrm{Var}S_N}} \leq x\right\} \xrightarrow[N\to\infty]{} \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du \,.$$

For large sample size N holds:

$$S_N \sim \mathcal{N}\left(\mu N, \sigma\sqrt{N}\right),$$

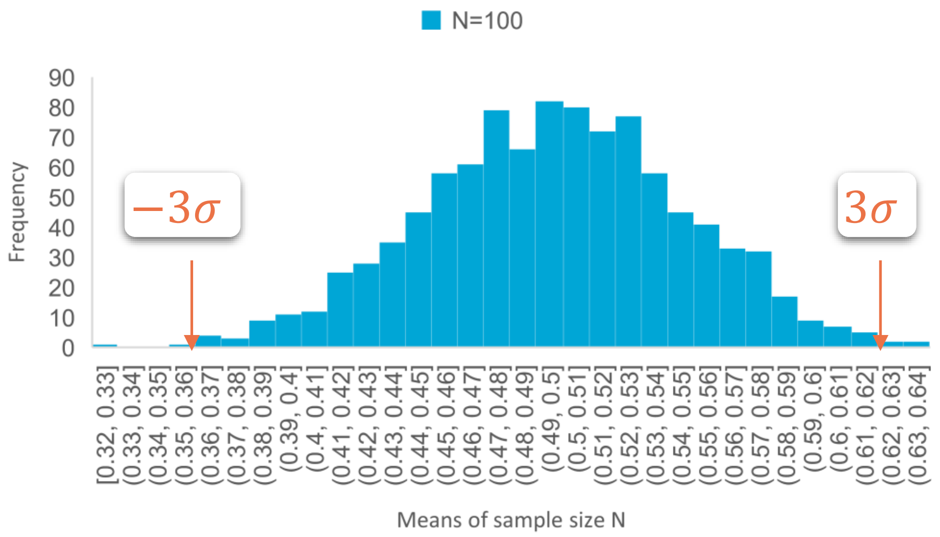
where μ and σ are the mean and standard deviation of the population we sample from



Central Limit Theorem

For large sample size N holds:

$$S_N \sim \mathcal{N}\left(\mu N, \sigma\sqrt{N}\right),$$



$$z = \frac{S_N/N - \mu}{\sigma/\sqrt{N}} \sim \mathcal{N}(0,1)$$

- Let X be a random variable with $\mu = 10$ and $\sigma = 4$. The sample size $N = 100$.
- Find the probability that the sample mean of these 100 observations is less than 9.
- $\mathrm{P}\left(\frac{S_N}{N} < 9\right) = \mathrm{P}\left(z < \frac{9-10}{\frac{4}{\sqrt{100}}}\right) = 0.0062$ from the standard normal distribution table.

Part III: Simulation-Based Approximation

Lecture 1: Estimating Uncertainty

Lecture 2: Advanced Sampling Methods

Part III: Simulation-Based Approximation

Lecture 1: Estimating Uncertainty

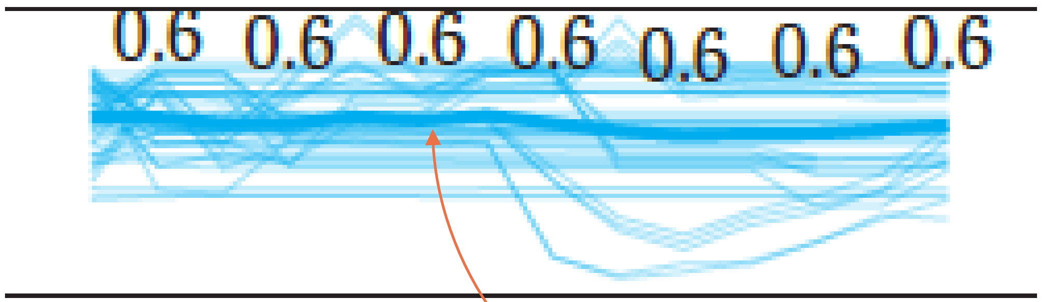
Lecture 2: Advanced Sampling Methods

1. Simulation use
   i. Wind farms
   ii. Confidence
2. More real problems
   i. Robot planning
   ii. Financial failures
3. Probabilistic sampling
   i. Conditional probabilities
   ii. Markov chains

Wind Farm Optimization

NP-Hard

Over one month of simulated time



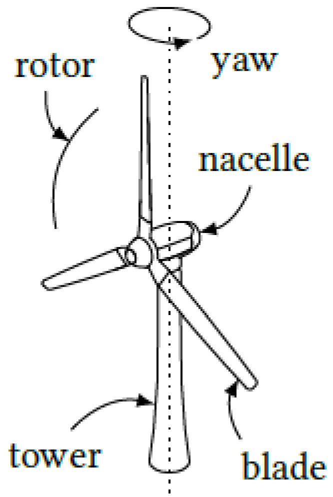


Figure 1: Turbine nomenclature.

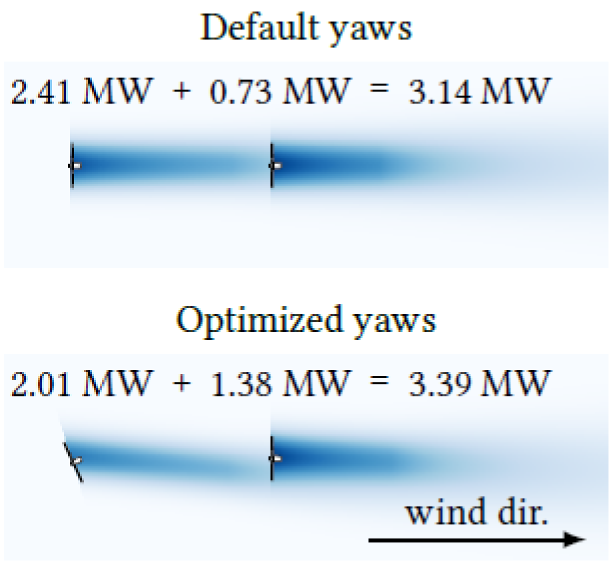


Figure 2: Overhead view of two wind turbines with yaw-based wake control. Darker areas have slower wind.

Mean normalized reward of reinforcement learning agents that control the turbine

Neustroev, G. (2022). Generalized Models of Sequential Decision-Making under Uncertainty. PhD Thesis.

Robot Planning with Reinforcement Learning

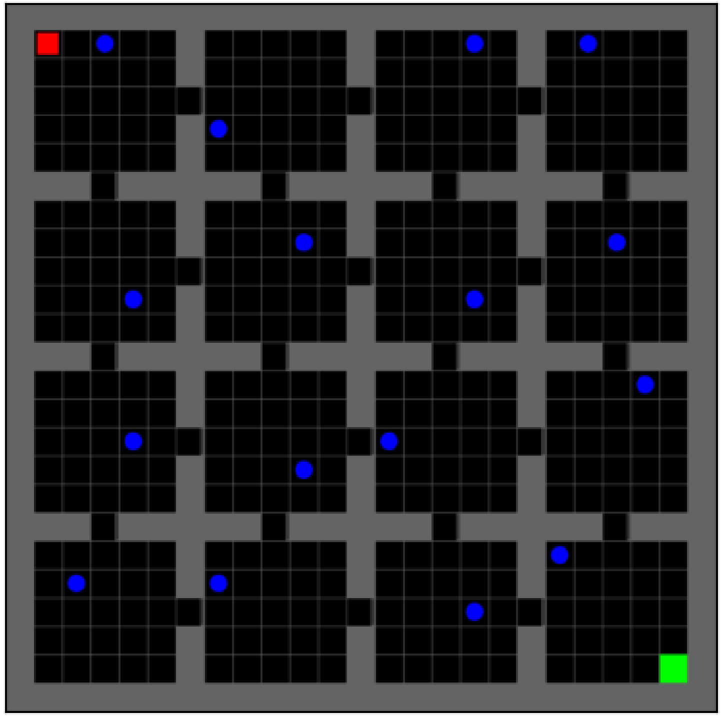
How to give confidence that this agent succeeds to reach green in 90% of the runs?

How large should the number of simulations be?

Collect statistics for random variable

$$S_N = \xi_1 + \cdots + \xi_N$$

$$\xi_i = \begin{cases} 1, \text{if reached green} \\ 0, \text{otherwise} \end{cases}$$



get to the green goal square

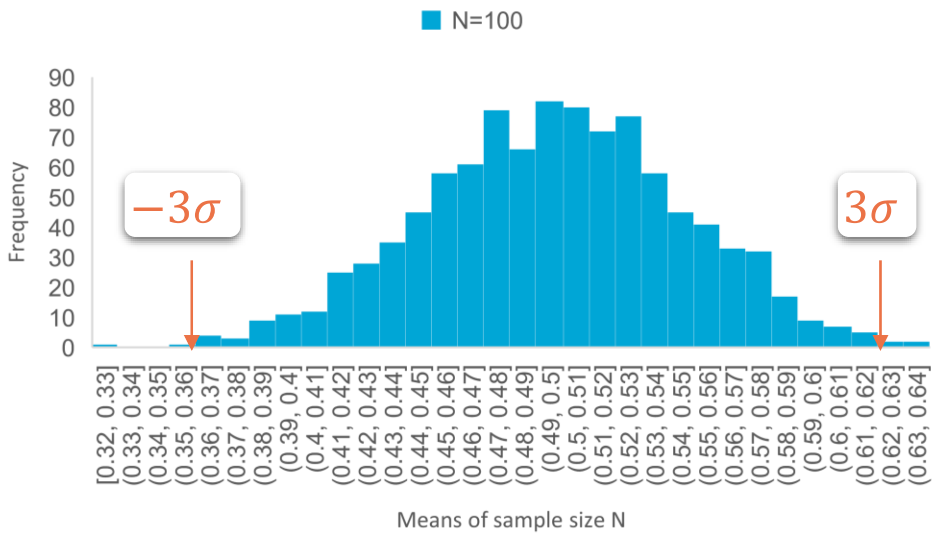
Pictures taken from MSc thesis by C. van Rijn, TU Delft 2023

Confidence Bounds

For large sample size N holds:
$S_N \sim \mathcal{N}(\mu N, \sigma\sqrt{N}),$

$$\mathrm{P}\left\{\mu - \frac{3\sigma}{\sqrt{N}} \leq \frac{S_N}{N} \leq \mu + \frac{3\sigma}{\sqrt{N}}\right\} \geq 0.997$$

Mean reachability success



(ε, δ) – approximation scheme:
$\mathrm{P}\{\mu - \varepsilon \leq \hat{\mu} \leq \mu + \varepsilon\} \geq 1 - \delta,$
where $\hat{\mu}$ approximates μ with absolute error ε and probability $1 - \delta$.

How Large is Enough

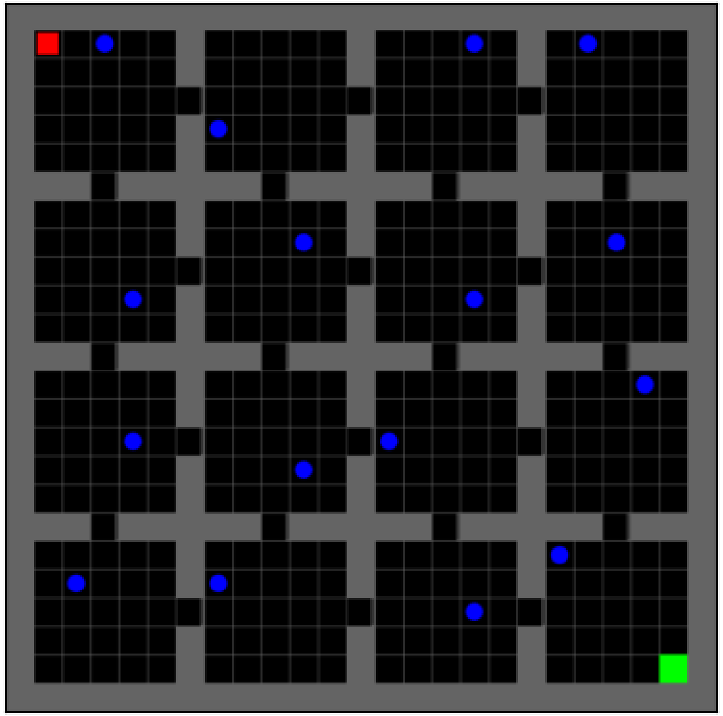
(ε, δ) – approximation scheme:
$P\{\mu - \varepsilon \leq \hat{\mu} \leq \mu + \varepsilon\} \geq 1 - \delta,$
where $\hat{\mu}$ approximates μ with absolute error ε and probability $1 - \delta$.

Hoeffding's inequality (for Bernoulli scheme):

$$\delta = P\{|\hat{\mu} - \mu| > \varepsilon\} \leq 2e^{-2\varepsilon^2 N}$$

(exponentially small tails)

$$N \geq \frac{Log\left(\frac{2}{\delta}\right)}{2\varepsilon^2}$$



get to the green goal square

Pictures taken from MSc thesis by C. van Rijn, TU Delft 2023

Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. The Annals of Mathematical Statistics, 493-507.

Recent failure

- [SVB](#) Financial Corp. was [seized by regulators on Friday](#). **(March 10, 2023)**

Reuters

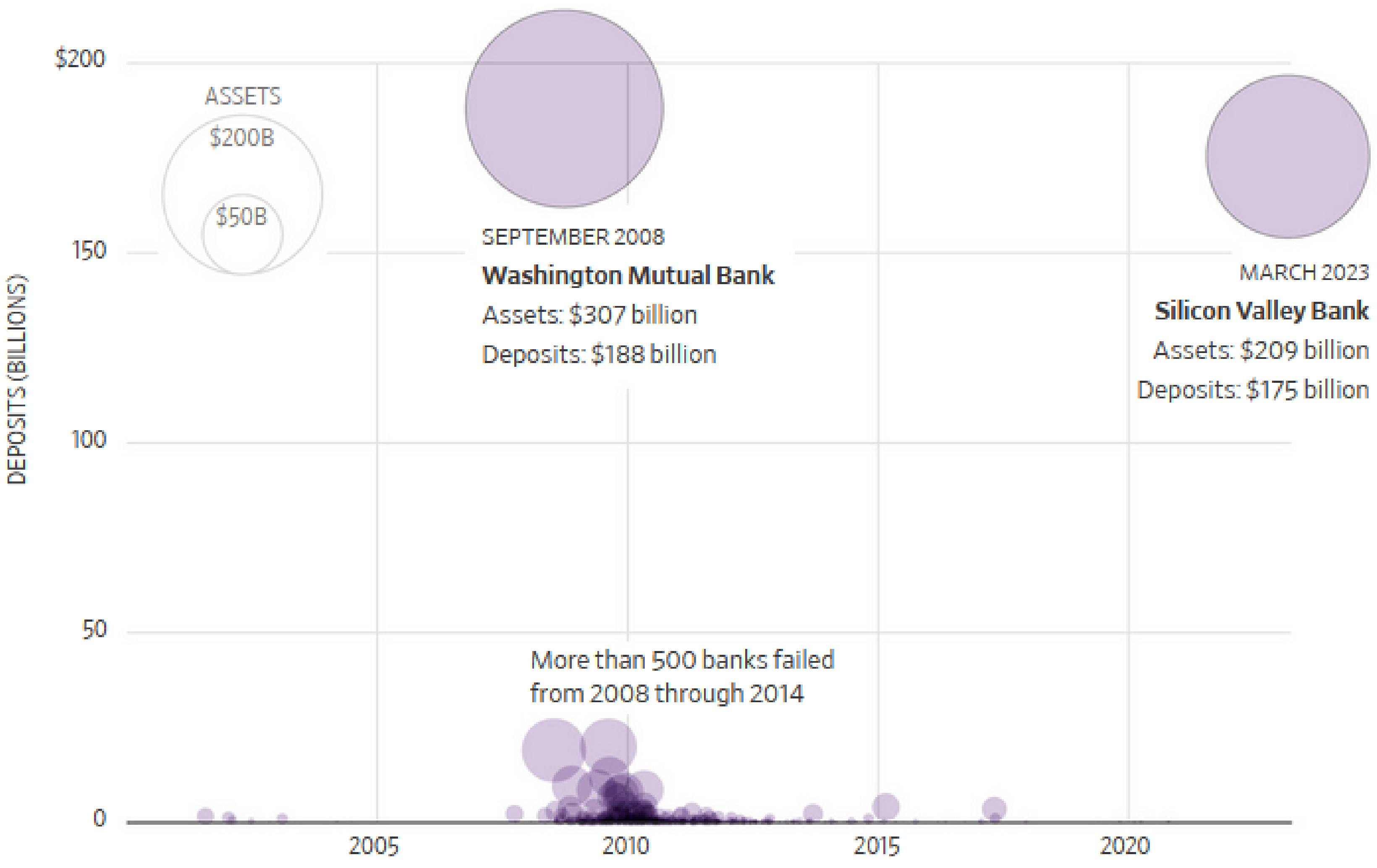
BANK RUN

Customers wait in line outside a branch of the Silicon Valley Bank in Wellesley, Massachusetts, U.S., March 13, 2023. REUTERS/Brian Snyder

Recent failure

- [SVB](#) Financial Corp. was seized by regulators on Friday. **(March 10, 2023)**
- *Ratings agency Moody's cut its outlook on the U.S. banking system to negative from stable "to reflect the rapid deterioration in the operating environment."* ***(March 14, 2023)***

Bank failures, 2001-23



Note: Figures for assets and deposits are estimates.
Source: Federal Deposit Insurance Corp.

Recent failure

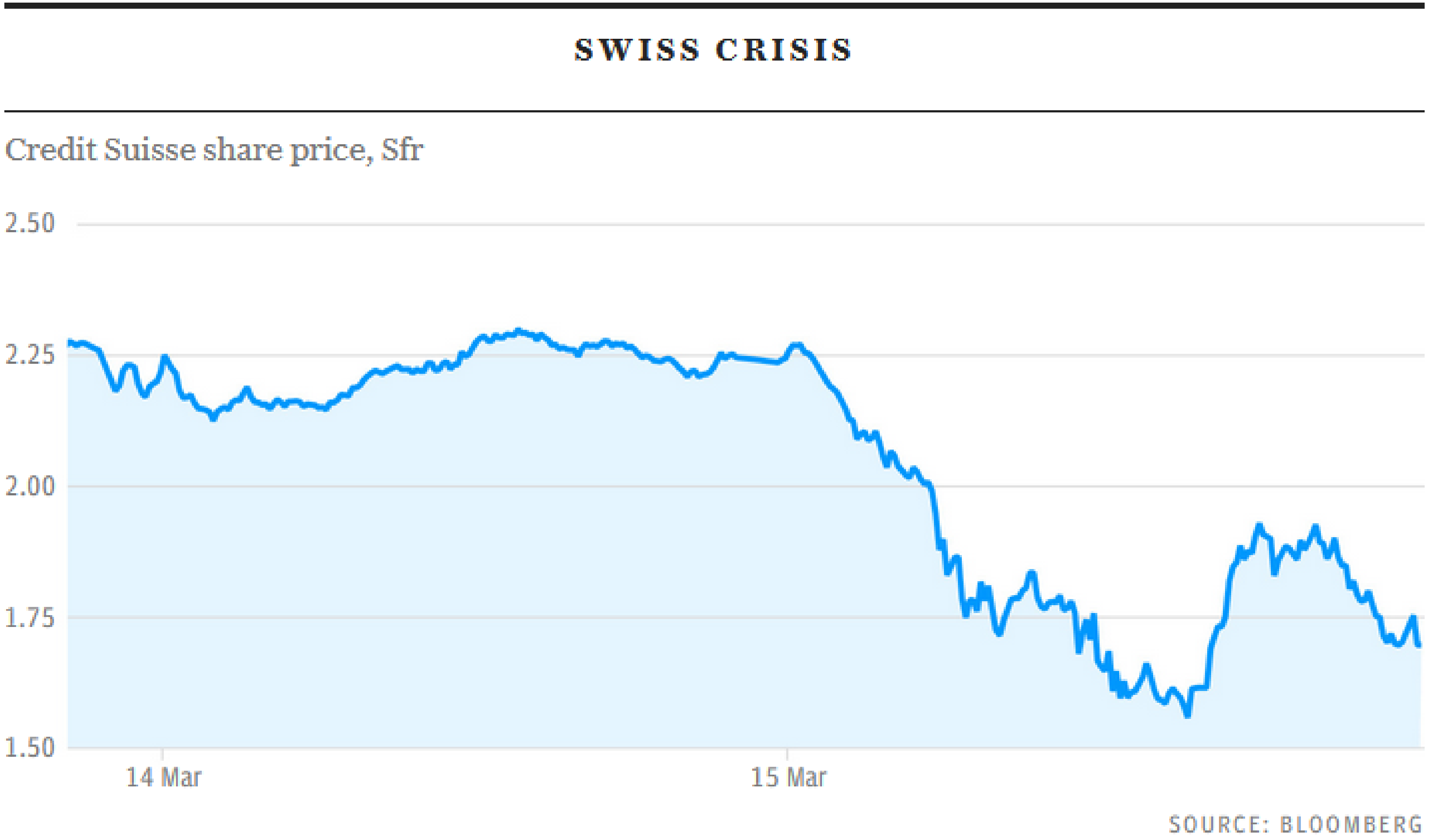
- SBNY -22.87% was closed by regulators on Sunday, the second massive bank failure in three days. **(March 12, 2023)**
- *“We were fine until the last couple of hours on Friday.”(Signature board member Barney Frank, **March 15, 2023**)*

Signature Bank became one of the crypto market’s leading banks.

