Splitting Techniques for Improving Sequential Monte Carlo

Slava Vaisman
(joint work with Dirk P. Kroese and Ilya B. Gertsbakh)

The University of Queensland and Ben-Gurion

r.vaisman@uq.edu.au, kroese@maths.uq.edu.au, eliahu@indigo.cs.bgu.ac.il

April 17, 2015

Overview

Sequential Monte Carlo

1. Simulating Self Avoiding Walks (polymers)
2. Molecular Simulation
3. Inference in Population Genetics
4. Finding Patterns in DNA Sequences
5. Counting 0/1 Tables
6. Approximating the Permanent
7. Network Reliability

The Crude Monte Carlo (CMC)

- Consider a random variable $\mathbf{X} \sim f(\mathbf{x})$ taking values in $\mathcal{X}$.
- A general objective of Monte Carlo simulation is to calculate

$$\ell = \mathbb{E}_f \left(H \left(\mathbf{X} \right) \right),$$

where $H : \mathcal{X} \to \mathbb{R}$ is a real-valued function.

The Crude Monte Carlo (CMC)

- Consider a random variable $\mathbf{X} \sim f(\mathbf{x})$ taking values in $\mathcal{X}$.
- A general objective of Monte Carlo simulation is to calculate

$$\ell = \mathbb{E}_f \left(H \left(\mathbf{X} \right) \right),$$

where $H : \mathcal{X} \to \mathbb{R}$ is a real-valued function.

- The Crude Monte Carlo (CMC) estimator of ℓ is given by

$$\widehat{\ell} = \frac{1}{N} \sum_{i=1}^{N} H \left(\mathbf{X}_i \right),$$

where $\mathbf{X}_i$ for $i = 1, \ldots, N$, are independent copies of random variable generated from $f(\mathbf{x})$.

The Crude Sequential Monte Carlo (CSMC)

- Suppose that the vector $\mathbf{X} \in \mathcal{X}$ is decomposable

$$\mathbf{X} = (X_1, X_2, \ldots, X_T), \quad 1 \leqslant T \leqslant n,$$

where for each $t = 1, \ldots, T$, X_t can be multidimensional.

The Crude Sequential Monte Carlo (CSMC)

- Suppose that the vector $\mathbf{X} \in \mathcal{X}$ is decomposable

$$\mathbf{X} = (X_1, X_2, \ldots, X_T), \quad 1 \leqslant T \leqslant n,$$

  where for each $t = 1, \ldots, T$, X_t can be multidimensional.
- Suppose that $f(\mathbf{x}_{1:t})$ is decomposable:

$$f(\mathbf{x}_{1:t}) = f_1(x_1) f_2(x_2 \mid x_1) \cdots f_t(x_t \mid x_1, \ldots, x_{t-1}), \quad 1 \leqslant t \leqslant n.$$

The Crude Sequential Monte Carlo (CSMC)

- Suppose that the vector $\mathbf{X} \in \mathcal{X}$ is decomposable

$$\mathbf{X} = (X_1, X_2, \ldots, X_T), \quad 1 \leqslant T \leqslant n,$$

  where for each $t = 1, \ldots, T$, X_t can be multidimensional.
- Suppose that $f(\mathbf{x}_{1:t})$ is decomposable:

$$f(\mathbf{x}_{1:t}) = f_1(x_1) f_2(x_2 \mid x_1) \cdots f_t(x_t \mid x_1, \ldots, x_{t-1}), \quad 1 \leqslant t \leqslant n.$$

- **Sequential Generation of X**:

$$\mathbf{X} = ().$$

The Crude Sequential Monte Carlo (CSMC)