PHOTO: MARISSA ALPER FOR THE WALL STREET JOURNAL

Recent failure

- Trading in Credit Suisse shares was halted as they fell as much as 21% on Wednesday. **(March 15, 2023)**
- Europe's 17th largest lender**.**
- *European banks were rocked by another turbulent session on Wednesday, plunging as much as 10% amid concerns over the global fallout.*

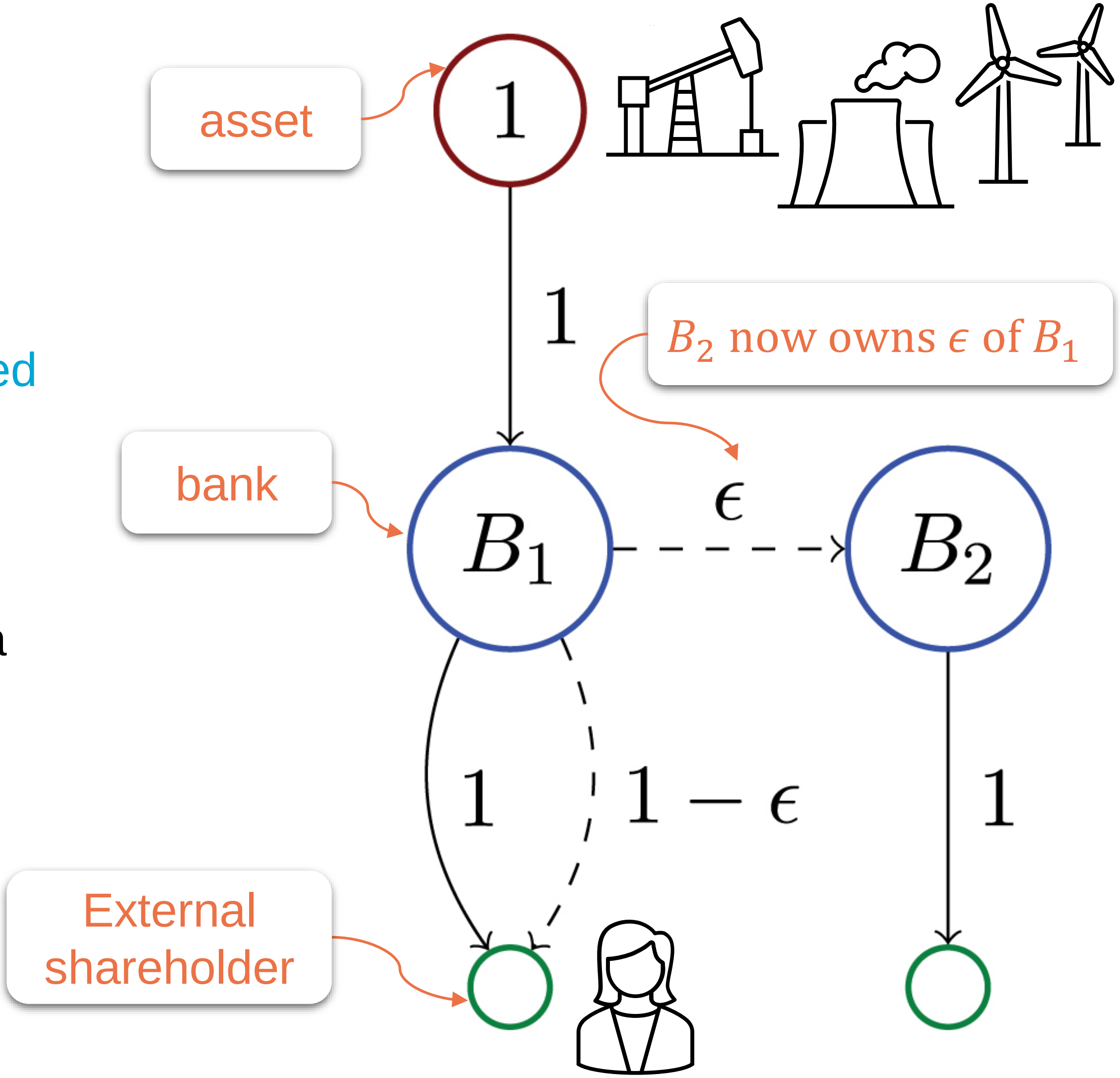


Predicting Failures is Hard

… NP-Hard

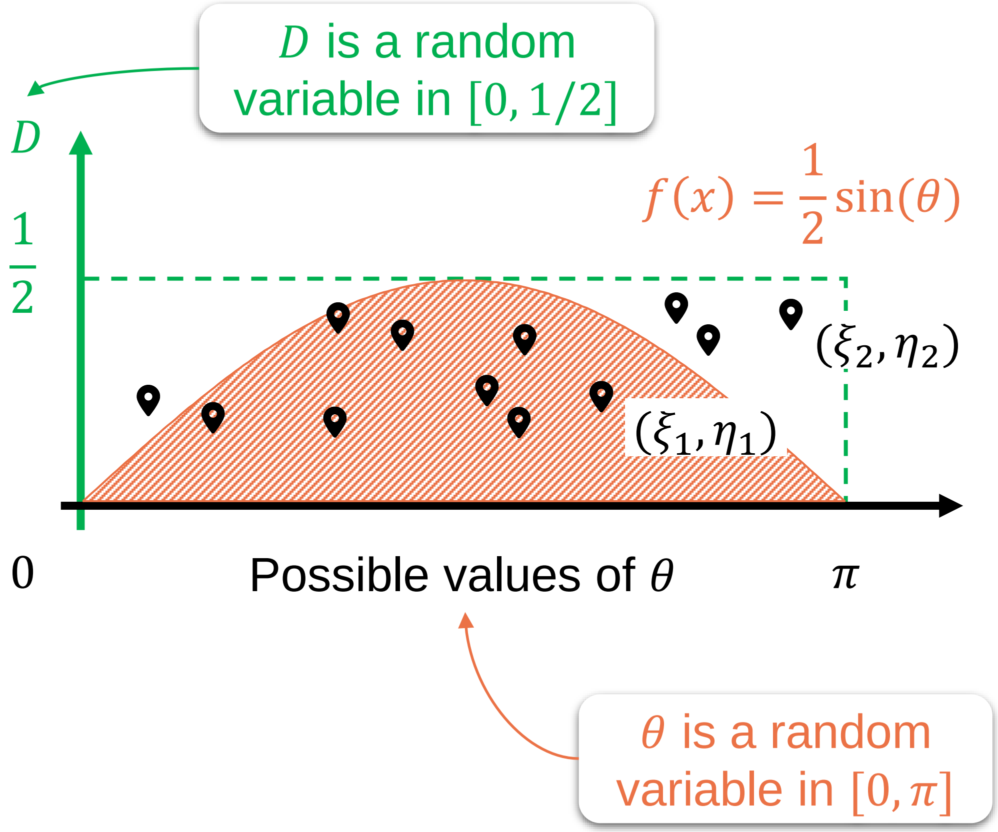
What is the maximum number of failures that can be caused by a drop of this size in asset values?

- An ε change in bank holding can change the market value of the bank by ε.
- Hardness result is based on the hardness of finding a maximum balanced clique in a bipartite graph.
- This complexity arises not from cycles, but from the nonlinear dynamics that occur when a bank drops below its critical threshold value.



Hemenway, Brett and Khanna, Sanjeev. 'Sensitivity and Computational Complexity in Financial Networks'. 2016

Monte-Carlo Method for Integration

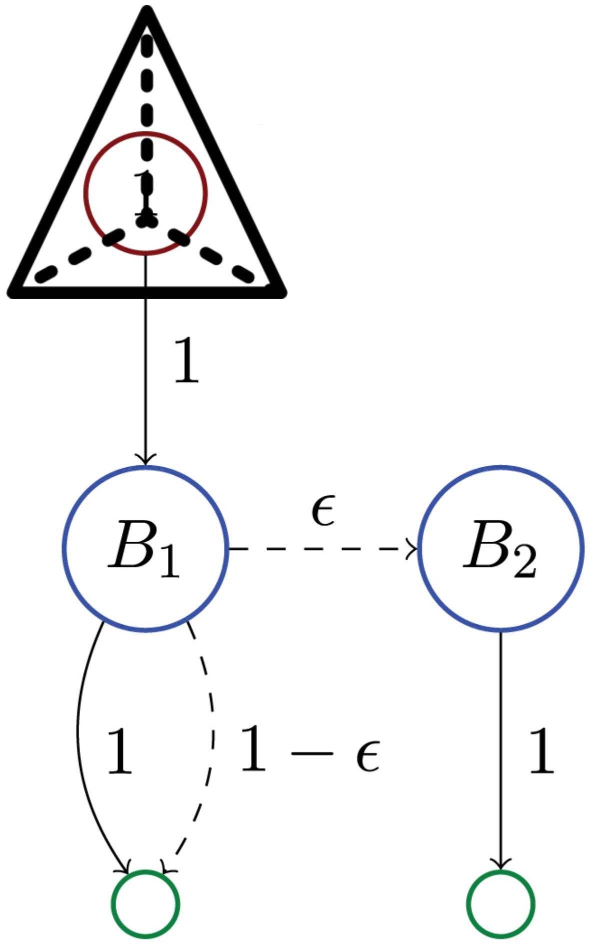


Find the area bounded by the curve $f(x)$

- Let a point fall at random in $[0, \pi] \times [0, 1/2]$.
- What is the probability p that the point falls in the area under the curve?
- We model two independent random variables ξ uniformly distributed on $[0, \pi]$ and η uniformly distributed on $[0, 1/2]$.
- Did event $f(\xi_1) < \eta_1$ happen? Yes. 🏆
- $\forall (\xi_i, \eta_i), i = 2, \dots, N$ check $f(\xi_i) < \eta_i$.
- $\frac{\sum 🏆}{N} \approx p = \frac{\int_0^\pi f(x)dx}{\pi/2}$.

Error $O\left(\sqrt{\frac{D}{N}}\right), D = const$

Multi-Dimensional Sampling



Sample from

$$X = \{ (x_1, x_2, \dots, x_D) \mid 0 \leq x_i \leq 1,$$
$$x_1 + x_2 + \dots + x_D = 1\}.$$

D is the dimension of the simplex.

D simplex is a convex hull, where each point on the surface describes a probability distribution over D outcomes.

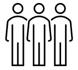
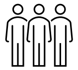
Joint Probability Distribution

- Specify a real number for each assignment (outcome) over a set of random variables $\xi_1, \xi_2, \dots, \xi_n$

Must hold:

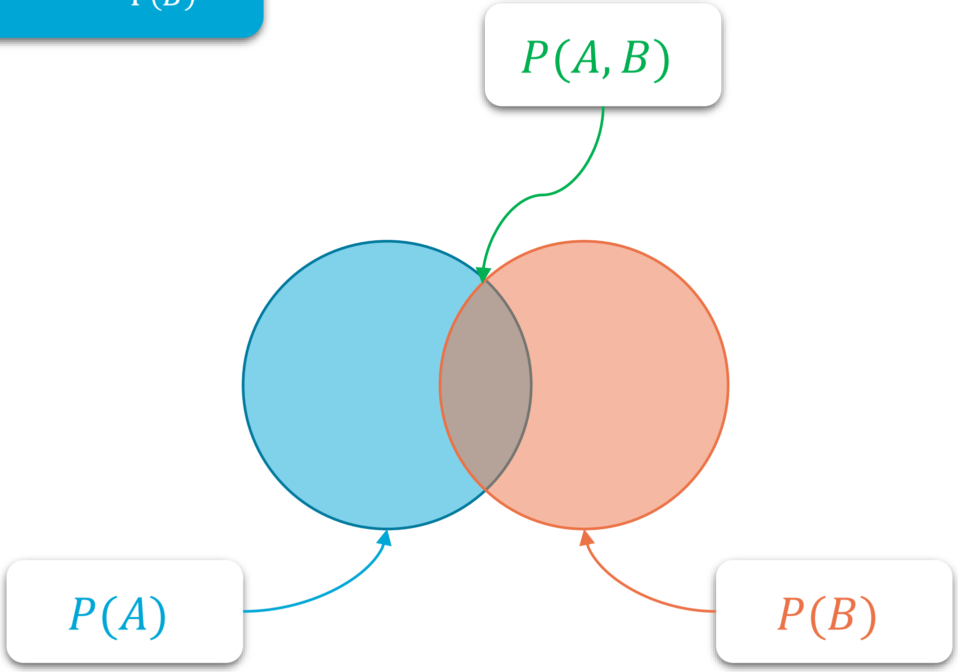
$$\mathrm{P}(\xi_1 = x_1, \xi_2 = x_2, \dots, \xi_n = x_n) \geq 0$$

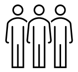
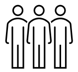
$$\sum_{(x_1, x_2, \dots, x_n)} \mathrm{P}(x_1, x_2, \dots, x_3) = 1$$

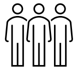
C	T	$\mathrm{P}(C, T)$
	$x_2 < 10\ min$	0.1
	$10\ min \leq x_2$	0.1
	$x_2 < 10\ min$	0.4
	$10\ min \leq x_2$	0.4

Conditional Probability Distribution

$$P(A|B) = \frac{P(A,B)}{P(B)}$$

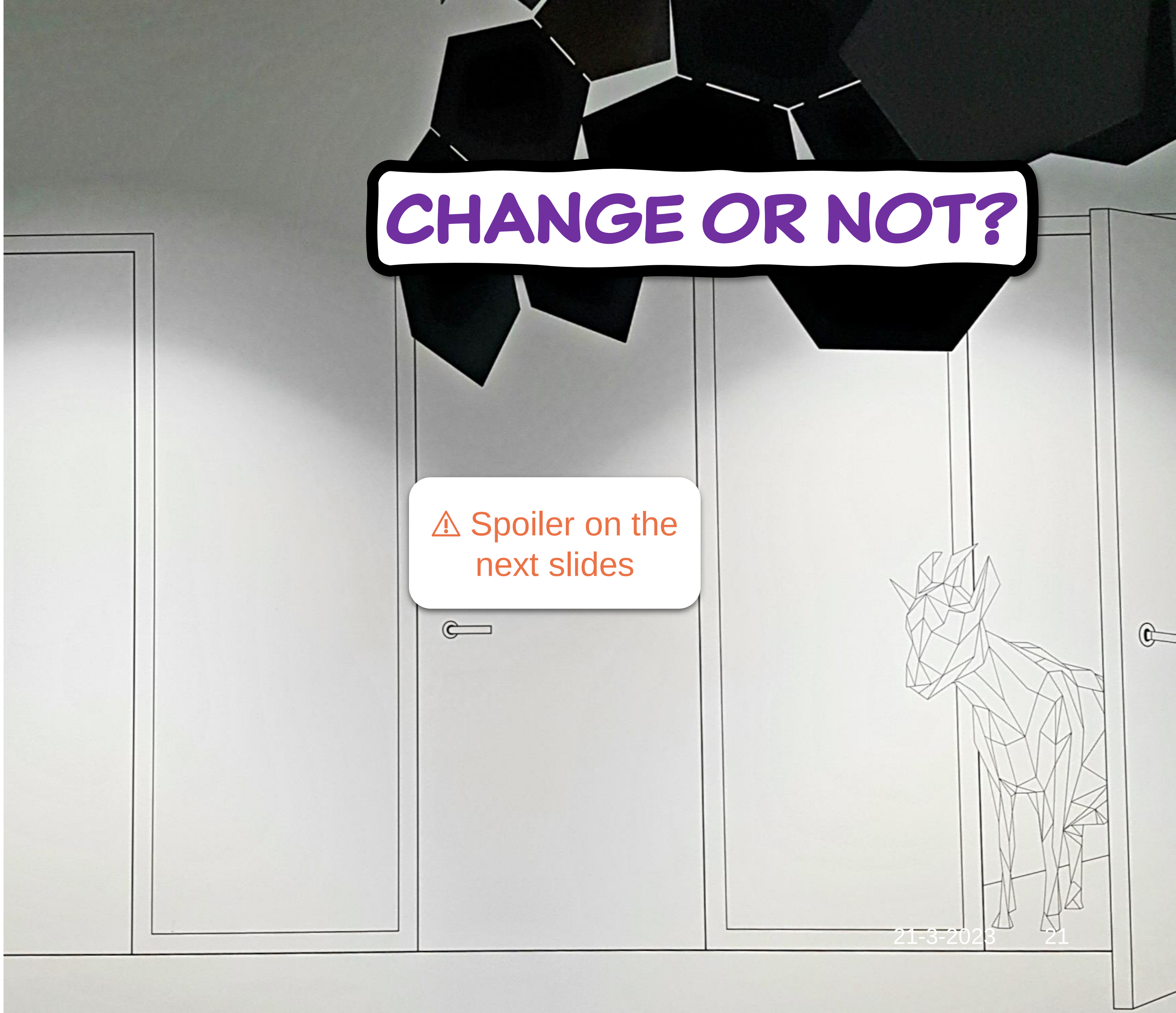


C	T	$P(C,T)$
	$x_2 < 10\ min$	0.1
	$10\ min \leq x_2$	0.1
	$x_2 < 10\ min$	0.4
	$10\ min \leq x_2$	0.4

$P(x_2 < 10|$  $) = ?$

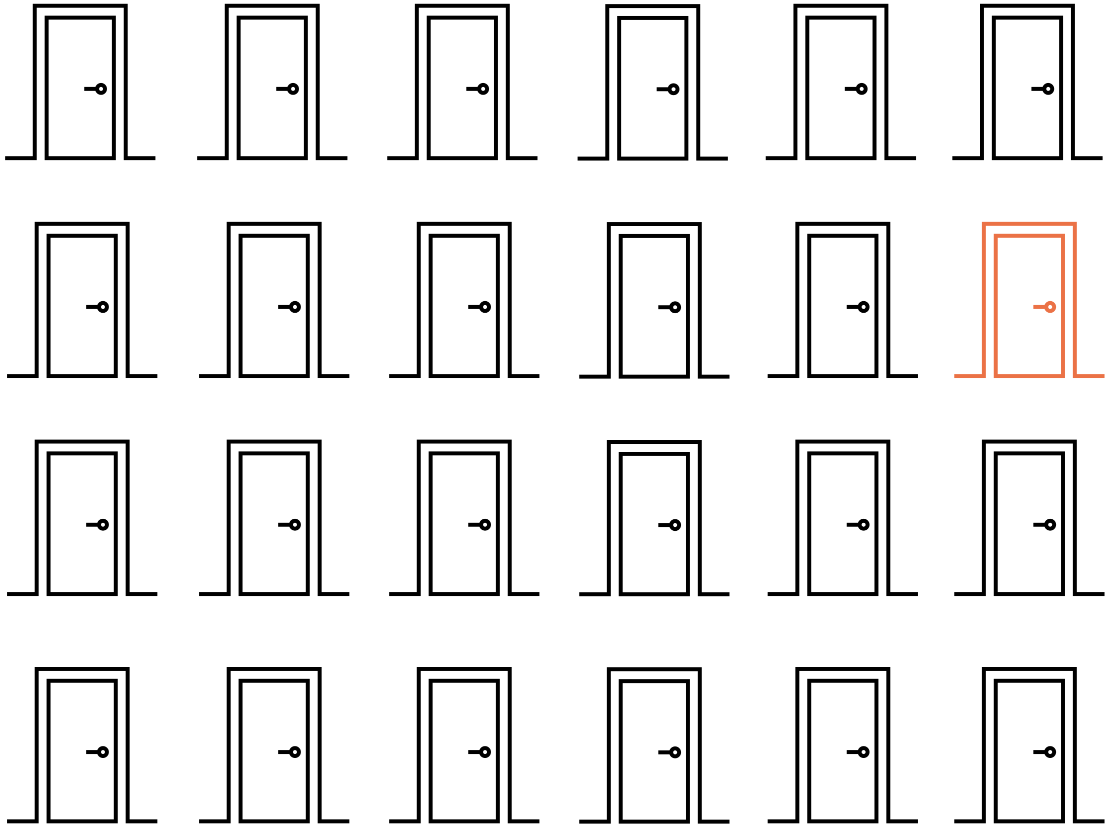
The Goat Problem

- Loosely based on an American TV Show from 1975 and dubbed Monty Hall Problem.
- Solved by Marilyn vos Savant in 1990.
- The Game:
  - There is a prize behind one of three doors and goats behind the rest.
  - We pick a door.
  - Then Monty open one door among the ones we did not pick and reveals one goat.
  - We are allowed to change.



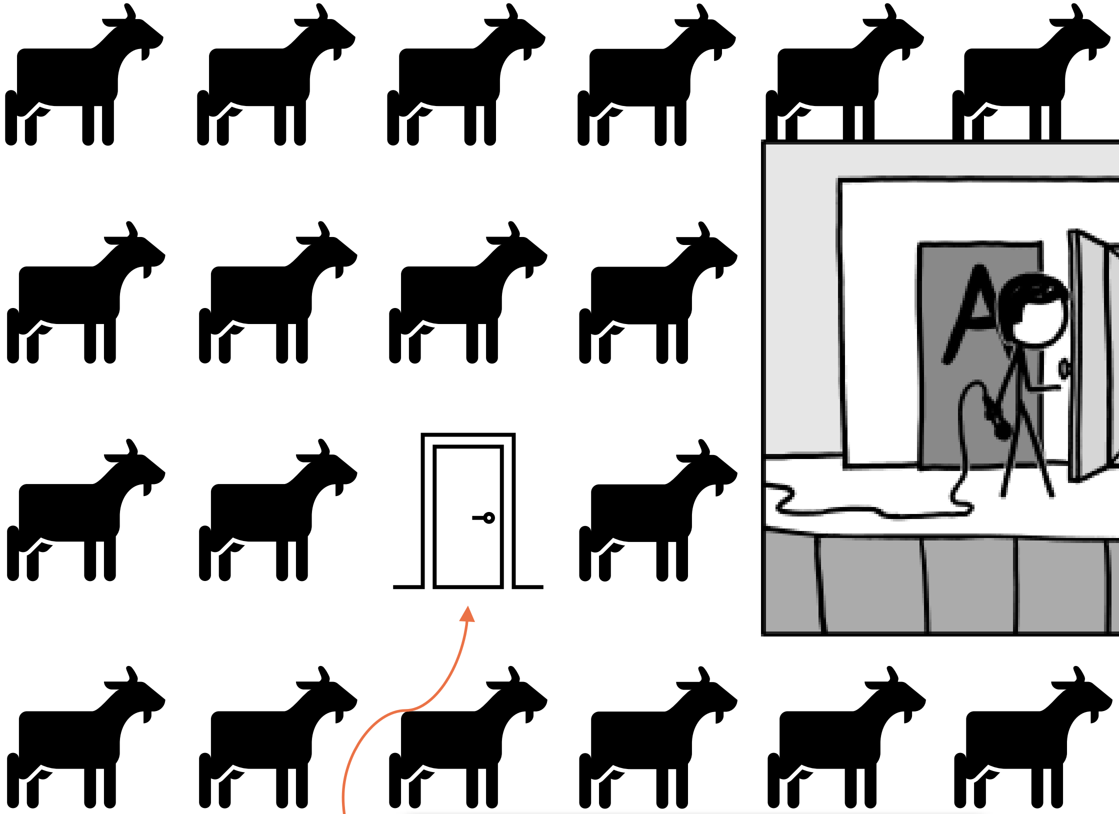
The Goat Problem

CHANGE OR NOT?

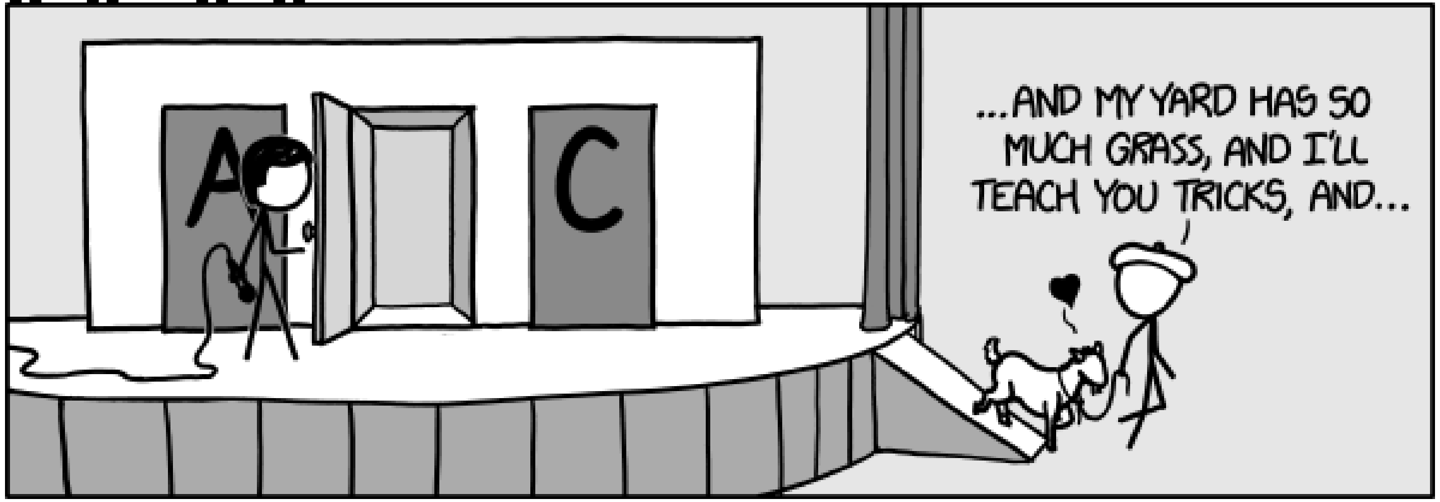


We pick this one

The Goat Problem



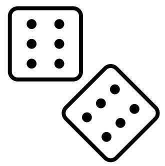
CHANGE!

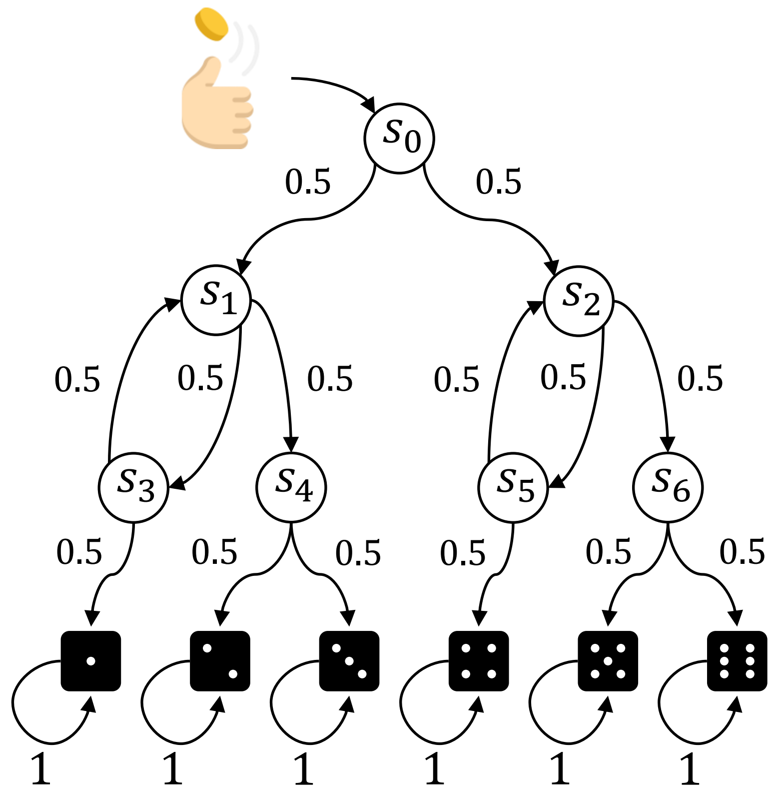


The prize is here with higher probability

https://imgs.xkcd.com/comics/monty_hall.png

Simulating a Dice (with only fair coins)



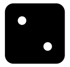


[Knuth & Yao, 1976]

Heads = “go left”; tails = “go right”.

Does this model a six-sided dice?

Reachability Probabilities (Knuth-Yao’s Die):

- Consider the event $A =$ 
- We model: $\xi = I_A(\omega)$,

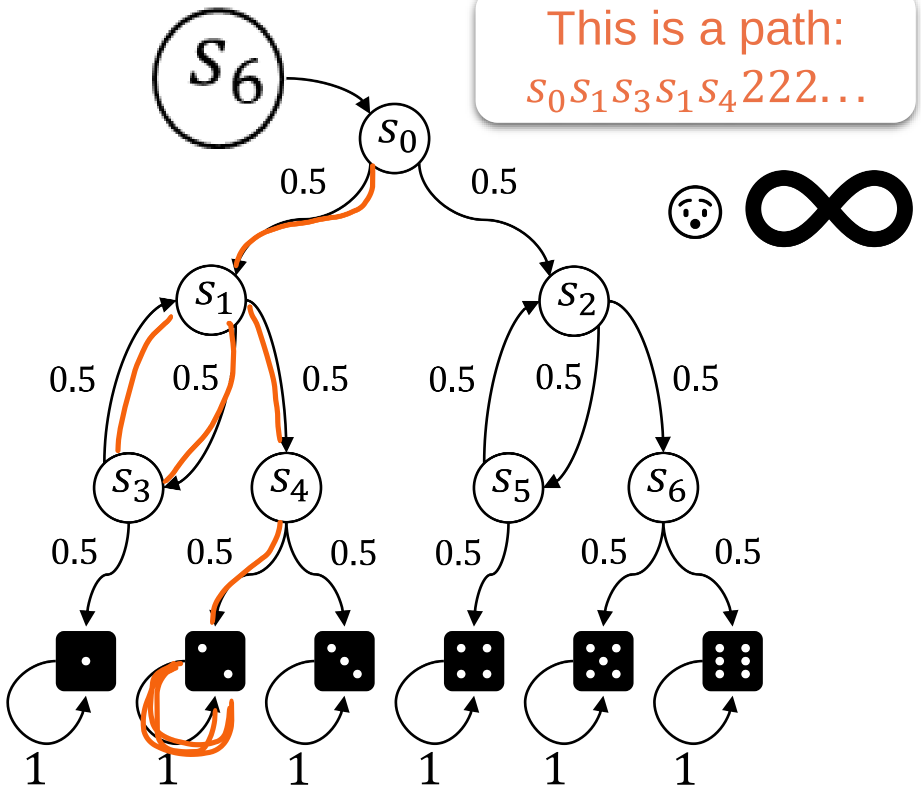
where $I_A(\omega) = \begin{cases} 1, \omega \in A, \\ 0, \omega \notin A. \end{cases}$

What is the probability of reaching A from s_0?

$$p_{s_4} = \frac{1}{2}, p_{s_1} = \frac{1}{3}, p_{s_3} = \frac{1}{6}, \text{and } p_{s_0} = \frac{1}{6}.$$

Joost-Pieter Katoen. 2016. The Probabilistic Model Checking Landscape. In Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science (LICS ’16).

Markov Chains



Discrete-time Markov chains are the simplest possible probabilistic models.

$$\mathrm{P}\{\boxed{s_0 s_1 s_4} 2^\omega, \boxed{s_0 s_1 s_4} 3^\omega\} = \mathrm{P}\{s_0 \to s_1\} \cdot \mathrm{P}\{s_1 \to s_4\} = \frac{1}{4}$$

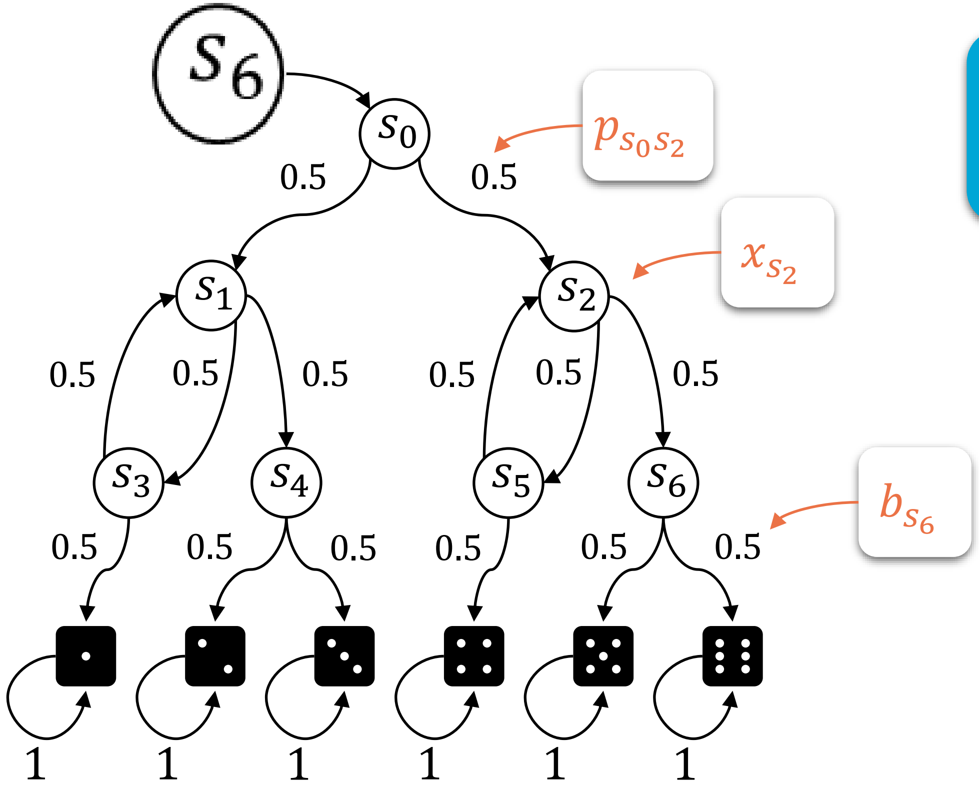
Eventually ⚁ $= \bigcup_{k \in \mathbb{N}} s_0 (s_1 s_3)^k s_1 s_4 2$

$$\sum_{k=0}^{\infty} \mathrm{P}\left(s_0 (s_1 s_3)^k s_1 s_4 2\right) = \frac{1}{8} \sum_{k=0}^{\infty} \left(\frac{1}{4}\right)^k = \frac{1}{8} \cdot \frac{1}{1 - \frac{1}{4}} = \frac{1}{6}$$

Geometric series

Katoen, J. P. (2013). Model checking meets probability: A gentle introduction. In *Engineering dependable software systems* (pp. 177-205). IOS Press.

Markov Chains



Katoen, J. P. (2013). Model checking meets probability: A gentle introduction. In *Engineering dependable software systems* (pp. 177-205). IOS Press.

Reachability probabilities are unique solutions of a given linear equation system.

- Let S be a set of states from which we can reach any side of the dice with > 0 steps
- Let matrix $\boldsymbol{P} = (p_{st})_{s,t \in S}$ be transition probabilities between states in S
- Let vector $\boldsymbol{b} = (b_s)_{s \in S}$ be probabilities to reach any side of the dice in a single step
- Let vector $\boldsymbol{x} = (x_s)_{s \in S}$ be probabilities to eventually reach any side of the dice from states in S

Then $\boldsymbol{x}$ is a unique solution of $(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{x} = \boldsymbol{b}$.

Algorithms for NP-Hard Problems

Part III: Simulation-Based Approximation

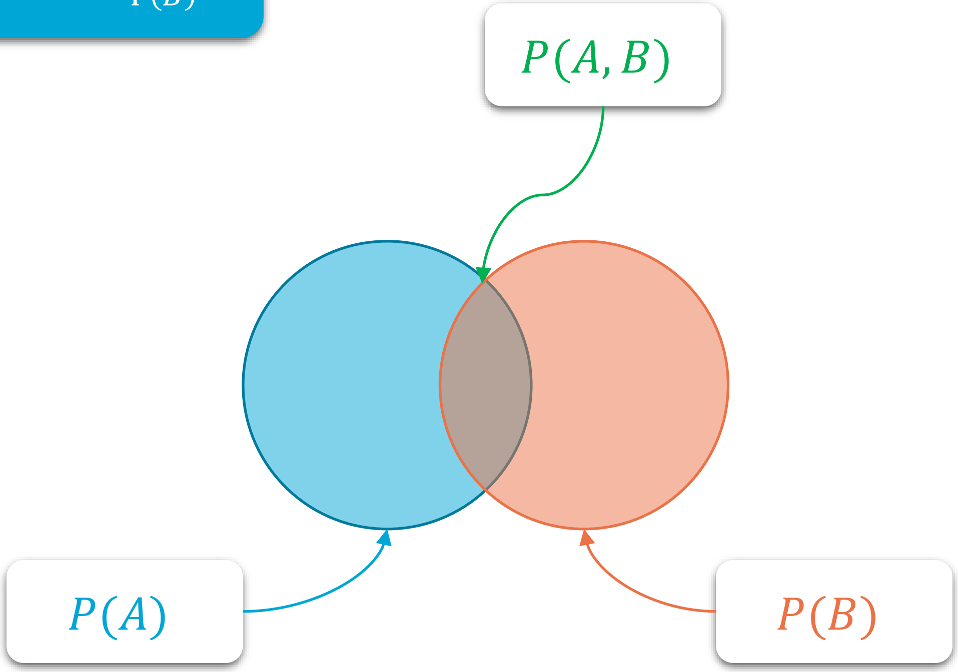
Lecture 3: Advanced Sampling Methods Continued

Instructor: Dr. Anna Lukina

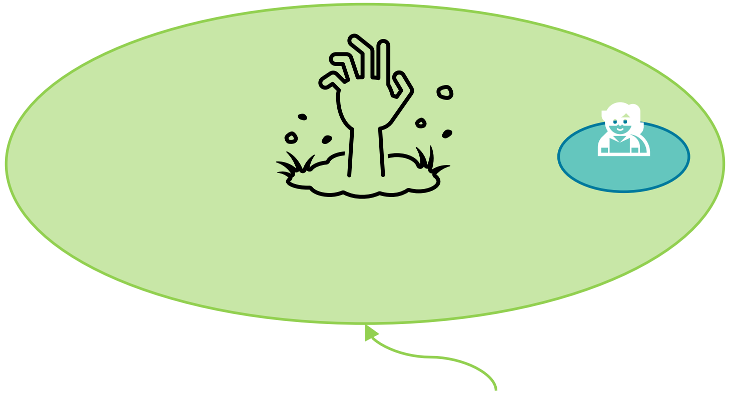
a.lukina@tudelft.nl

Conditional Probability Warm-up

$$P(A|B) = \frac{P(A,B)}{P(B)}$$



- 0.01% of population are immune to disease X.
- A new test for immunity has proved to be 99% accurate.
- You took this test, and it came back positive.
- What are the chances you are immune?