- Suppose that the vector $\mathbf{X} \in \mathcal{X}$ is decomposable

$$\mathbf{X} = (X_1, X_2, \ldots, X_T), \quad 1 \leqslant T \leqslant n,$$

  where for each $t = 1, \ldots, T$, X_t can be multidimensional.
- Suppose that $f(\mathbf{x}_{1:t})$ is decomposable:

$$f(\mathbf{x}_{1:t}) = f_1(x_1) f_2(x_2 \mid x_1) \cdots f_t(x_t \mid x_1, \ldots, x_{t-1}), \quad 1 \leqslant t \leqslant n.$$

- **Sequential Generation of X**:

$$\mathbf{X} = (x_1).$$

The Crude Sequential Monte Carlo (CSMC)

- Suppose that the vector $\mathbf{X} \in \mathcal{X}$ is decomposable

$$\mathbf{X} = (X_1, X_2, \ldots, X_T), \quad 1 \leqslant T \leqslant n,$$

  where for each $t = 1, \ldots, T$, X_t can be multidimensional.
- Suppose that $f(\mathbf{x}_{1:t})$ is decomposable:

$$f(\mathbf{x}_{1:t}) = f_1(x_1) f_2(x_2 \mid x_1) \cdots f_t(x_t \mid x_1, \ldots, x_{t-1}), \quad 1 \leqslant t \leqslant n.$$

- **Sequential Generation of X**:

$$\mathbf{X} = (x_1, x_2).$$

The Crude Sequential Monte Carlo (CSMC)

- Suppose that the vector $\mathbf{X} \in \mathcal{X}$ is decomposable

$$\mathbf{X} = (X_1, X_2, \ldots, X_T), \quad 1 \leqslant T \leqslant n,$$

  where for each $t = 1, \ldots, T$, X_t can be multidimensional.
- Suppose that $f(\mathbf{x}_{1:t})$ is decomposable:

$$f(\mathbf{x}_{1:t}) = f_1(x_1) f_2(x_2 \mid x_1) \cdots f_t(x_t \mid x_1, \ldots, x_{t-1}), \quad 1 \leqslant t \leqslant n.$$

- **Sequential Generation of X**:

$$\mathbf{X} = (x_1, x_2, x_3).$$

The Crude Sequential Monte Carlo (CSMC)

- Suppose that the vector $\mathbf{X} \in \mathcal{X}$ is decomposable

$$\mathbf{X} = (X_1, X_2, \ldots, X_T), \quad 1 \leqslant T \leqslant n,$$

  where for each $t = 1, \ldots, T$, X_t can be multidimensional.
- Suppose that $f(\mathbf{x}_{1:t})$ is decomposable:

$$f(\mathbf{x}_{1:t}) = f_1(x_1) f_2(x_2 \mid x_1) \cdots f_t(x_t \mid x_1, \ldots, x_{t-1}), \quad 1 \leqslant t \leqslant n.$$

- **Sequential Generation of X**:

$$\mathbf{X} = (x_1, x_2, x_3, x_4, \cdots).$$

The Crude Sequential Monte Carlo (CSMC)

- Suppose that the vector $\mathbf{X} \in \mathcal{X}$ is decomposable

$$\mathbf{X} = (X_1, X_2, \ldots, X_T), \quad 1 \leqslant T \leqslant n,$$

  where for each $t = 1, \ldots, T$, X_t can be multidimensional.
- Suppose that $f(\mathbf{x}_{1:t})$ is decomposable:

$$f(\mathbf{x}_{1:t}) = f_1(x_1) f_2(x_2 \mid x_1) \cdots f_t(x_t \mid x_1, \ldots, x_{t-1}), \quad 1 \leqslant t \leqslant n.$$