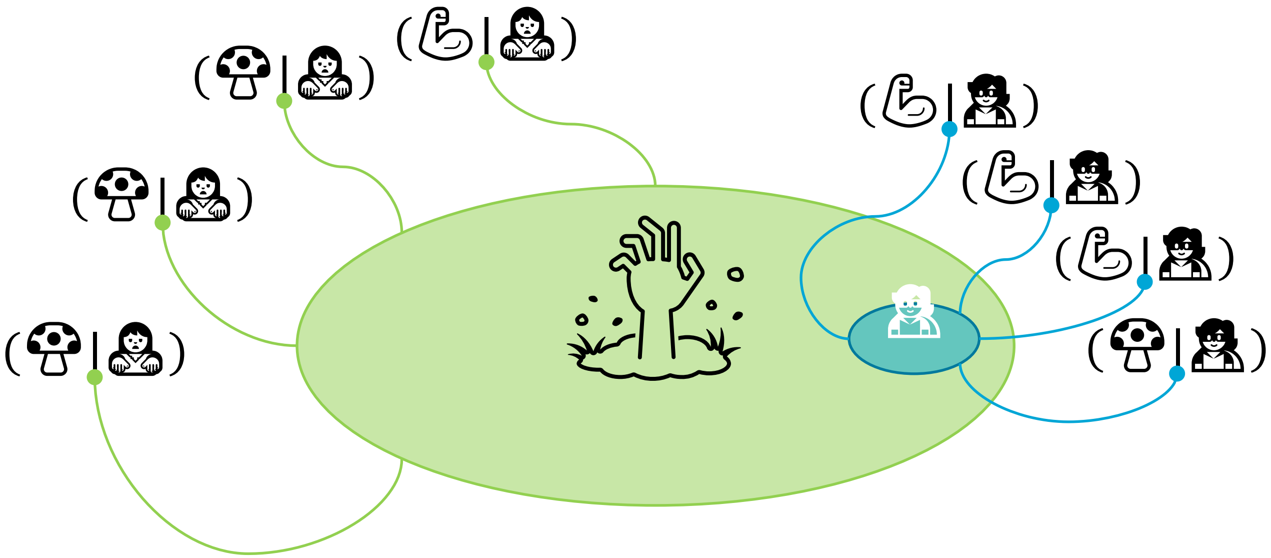


Conditional Probability Warm-up

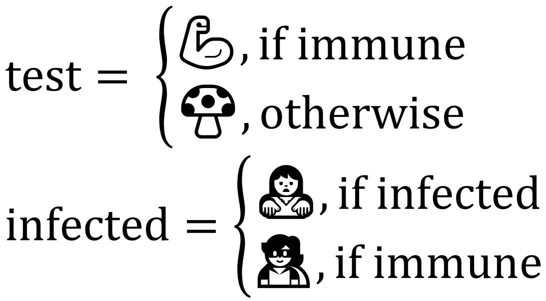
$$P(A|B) = \frac{P(A,B)}{P(B)}$$

What does it mean to be 99% accurate?

- 0.01% of population are immune to disease X.
- A new test for immunity has proved to be 99% accurate.
- You took this test, and it came back positive.
- What are the chances you are immune?



Define two random variables:



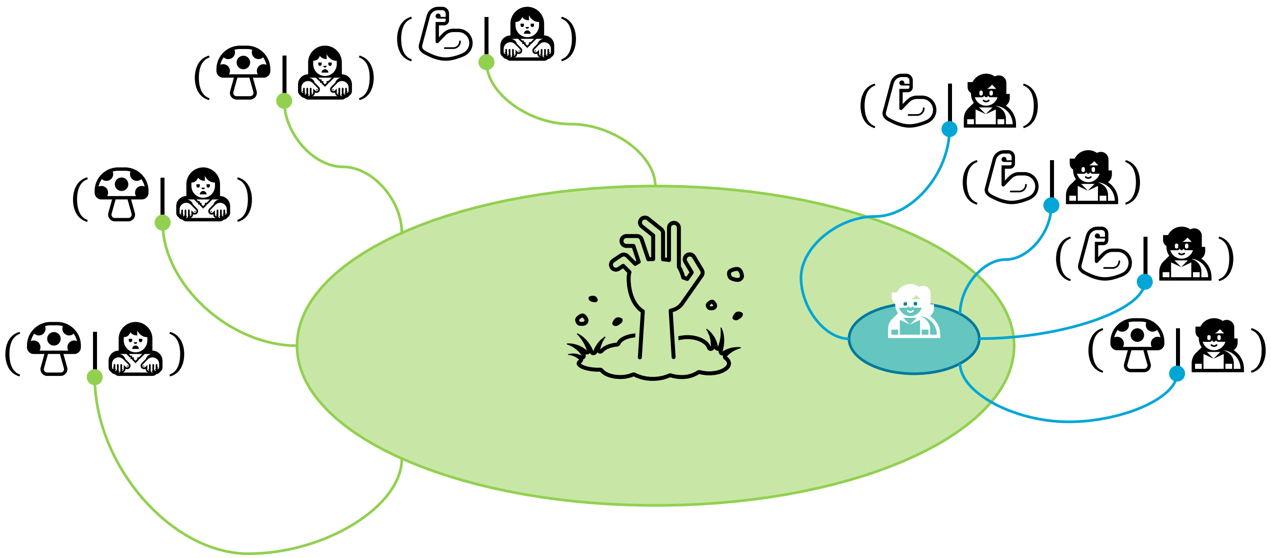
⚠ Spoiler on the next slide

Conditional Probability Warm-up

$$P(🧑‍🔬 \mid 💪) = \frac{P(🧑‍🔬, 💪)}{P(💪)}$$

What does it mean to be 99% accurate?

- 0.01% of population are immune to disease X.
- A new test for immunity has proved to be 99% accurate.
- You took this test, and it came back positive.
- What are the chances you are immune?



$$P(💪) = P(💪, 🧑‍🔬) + P(💪, 🙍‍♀️)$$

$$P(💪 \mid 🧑‍🔬) \cdot P(🧑‍🔬) \quad P(💪 \mid 🙍‍♀️) \cdot P(🙍‍♀️)$$

$$0.99 \quad 0.0001 \quad 0.01 \quad 0.9999$$

$$P(🧑‍🔬 \mid 💪) = \frac{0.000099}{0.010098} \approx 0.009 \rightarrow 0.9\%$$

Part III: Simulation-Based Approximation

Lecture 1: Estimating Uncertainty

Lecture 2: Advanced Sampling Methods

Lecture 3: Advanced Sampling Methods Continued

Part III: Simulation-Based Approximation

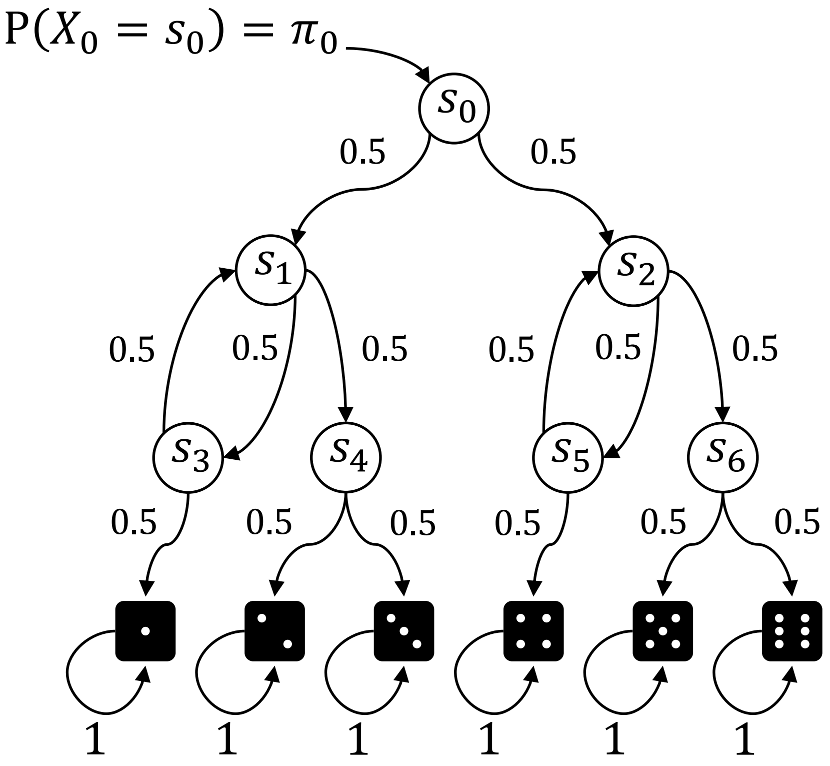
Lecture 1: Estimating Uncertainty

Lecture 2: Advanced Sampling Methods

Lecture 3: Advanced Sampling Methods Continued

1. Markov Chain Monte Carlo
   i. Markov Chains
   ii. Random Walk Metropolis
2. Rare event estimation
   1. Importance Sampling
   2. Importance Splitting

Markov Chains



Katoen, J. P. (2013). Model checking meets probability: A gentle introduction. In *Engineering dependable software systems* (pp. 177-205). IOS Press.

S is a countable set of states $s_i, i \in \mathbb{N}$.

We study a sequence of random variables $\mathbf{X} = (X_0, X_1, \ldots)$ that take values in S.

Markov property:
the past and future are conditionally independent, given the present.

$\boldsymbol{X} = (X_0, X_1, X_2, \ldots)$ is a Markov chain if

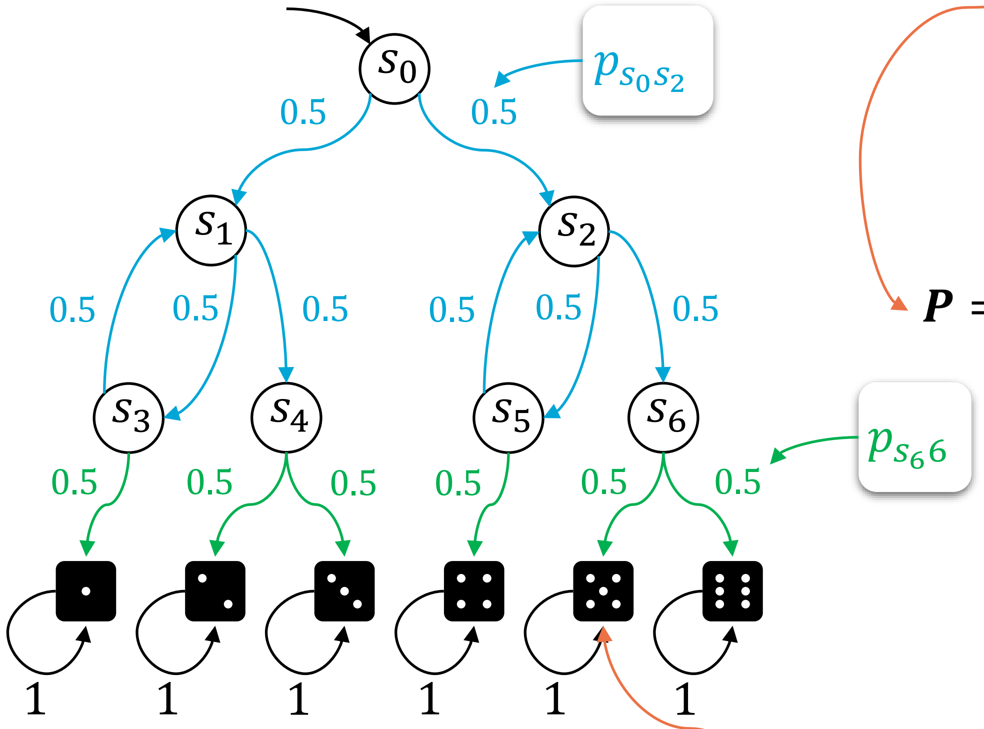
$$P(X_{n+k} = s \mid X_n, X_{n-1}, \ldots, X_0) = P(X_{n+k} = s \mid X_n)$$

for every $n, k \in \mathbb{N}$ and $s \in S$.

$$\Updownarrow$$

$$\mathrm{P}(X_0 = s_0, \ldots, X_n = s_n) = \pi_0 \cdot p_{s_0 s_1} \cdot \ldots \cdot p_{s_{n-1} s_n}.$$

Markov Chains



Absorbing state

Transition matrix

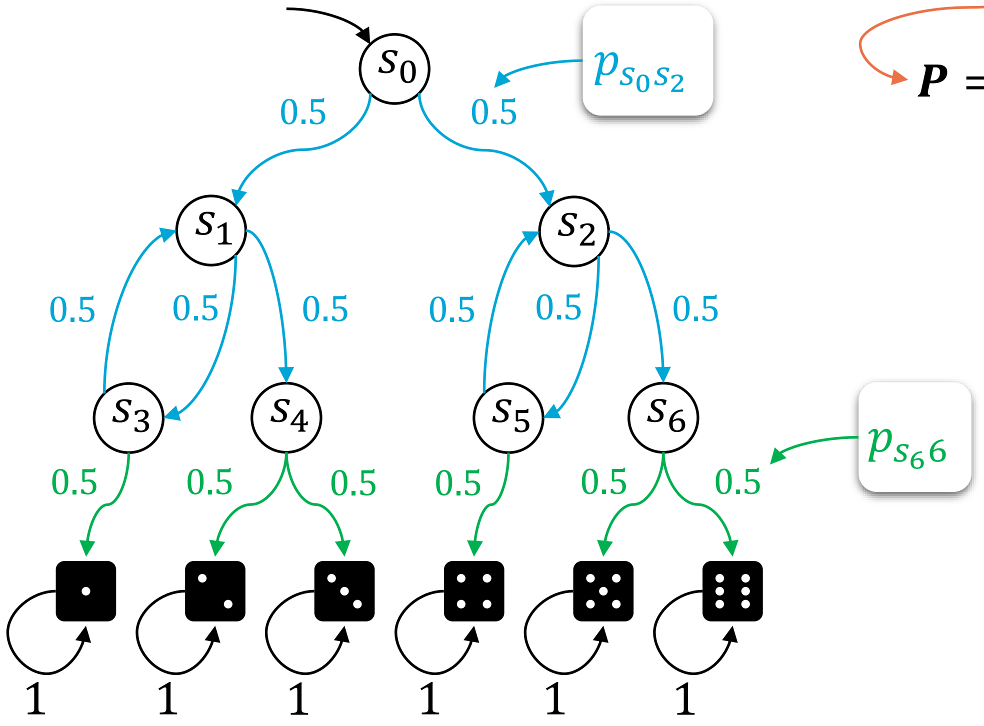
$$
\boldsymbol{P} = \begin{pmatrix}
p_{s_0 s_0} & p_{s_0 s_1} & p_{s_0 s_2} & p_{s_0 s_3} & p_{s_0 s_4} & p_{s_0 s_5} & p_{s_0 s_6} & 0 & & 0 \\
p_{s_1 s_0} & p_{s_1 s_1} & p_{s_1 s_2} & p_{s_1 s_3} & p_{s_1 s_4} & p_{s_1 s_5} & p_{s_1 s_6} & 0 & & 0 \\
p_{s_2 s_0} & p_{s_2 s_1} & p_{s_2 s_2} & p_{s_2 s_3} & p_{s_2 s_4} & p_{s_2 s_5} & p_{s_2 s_6} & 0 & & 0 \\
p_{s_3 s_0} & p_{s_3 s_1} & p_{s_3 s_2} & p_{s_3 s_3} & p_{s_3 s_4} & p_{s_3 s_5} & p_{s_3 s_6} & p_{s_3 1} & & p_{s_3 6} \\
p_{s_4 s_0} & p_{s_4 s_1} & p_{s_4 s_2} & p_{s_4 s_3} & p_{s_4 s_4} & p_{s_4 s_5} & p_{s_4 s_6} & p_{s_4 1} & \cdots & p_{s_4 6} \\
p_{s_5 s_0} & p_{s_5 s_1} & p_{s_5 s_2} & p_{s_5 s_3} & p_{s_5 s_4} & p_{s_5 s_5} & p_{s_5 s_6} & p_{s_5 1} & & p_{s_5 6} \\
p_{s_6 s_0} & p_{s_6 s_1} & p_{s_6 s_2} & p_{s_6 s_3} & p_{s_6 s_4} & p_{s_6 s_5} & p_{s_6 s_6} & p_{s_6 1} & & p_{s_6 6} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & & \cdots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & & 1
\end{pmatrix}
$$

$$\forall s_i \in S \sum_{s_j \in S} p_{s_i s_j} = 1$$

$p_{s_i s_j} \geq 0$ and each row is a distribution over S

Katoen, J. P. (2013). Model checking meets probability: A gentle introduction. In *Engineering dependable software systems* (pp. 177-205). IOS Press.

Markov Chains



Transition matrix

$$P = \begin{pmatrix} A & B \\ 0 & I \end{pmatrix}$$

$$\mathrm{P}(X_1 = s_j | X_0 = s_i) = p_{s_i s_j}$$

$$\mathrm{P}(X_1 = s_j) = \sum_i \mathrm{P}(X_1 = s_j, X_0 = s_i) =$$

$$= \sum_i \mathrm{P}(X_1 = s_j | X_0 = s_i) \cdot \mathrm{P}(X_0 = s_i) =$$

$$= \sum_i \pi_i p_{s_i s_j} = \{j = 1\} = \frac{1}{2}$$

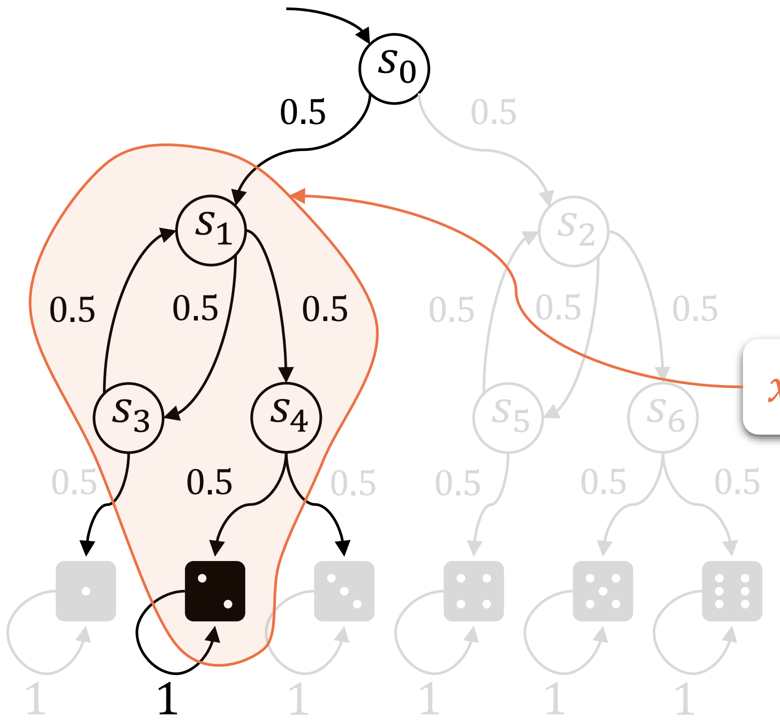
$$\mathrm{P}(X_2 = s_j | X_1 = s_i) = p_{s_i s_j}$$

$$\mathrm{P}(X_2 = s_j) = \sum_{i.k} \mathrm{P}(X_2 = s_j, X_1 = s_k, X_0 = s_i) =$$

$$= \sum_{i.k} \pi_i p_{s_i s_k} p_{s_k s_j} = (\pi \boldsymbol{P}^2)_j = \{j = 3\} = \frac{1}{4}$$

Katoen, J. P. (2013). Model checking meets probability: A gentle introduction. In *Engineering dependable software systems* (pp. 177-205). IOS Press.

Markov Chains



Transition matrix for eventually reaching 2 on the dice

$$P = \begin{pmatrix} A & B \\ 0 & I \end{pmatrix} \Rightarrow (A)_2 = \begin{pmatrix} p_{s_0 s_0} & p_{s_0 s_1} & p_{s_0 s_3} & p_{s_0 s_4} \\ p_{s_1 s_0} & p_{s_1 s_1} & p_{s_1 s_3} & p_{s_1 s_4} \\ p_{s_3 s_0} & p_{s_3 s_1} & p_{s_3 s_3} & p_{s_3 s_4} \\ p_{s_4 s_0} & p_{s_4 s_1} & p_{s_4 s_3} & p_{s_4 s_4} \end{pmatrix}, x = \begin{pmatrix} x_{s_0} \\ x_{s_1} \\ x_{s_3} \\ x_{s_4} \end{pmatrix}$$

$x_{s_1} = \mathrm{P}(s_1 \to 2)$

$$\mathrm{P}(X_k = 2) = \sum_{i_0,\dots,i_{k-1}} \pi_{i_0} p_{s_{i_0} s_{i_1}} \dots p_{s_{i_{k-1}} 2} = \left(P^k\right)_2$$

$$S = I + (P)_2 + (P^2)_2 + \cdots$$
$$S - I = (P)_2^T S (P)_2$$
$$(P)_2^T S (P)_2 - S = -I$$

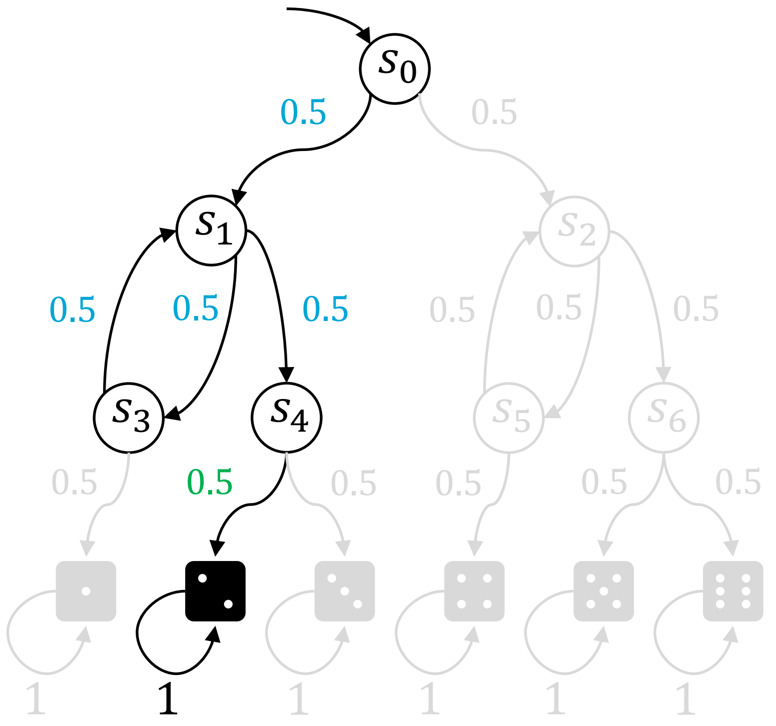
There is a unique S iff $x_{k+1} = (P)_2 x_k$ is stable (Lyapunov)

Katoen, J. P. (2013). Model checking meets probability: A gentle introduction. In *Engineering dependable software systems* (pp. 177-205). IOS Press.

Markov Chains

Eventually $= \bigcup_{k=0}^{\infty} s_0(s_1 s_3)^k s_1 s_4 2$

Reachability probabilities are unique solutions of a given linear equation system for reaching .



$$\begin{pmatrix} x_{s_0} \\ x_{s_1} \\ x_{s_3} \\ x_{s_4} \end{pmatrix} = \begin{pmatrix} 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} x_{s_0} \\ x_{s_1} \\ x_{s_3} \\ x_{s_4} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1/2 \end{pmatrix}$$

$(\boldsymbol{A})_2$ is stable since its eigenvalues are < 1

$$\sum_{k \in \mathbb{N}} \left(\boldsymbol{A}^k\right)_2 = (\boldsymbol{I} - (\boldsymbol{A})_2)^{-1}$$

Then $\boldsymbol{x}$ is a unique solution of $(\boldsymbol{I} - (\boldsymbol{A})_2)\boldsymbol{x} = (\boldsymbol{B})_2$.

Katoen, J. P. (2013). Model checking meets probability: A gentle introduction. In *Engineering dependable software systems* (pp. 177-205). IOS Press.

Markov Chain Monte Carlo

$$P(\theta|X) = \frac{P(X|\theta) \cdot P(\theta)}{\cancel{P(X)}}$$

How do we do Monte Carlo if we can't draw independent samples from posterior?

Sample from $P(X|\theta) \cdot P(\theta)$

Dependent sampling: Random Walk Metropolis

In Random Walk Metropolis, Markov chain must be reversible: $P(\theta_{t+1}, \theta_t) = P(\theta_t, \theta_{t+1})$

$$\begin{cases} \theta_t = \theta_t', \text{ if } r > u \sim U(0,1) \\ \theta_t = \theta_{t-1}, \text{otherwise} \end{cases}$$

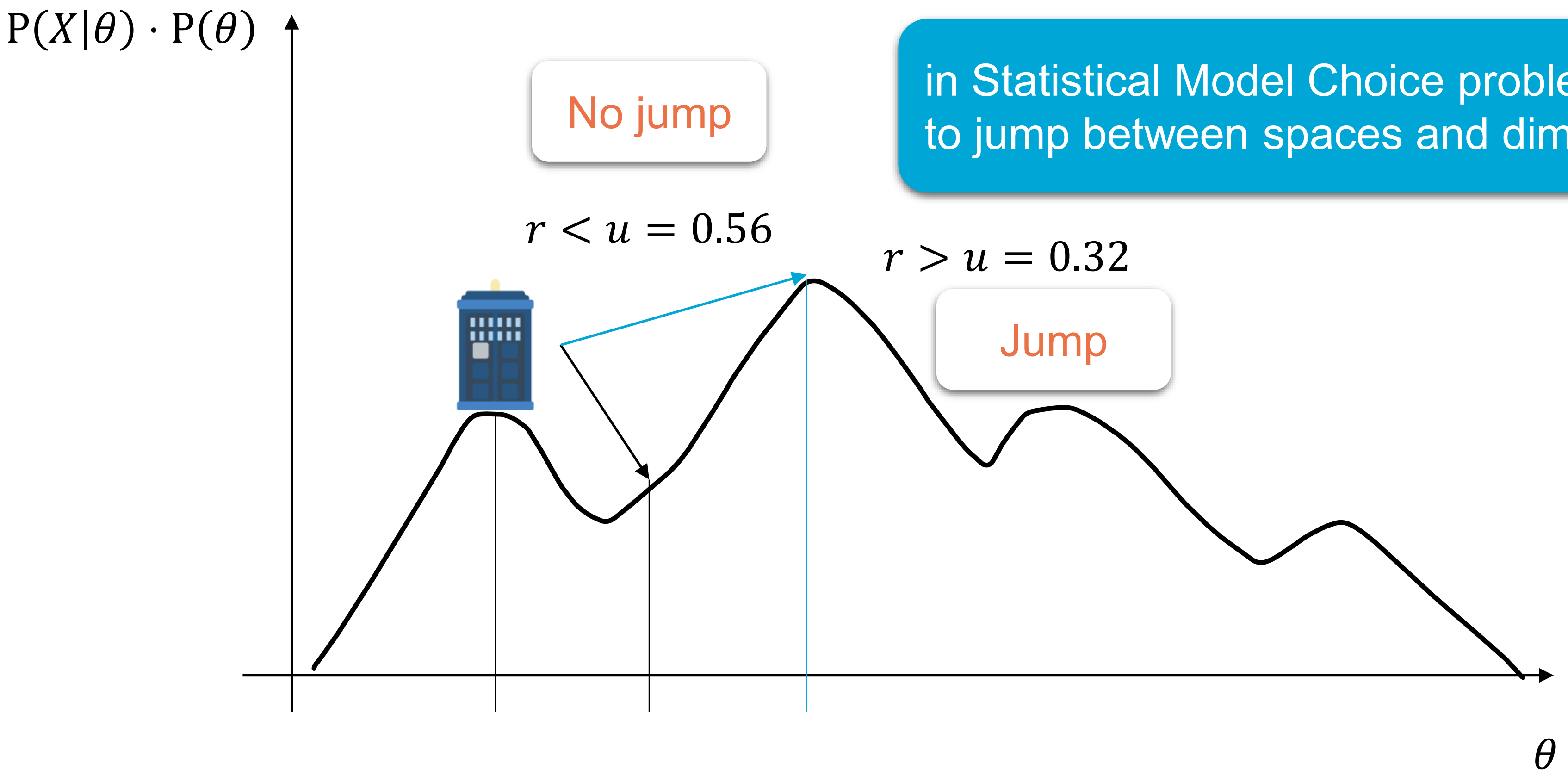
Algorithm:

Draw $\theta_0 \sim \pi(\theta)$, an arbitrary distribution

For each iteration t:

1. Propose a new place to step on $\theta_t' \sim \mathcal{N}(\theta_{t-1}, \sigma)$
2. Compute $r = \frac{P(X|\theta_t') \cdot P(\theta_t')}{P(X|\theta_{t-1}) \cdot P(\theta_{t-1})}$

Markov Chain Monte Carlo



in Statistical Model Choice problems, it is often necessary to jump between spaces and dimensions

http://www.statslab.cam.ac.uk/~rrw1/markov/M.pdf

Random K-SAT Counting

- K-SAT counting problems for $K \geq 2$ are NP-hard.
- A K-RSAT(m, n) contains m clauses of length K corresponding to n variables.
- Each clause is drawn uniformly from the set of $\binom{n}{K}2^K$ clauses, independently of the other clauses.
- $\beta = \frac{m}{n}$ – clause density is crucial.

INTERESTING ARE RARE

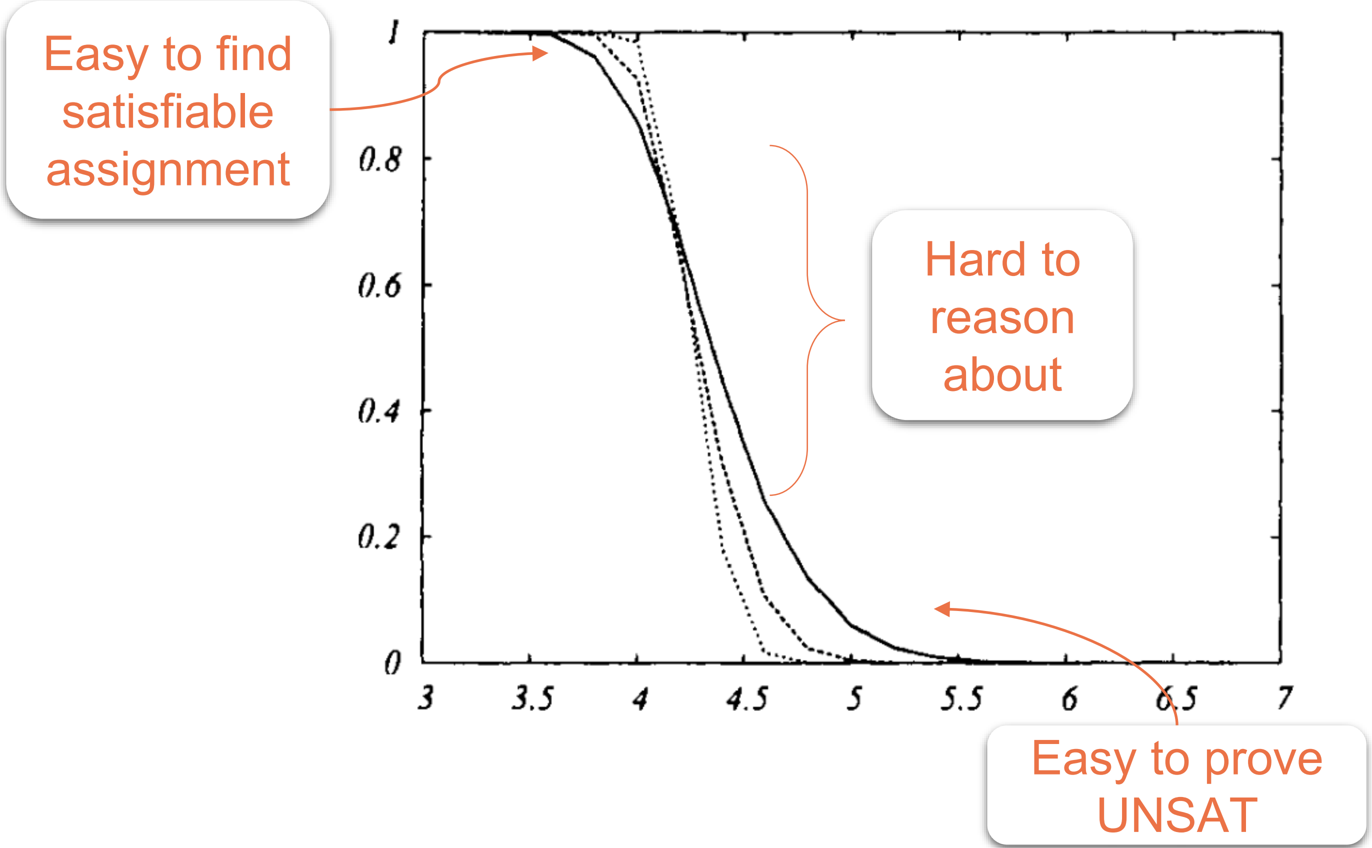
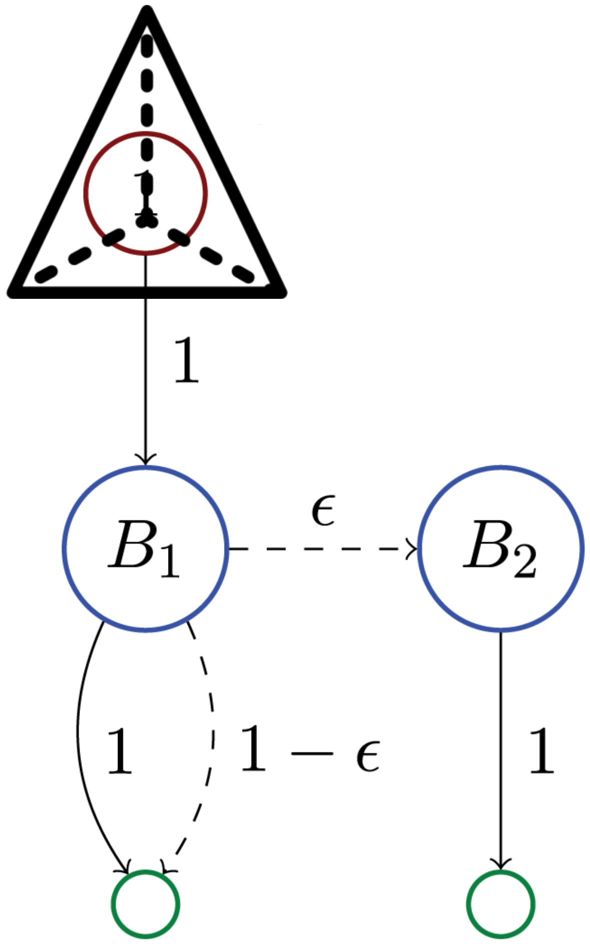


Figure 9.2 The probability that a K-RSAT(m, n) problem has a solution as a function of the clause density β for $n = 50$, 100, and 200.

Rubinstein RY, Kroese DP. Simulation and the Monte Carlo Method. 2nd ed. New York: John Wiley & Sons; 2007.

Multi-Dimensional Sampling



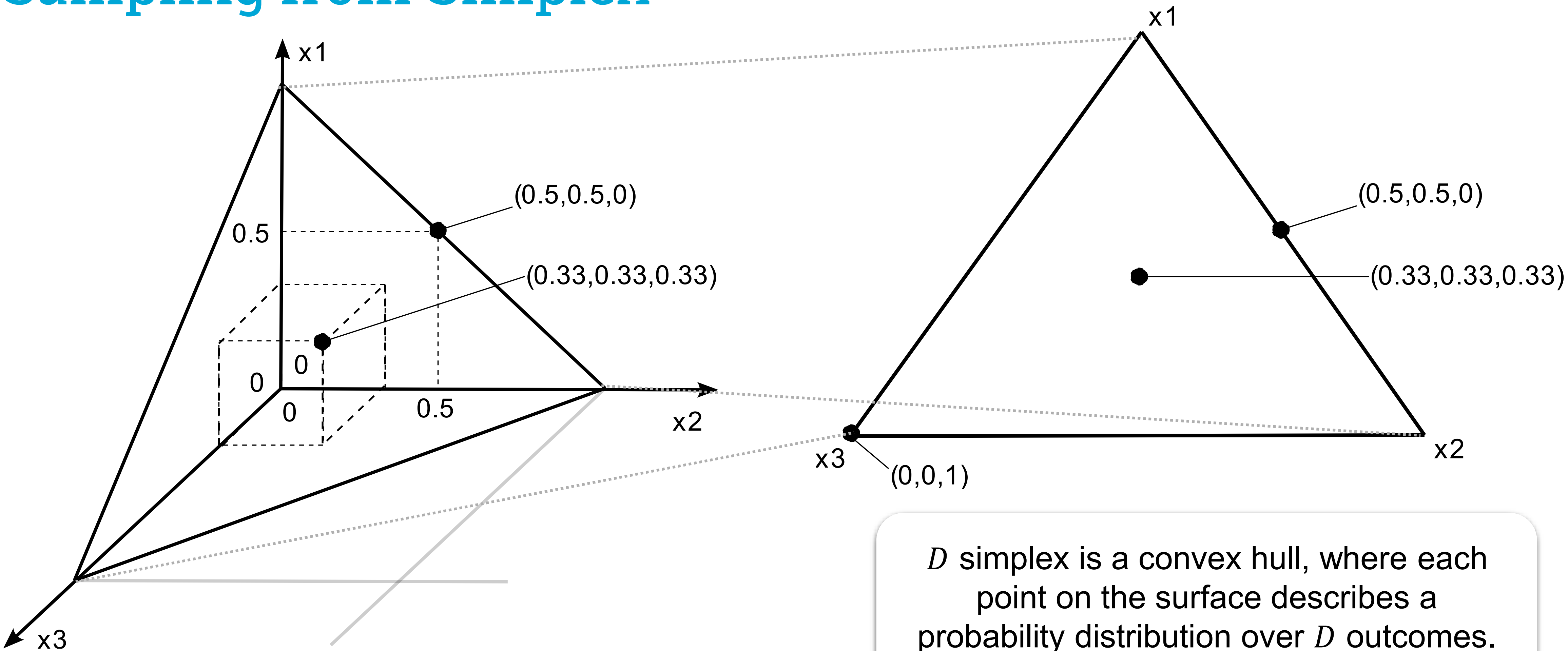
Sample from

$$X = \{ (x_1, x_2, \dots, x_D) \mid 0 \le x_i \le 1,$$
$$x_1 + x_2 + \dots + x_D = 1\}.$$

D is the dimension of the simplex.

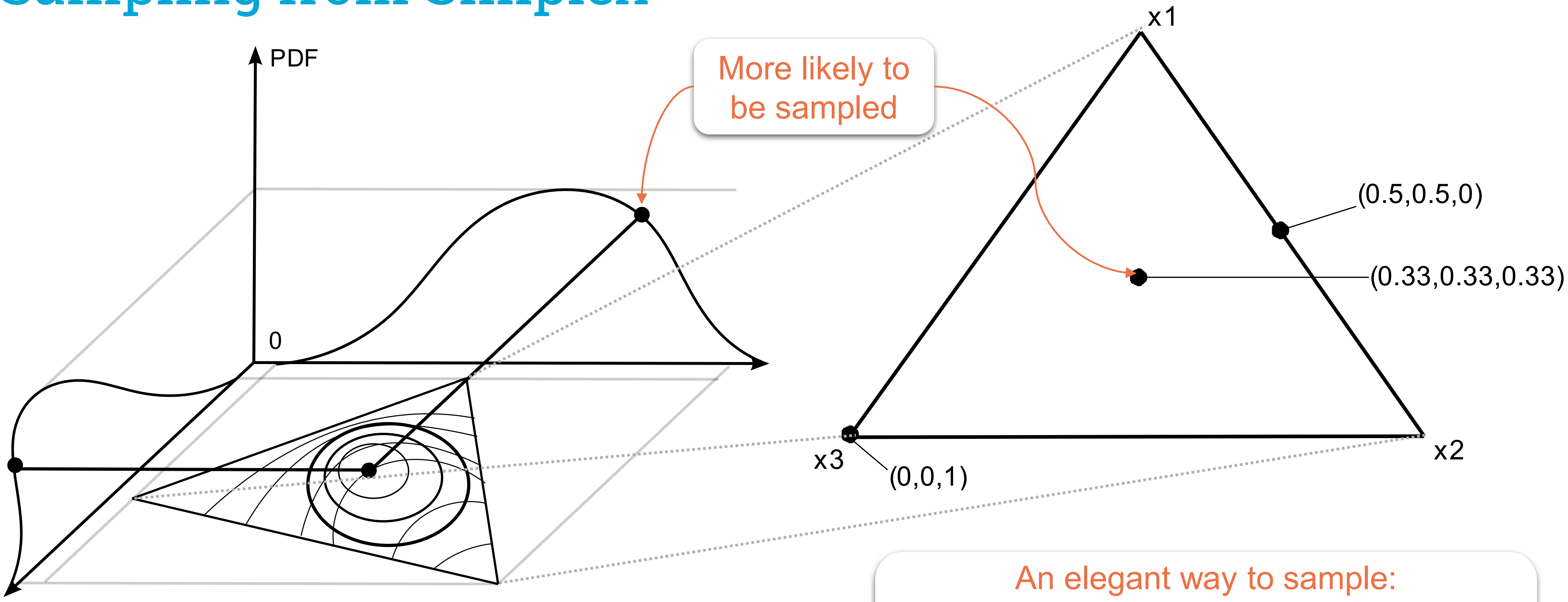
D simplex is a convex hull, where each point on the surface describes a probability distribution over D outcomes.

Sampling from Simplex



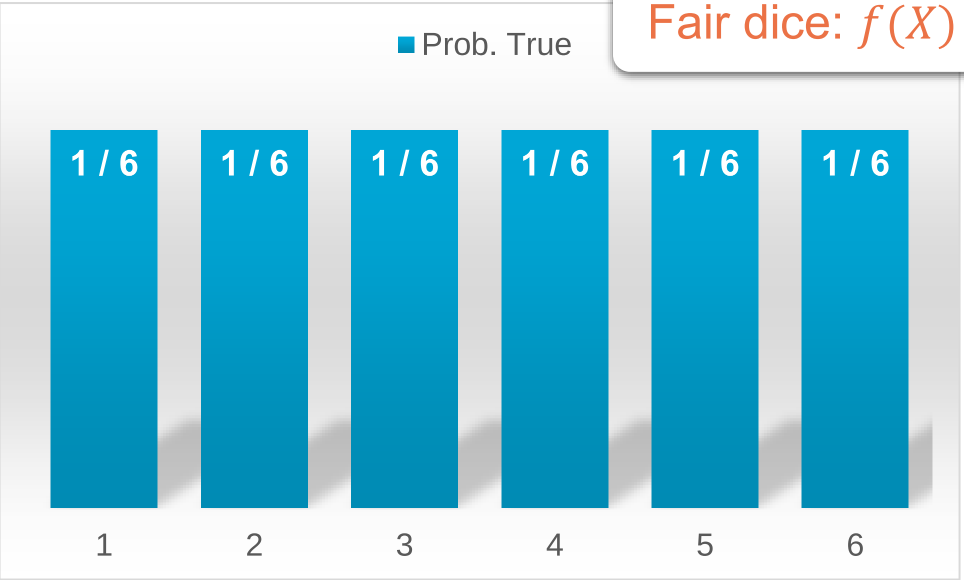
D simplex is a convex hull, where each point on the surface describes a probability distribution over D outcomes.

Sampling from Simplex

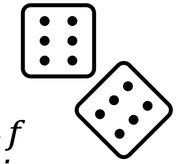


An elegant way to sample:
Sample X from $[0,1]$ uniformly and return $-\log(X)$. Take D samples, and then normalize.

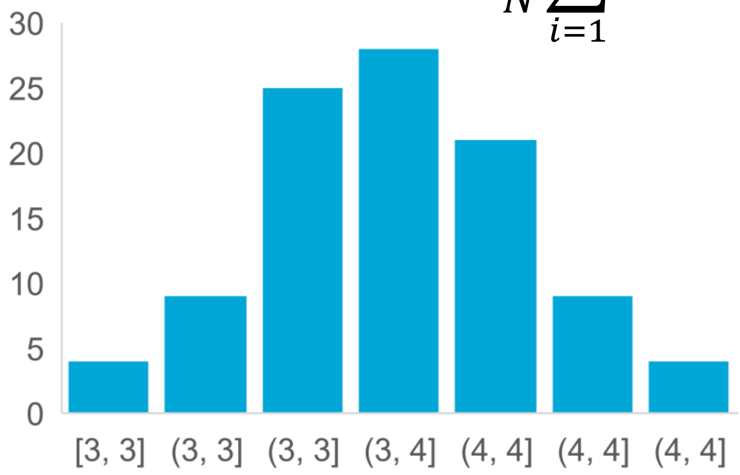
Importance Sampling



Fair dice: $f(X)$



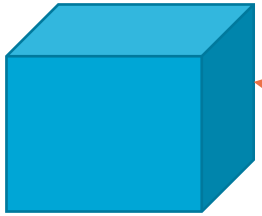
mean $\approx \frac{1}{N}\sum_{i=1}^{N} X_i^f$



If we have $f(X)$ we can easily calculate:

$$E_f[X] = \sum_x x \cdot f(x) = 1 \cdot \frac{1}{6} + \cdots + 6 \cdot \frac{1}{6} = 3\frac{1}{2}$$

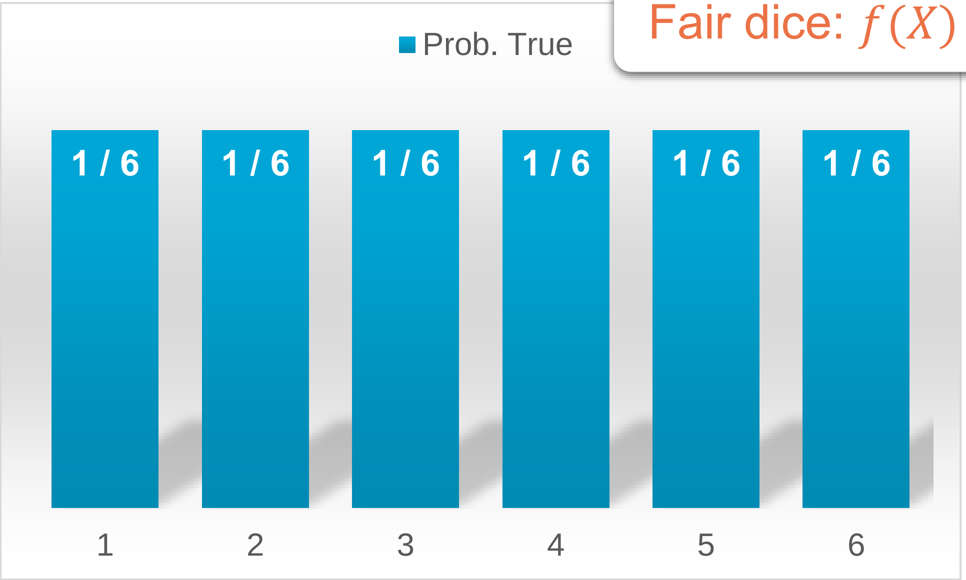
Estimate mean by shaking the box and observing the result



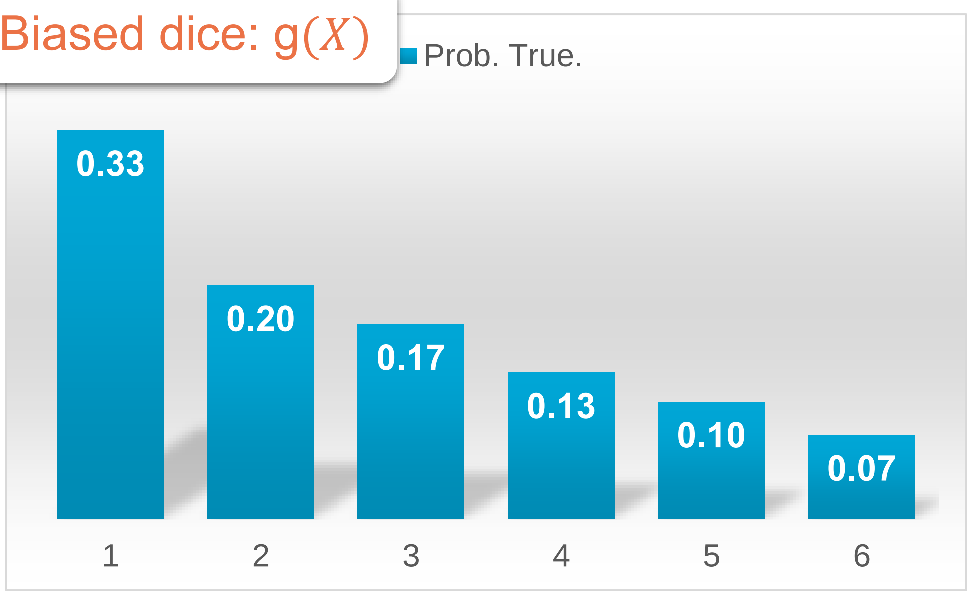
Do not see fair dice

Importance Sampling

Fair dice: $f(X)$



Biased dice: g(X)



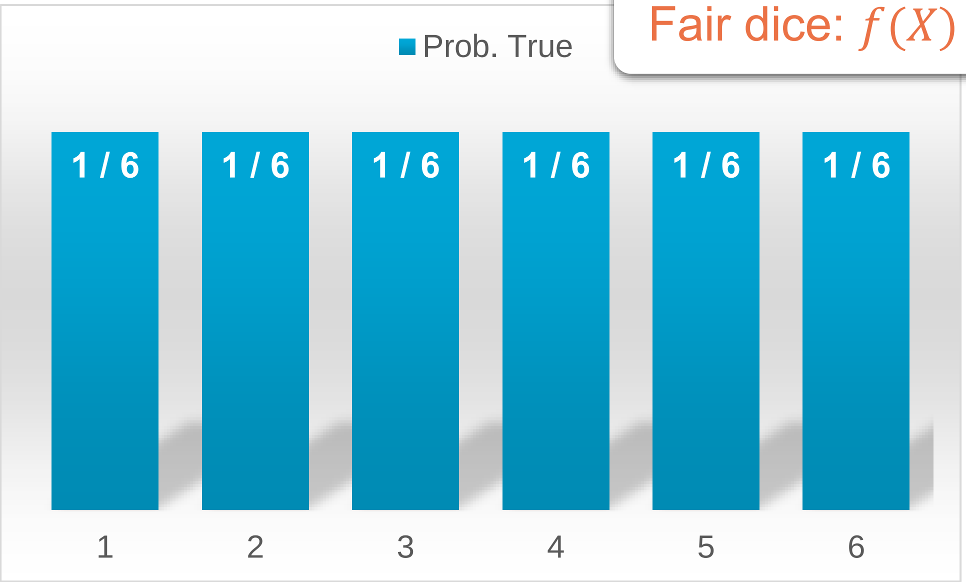
Estimate $\mathrm{E}_g[X]$ if we can't simulate $g(X)$ but only $f(X)$

If we have $g(X)$ we can easily calculate:

$$\mathrm{E}_g[X] = \sum_x x \cdot g(x) \cong 2.66$$

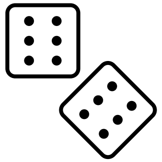
$$\approx \frac{1}{N} \sum_{i=1}^{N} X_i^g$$

Importance Sampling



Fair dice: $f(X)$

Estimate $\mathrm{E}_g[X]$ if we can't simulate $g(X)$ but only $f(X)$



$$\mathrm{E}_g[X] = \sum_x x \cdot g(x)$$

$$= \sum_x x \cdot \frac{g(x)}{f(X)} \cdot f(X)$$

$$= \mathrm{E}_f\left[X \cdot \frac{g(X)}{f(X)}\right]$$

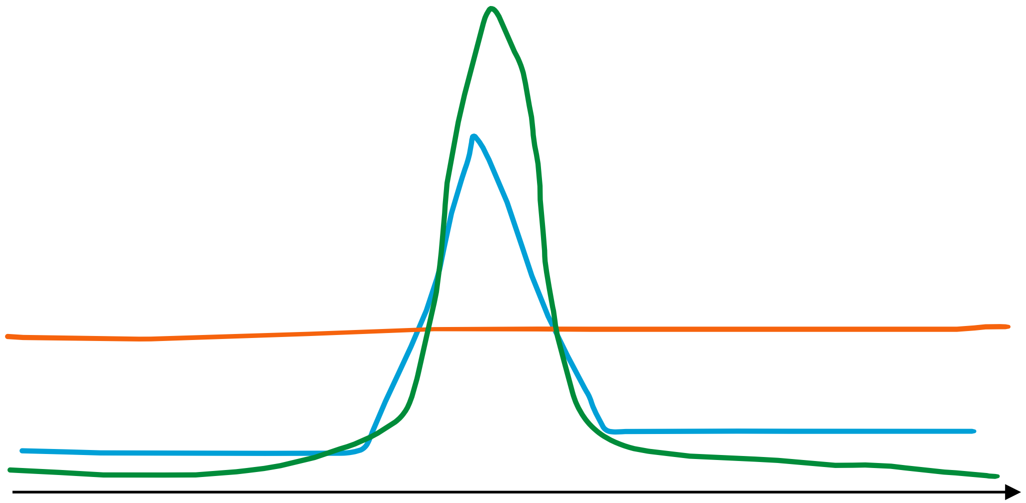
Throw fair dice and approximate

$$\approx \frac{1}{N} \sum_{i=1}^{N} X_i^f \cdot \boxed{\frac{g(X_i^f)}{f(X_i^f)}}$$

Weight ω_i

To see simulation of this procedure: https://youtu.be/V8f8ueBc9sY

Importance Sampling Variance



Choose $g(X)$ close to $f(X)$ 🤔

$$\mu = \mathrm{E}_g[X] = ?$$

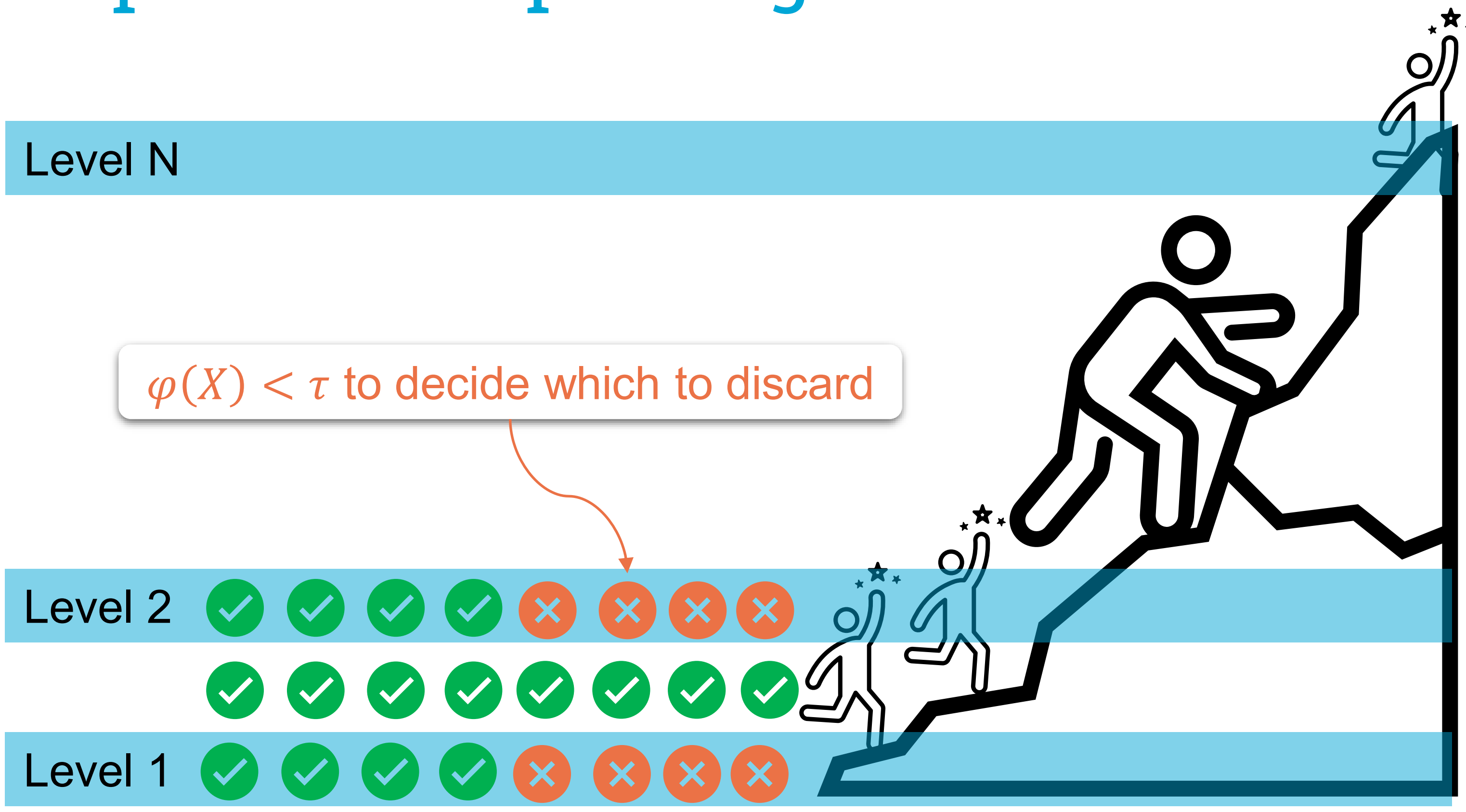
Will have zero variance

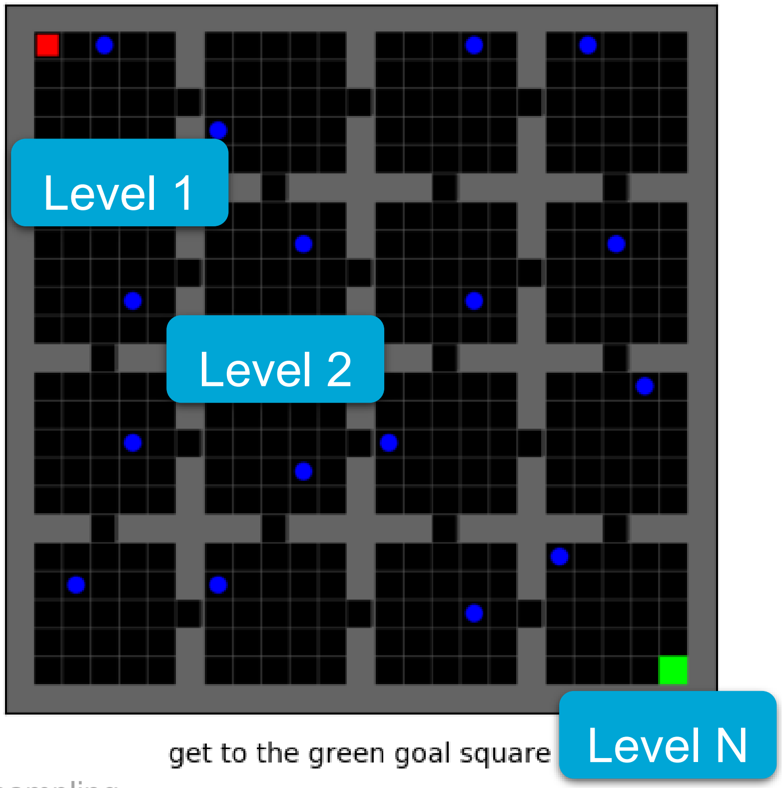
$$f(X) = \frac{g(X)X}{\mu}$$

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} X_i^f \cdot \frac{g(X_i^f)\mu}{g(X_i^f)X_i^f} = \frac{1}{N}N\mu = \mu$$

$$\mathrm{Var}(\hat{\mu}) = 0$$ 👍

Importance Splitting





Picture of the grid is taken from MSc thesis by C. van Rijn, TU Delft 2023

Kahn, H. and Harris, T. E. (1951). Estimation of particle transmission by random sampling. https://dornsifecms.usc.edu/assets/sites/520/docs/kahnharris.pdf

Ma, H., Blom, H.A.P. Random Assignment Versus Fixed Assignment in Multilevel Importance Splitting for Estimating Stochastic Reach Probabilities. Methodol Comput Appl Probab 24, 2313–2338 (2022). https://doi.org/10.1007/s11009-021-09892-4

Algorithms for NP-Hard Problems

Part III: Simulation-Based Approximation

Lecture 4: Optimization and Planning

Instructor: Dr. Anna Lukina

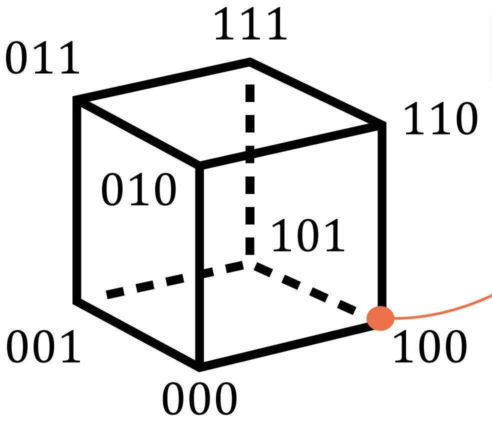
a.lukina@tudelft.nl

TU Delft

https://xkcd.com/287/

The Easiest of Hard Problems

Classic Monte-Carlo estimator



$\boldsymbol{x} \in \{0,1\}^m$

Random experiment:

$$X_i = \begin{cases} 2^m, \text{if } \boldsymbol{a} \cdot \boldsymbol{x} \leq b \\ 0, \text{otherwise} \end{cases}$$

$$\mathrm{E}X_i = \frac{|\Omega|}{2^m} \cdot 2^m$$

In practice:

$$\boldsymbol{a} = (1, \dots, 1), b = \frac{m}{3}$$

Yields mean of 0

Problem:

- Given m items of size $a_i, i = 1, \dots, m$.
- A knapsack of weight limit $b \in \mathbb{N}$.
- Estimate the number of combination of items that can be fit into the knapsack.

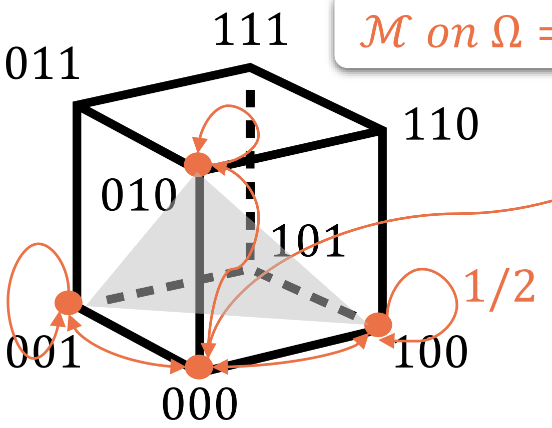
Model:

- $\boldsymbol{a} = (a_0, \dots, a_{m-1}) \in \mathbb{N}^m$
- Estimate the number of vectors $\boldsymbol{x} \in \{0,1\}^m$ satisfying $\boldsymbol{a} \cdot \boldsymbol{x} = \sum_{i=0}^{m-1} a_i x_i \leq b$.
- “Knapsack Solutions”: $\Omega = \{\boldsymbol{x} \in \{0,1\}^m : \boldsymbol{a} \cdot \boldsymbol{x} \leq b\}$

Mark Jerrum and Alistair Sinclair. 1996. The Markov chain Monte Carlo method: an approach to approximate counting and integration. Approximation algorithms for NP-hard problems. PWS Publishing Co., USA, 482–520.

The Easiest of Hard Problems

Markov Chain Monte Carlo estimator



$\mathcal{M}$ on $\Omega = \{\boldsymbol{x} \in \{0,1\}^m : \boldsymbol{a} \cdot \boldsymbol{x} \leq b\}$

A random walk on a Boolean hypercube $\Omega = \{\boldsymbol{x} \in \{0,1\}^m : \boldsymbol{a} \cdot \boldsymbol{x} \leq b\}$

$$\mathrm{E}X_i = \frac{|\Omega|}{2^m} \cdot 2^m$$

In practice:

$$\boldsymbol{a} = (1, \ldots, 1), b = \frac{m}{3}$$

Simulation:

- Start with a feasible solution $\boldsymbol{x} \in \Omega$.
- $\boldsymbol{x}' = \begin{cases} \boldsymbol{x} \in \Omega, \frac{1}{2} \\ (x_0, \ldots, 1 - x_i, \ldots, x_{m-1}), i \sim U\{0, \ldots, m-1\}, \frac{1}{2} \end{cases}$
- Random variable $Y = \boldsymbol{x}'$ if $\boldsymbol{a} \cdot \boldsymbol{x}' \leq b$
- $\mathcal{M}$ is ergodic and aperiodic Markov chain on Ω.
- Its stationary distribution is uniform over Ω.
- Works for arbitrary $\boldsymbol{a}, b$.

Mark Jerrum and Alistair Sinclair. 1996. The Markov chain Monte Carlo method: an approach to approximate counting and integration. Approximation algorithms for NP-hard problems. PWS Publishing Co., USA, 482–520.
B. Morris and A.J. Sinclair. Random walks on truncated cubes and sampling 0-1 knapsack solutions. SIAM J. Comput., 34(1):195-226, 2004.

Part III: Simulation-Based Approximation

Lecture 1: Estimating Uncertainty

Lecture 2: Advanced Sampling Methods

Lecture 3: Advanced Sampling Methods Continued

Lecture 4: Optimization and Planning

Part III: Simulation-Based Approximation

Lecture 1: Estimating Uncertainty

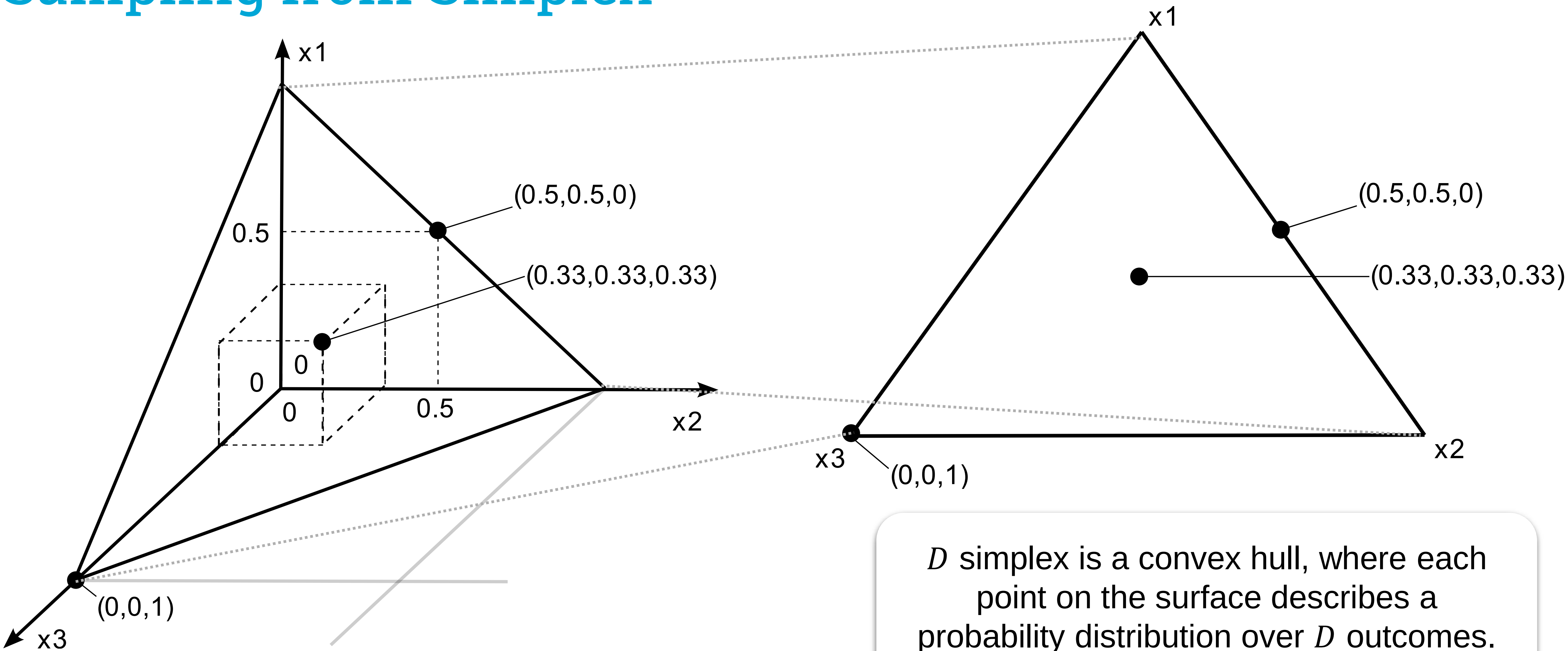
Lecture 2: Advanced Sampling Methods

Lecture 3: Advanced Sampling Methods Continued

Lecture 4: Optimization and Planning

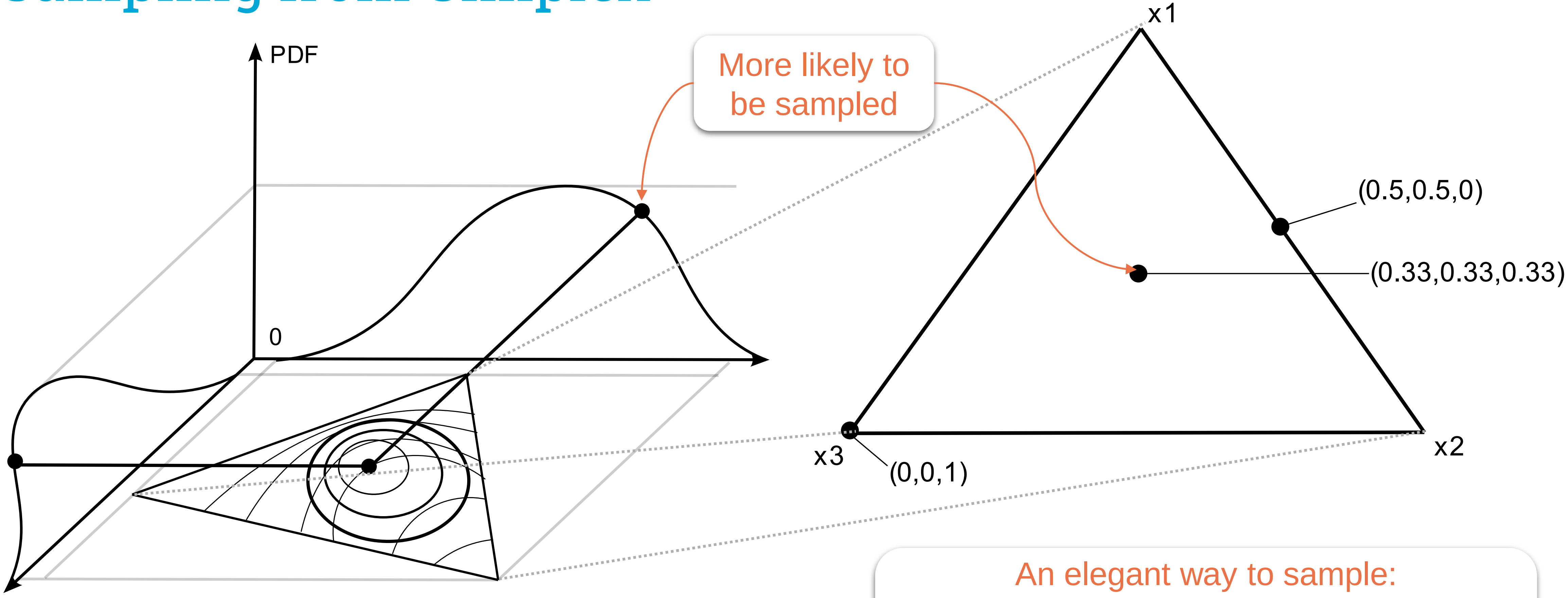
1. Rare event estimation
   i. Importance Sampling
   ii. Importance Splitting
2. Simulated Annealing
3. Overview

Sampling from Simplex



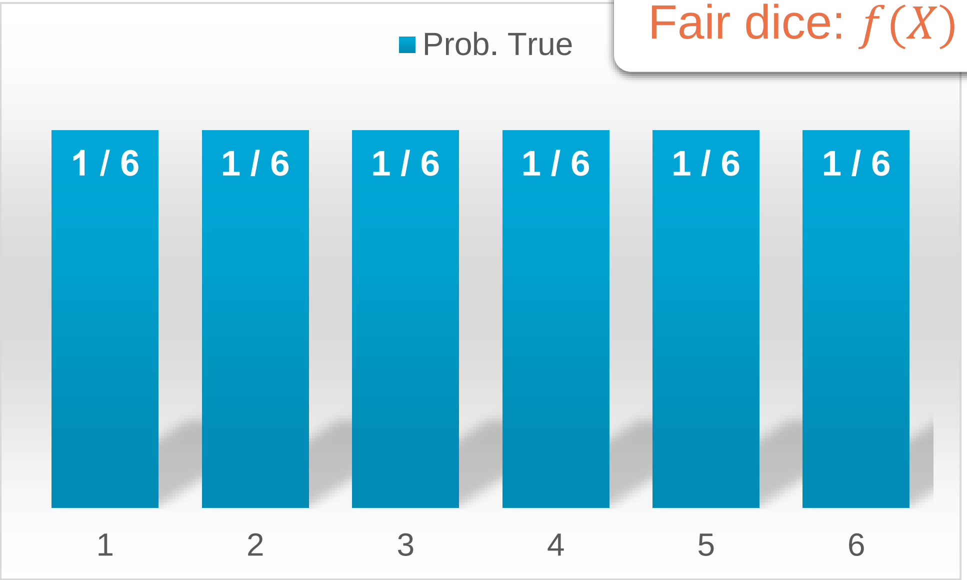
D simplex is a convex hull, where each point on the surface describes a probability distribution over D outcomes.

Sampling from Simplex



An elegant way to sample:
Sample X from $[0,1]$ uniformly and return $-\log(X)$. Take D samples, and then normalize.

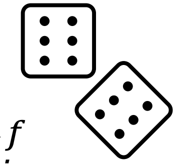
Importance Sampling



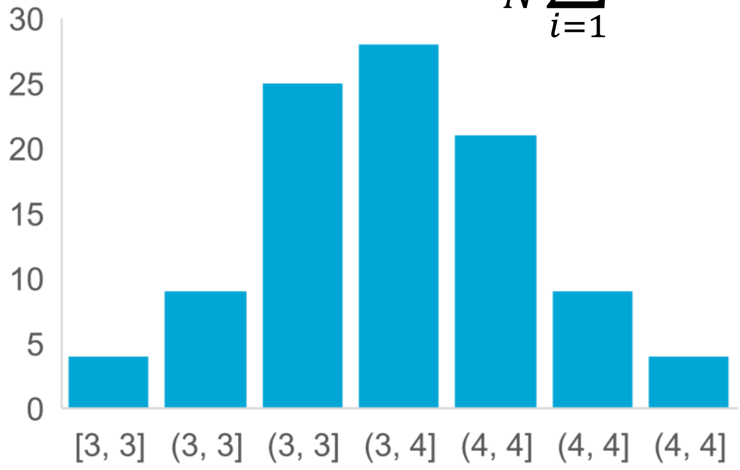
Fair dice: $f(X)$

If we have $f(X)$ we can easily calculate:

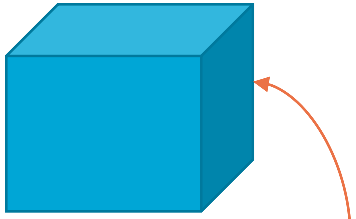
$$\mathrm{E}_f[X] = \sum_x x \cdot f(x) = 1 \cdot \frac{1}{6} + \cdots + 6 \cdot \frac{1}{6} = 3\frac{1}{2}$$



mean $\approx \frac{1}{N}\sum_{i=1}^{N} X_i^f$



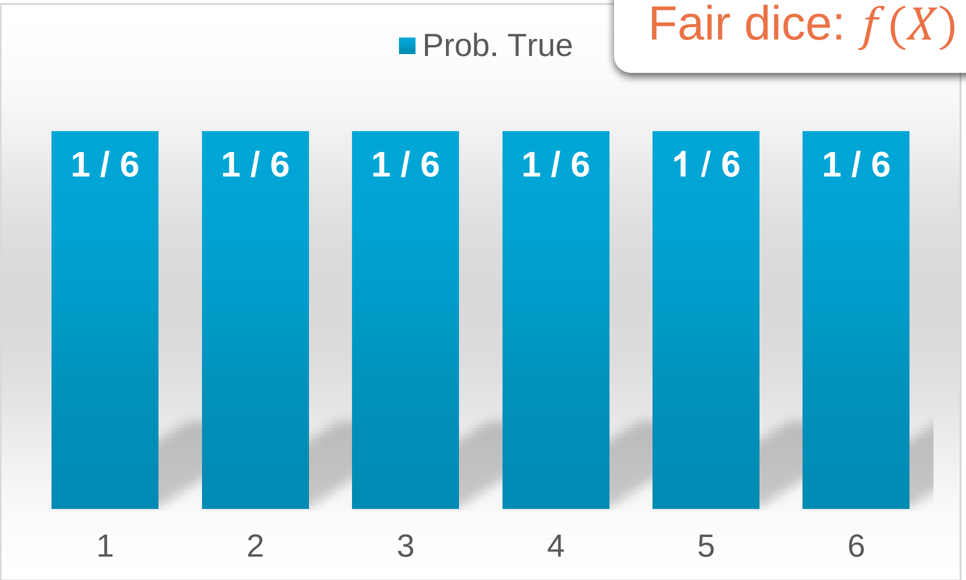
Estimate mean by shaking the box and observing the result



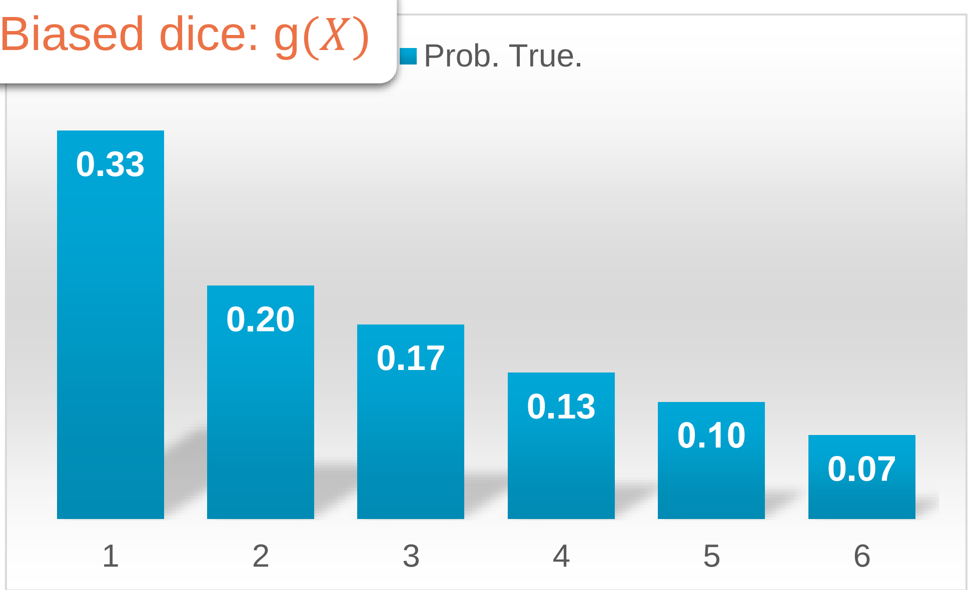
Do not see fair dice

Importance Sampling

Fair dice: $f(X)$



Biased dice: $g(X)$



Estimate $\mathrm{E}_g[X]$ if we can't simulate $g(X)$

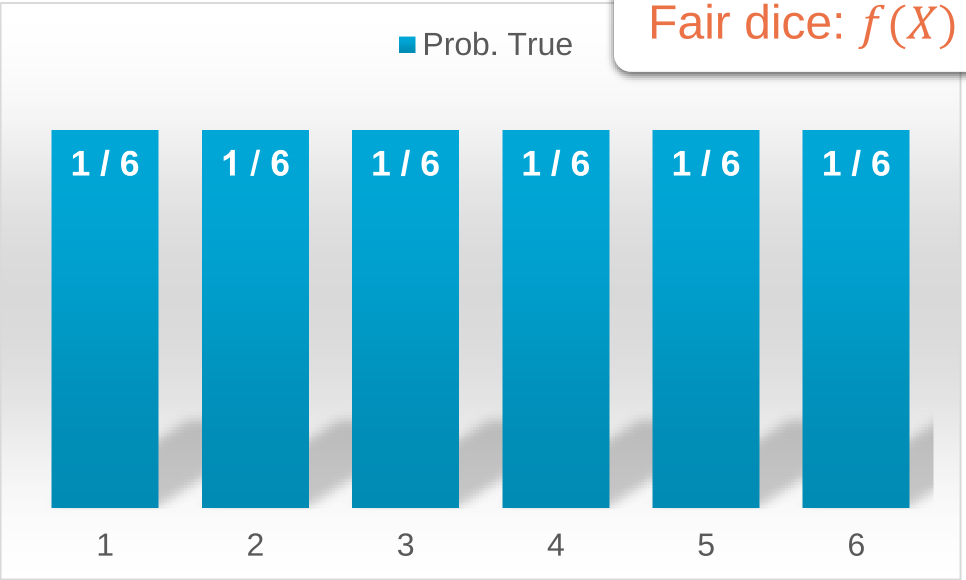
(can't draw samples X_i^g) but only $f(X)$

If we have $g(X)$ we can easily calculate:

$$\mathrm{E}_g[X] = \sum_x x \cdot g(x) \cong 2.66$$

$$\approx \frac{1}{N} \sum_{i=1}^{N} X_i^g$$

Importance Sampling



Fair dice: $f(X)$

Estimate $\mathrm{E}_g[X]$ if we can't simulate $g(X)$ but only $f(X)$

$$\mathrm{E}_g[X] = \sum_x x \cdot g(x)$$

$$= \sum_x x \cdot \frac{g(x)}{f(X)} \cdot f(X)$$

$$= \mathrm{E}_f\left[X \cdot \frac{g(X)}{f(X)}\right]$$

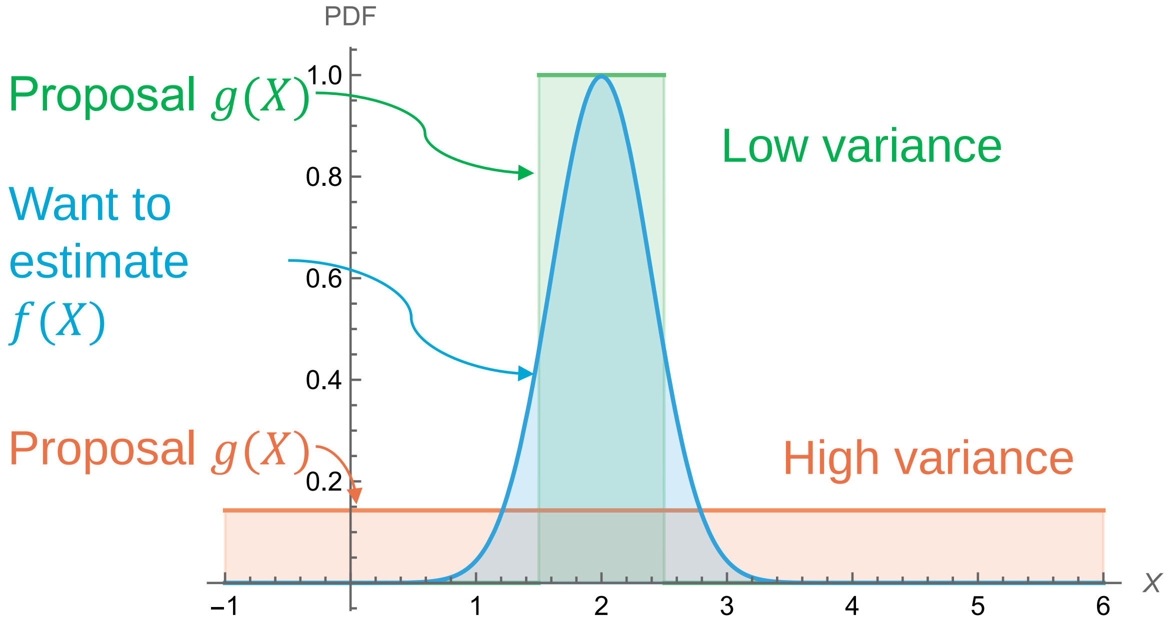
Throw fair dice and approximate

$$\approx \frac{1}{N} \sum_{i=1}^{N} X_i^f \cdot \boxed{\frac{g(X_i^f)}{f(X_i^f)}}$$

Weight ω_i

To see simulation of this procedure: https://youtu.be/V8f8ueBc9sY

Importance Sampling Variance



Choose $g(X)$ close to $f(X)$ 🤔

$$\mu = \mathrm{E}_g[X] = ?$$

Will have zero variance

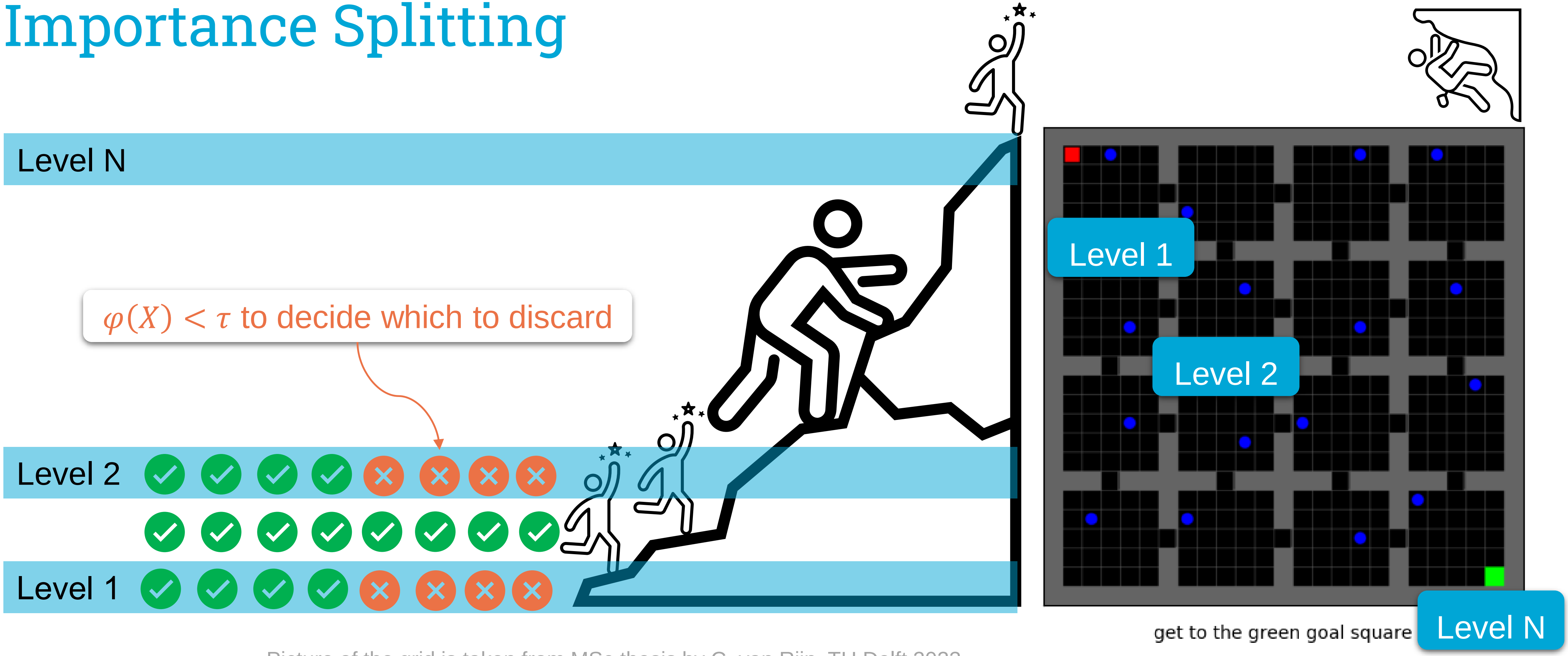
$$f(X) = \frac{g(X)X}{\mu}$$

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} X_i^f \cdot \frac{g(X_i^f)\mu}{g(X_i^f)X_i^f} = \frac{1}{N} N\mu = \mu$$

$$\mathrm{Var}(\hat{\mu}) = 0$$ 👍

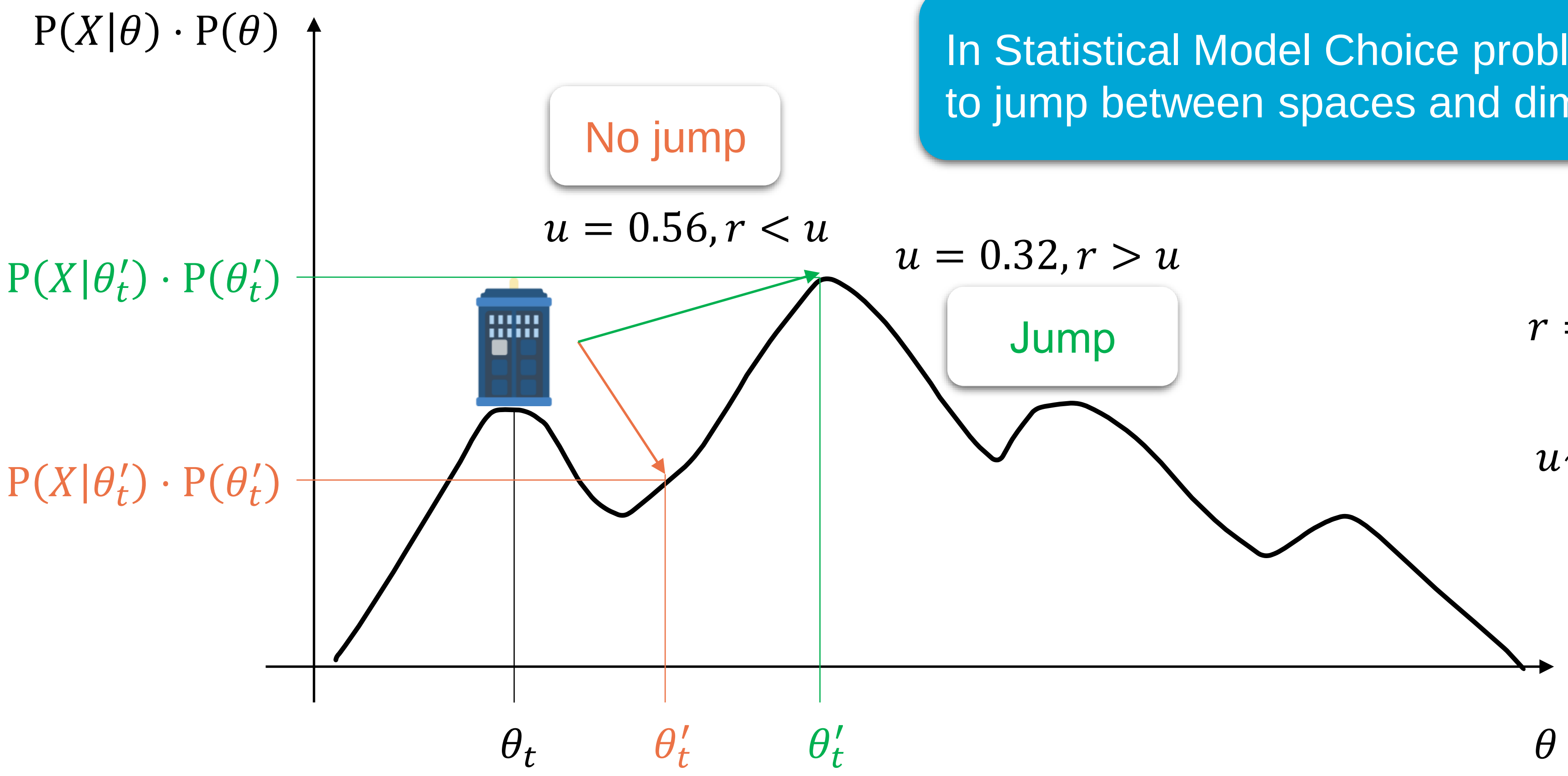
Kahn, H. and Harris, T. E. (1951). Estimation of particle transmission by random sampling. https://dornsifecms.usc.edu/assets/sites/520/docs/kahnharris.pdf

Importance Splitting



Picture of the grid is taken from MSc thesis by C. van Rijn, TU Delft 2023
Ma, H., Blom, H.A.P. Random Assignment Versus Fixed Assignment in Multilevel Importance Splitting for Estimating Stochastic Reach Probabilities. Methodol Comput Appl Probab 24, 2313–2338 (2022). https://doi.org/10.1007/s11009-021-09892-4

Random Walk Metropolis



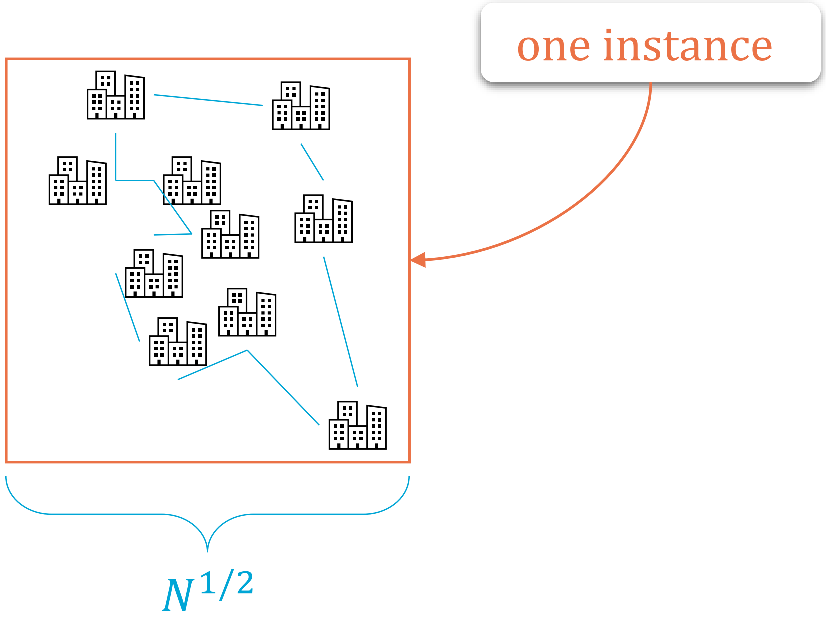
In Statistical Model Choice problems, it is often necessary to jump between spaces and dimensions

$$r = \frac{P(X|\theta_t') \cdot P(\theta_t')}{P(X|\theta_{t-1}) \cdot P(\theta_{t-1})}$$

$$u \sim U(0,1)$$

http://www.statslab.cam.ac.uk/~rrw1/markov/M.pdf

Travelling Salesperson



A solution is a list of the positions of N cities

Simulation:

- Choosing $N^{1/2}$ as square length keeps average tour length α independent of N

Greedy to bound tour length from above:

- Go to the nearest
- $\bar{\alpha} = 1.12$
- Variance decreases as $N^{-1/2}$

Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by Simulated Annealing. *Science*, *220*(4598), 671–680. http://www.jstor.org/stable/1690046

Simulated Annealing

Real cities are not uniformly distributed, but are clumped, with dense and sparse regions

Metropolis on fire:

- Randomly swap directions between two neighbor cities

Fighting local minima:

1. Metropolis exploration: bad swaps are allowed if $e^{-\frac{\Delta\alpha}{T}} > u,\ u \sim U(0,1)$ (i.e., $-\frac{\Delta\alpha}{T} \nearrow$ with $\Delta\alpha$ and $\searrow$ with T).
2. Temperature T: lower when converging.

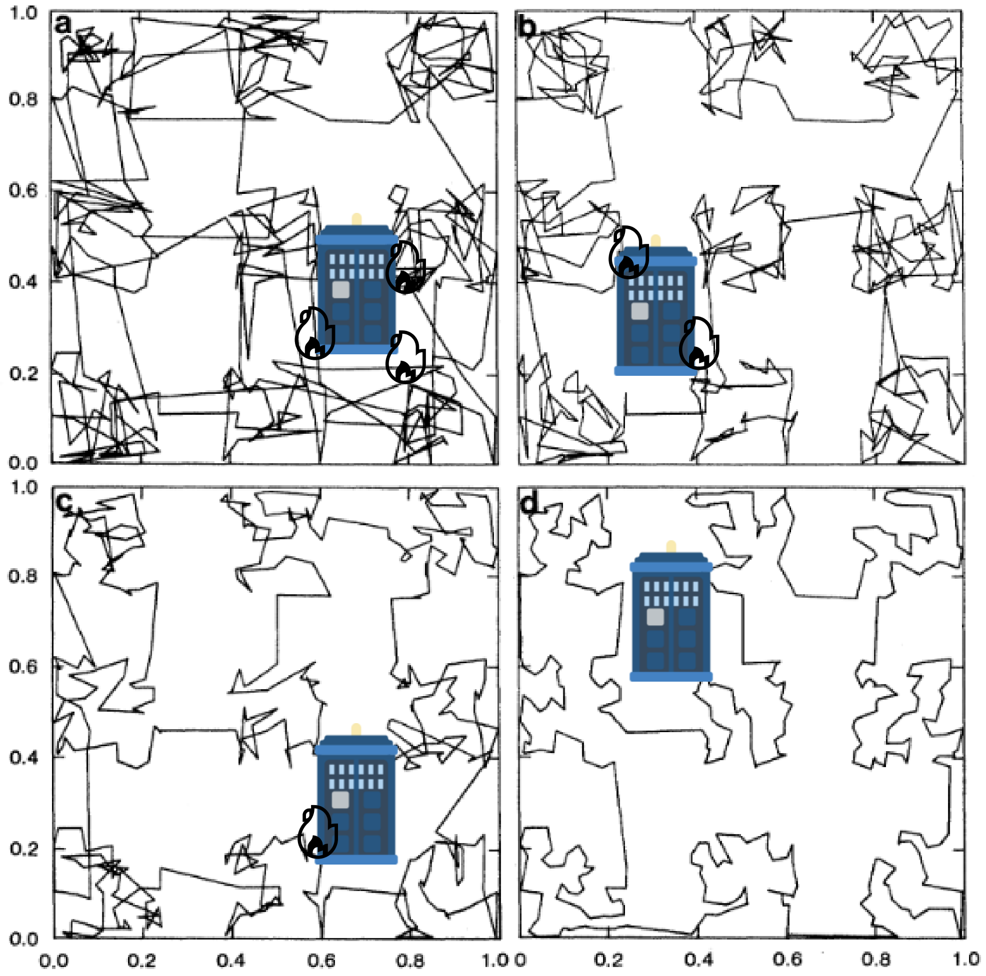


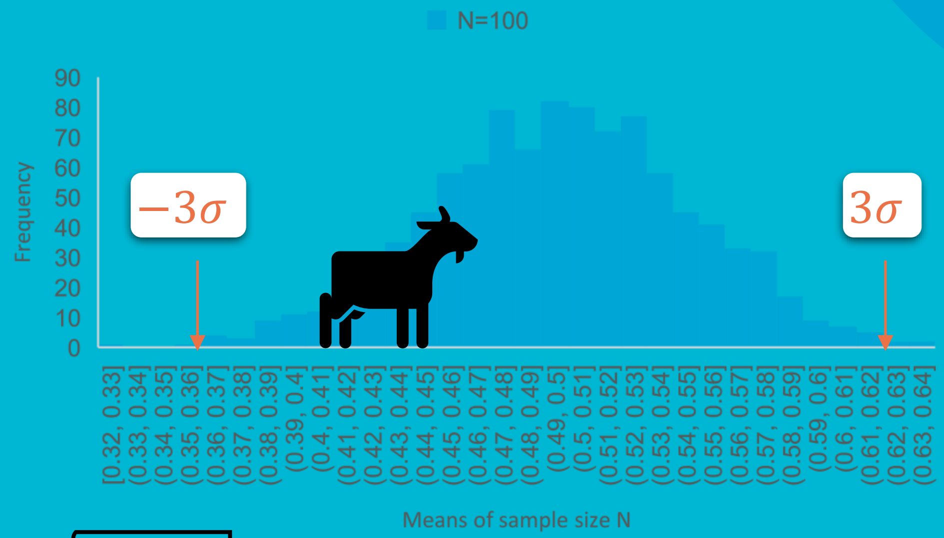
Fig. 9. Results at four temperatures for a clustered 400-city traveling salesman problem. The points are uniformly distributed in nine regions. (a) $T = 1.2$, $\alpha = 2.0567$; (b) $T = 0.8$, $\alpha = 1.515$; (c) $T = 0.4$, $\alpha = 1.055$; (d) $T = 0.0$, $\alpha = 0.7839$.

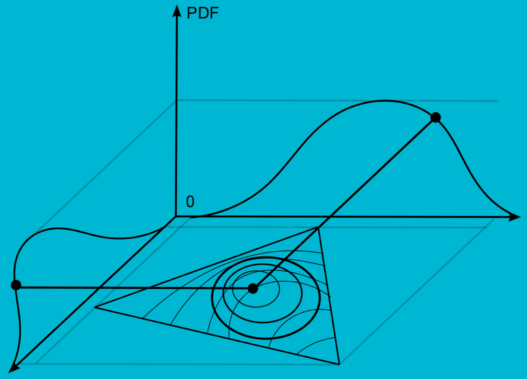
Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by Simulated Annealing. *Science*, *220*(4598), 671–680. http://www.jstor.org/stable/1690046

$$z = \frac{S_N/N - \mu}{\sigma/\sqrt{N}} \sim \mathcal{N}(0,1)$$

Monte-Carlo integration of peculiar shapes

- Don't forget confidence interval!
- $N \geq \frac{\log\left(\frac{2}{\delta}\right)}{2\varepsilon^2}$





Can't sample from the distribution:

- Sample from another one close to it and normalize
- Or split into levels and sample anew from each

NP-Hard problems to solve with simulations:

- Train shunting with uncertain delays
- Reachability probability of rare events in planning
- Estimation in multi-dimensional space

Can't sample iid:

- Sample conditionally from ergodic, aperiodic, symmetric Markov chains with stationary distribution as the target

Can't get out of local minima:

- Increase the temperature

Thanks for being part of this year’s simulated annealing of Part III, your feedback brings us closer to the globally optimal course! We will soon be able to reduce the temperature 🔥