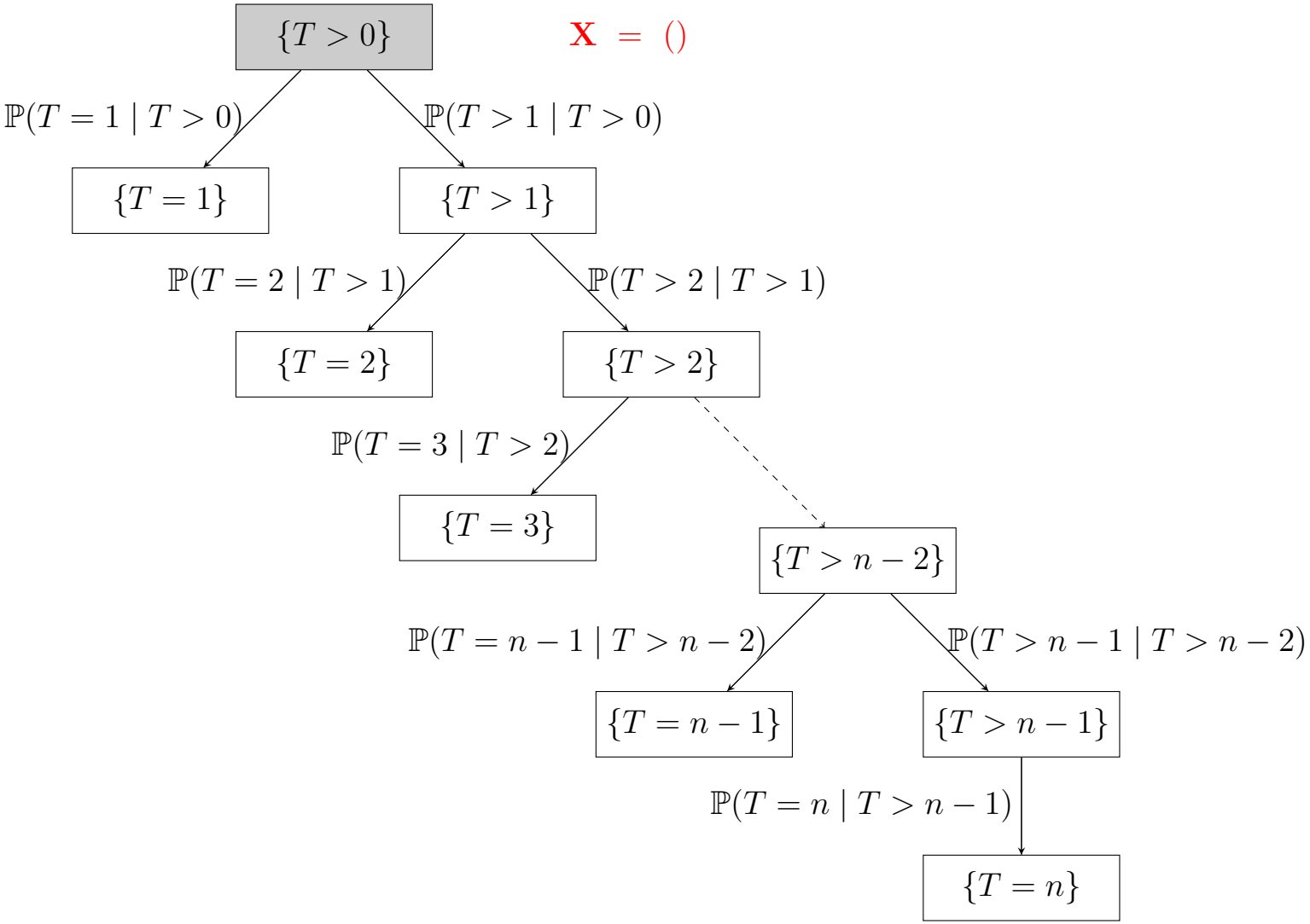
- **Sequential Generation of X**:

$$\mathbf{X} = (x_1, x_2, x_3, x_4, \cdots).$$

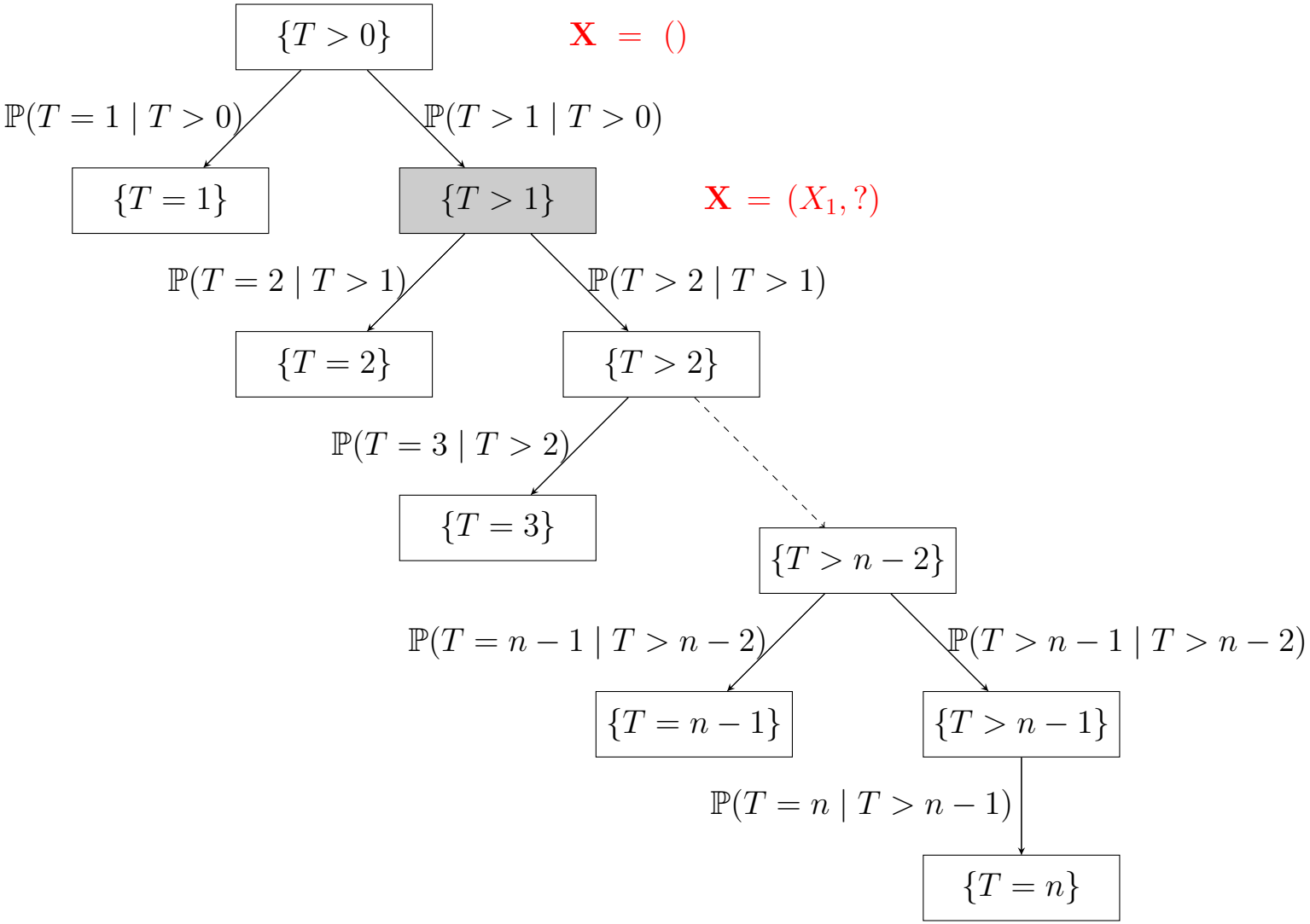
- **Example - samples of different lengths**. A fair coin that is tossed repeatedly until the first "Head" appears or, until n tosses have been made. Then,

$$\mathcal{X} = \{H, TH, TTH, \ldots, \underbrace{T \cdots, T}_{n \text{ times}}\}.$$

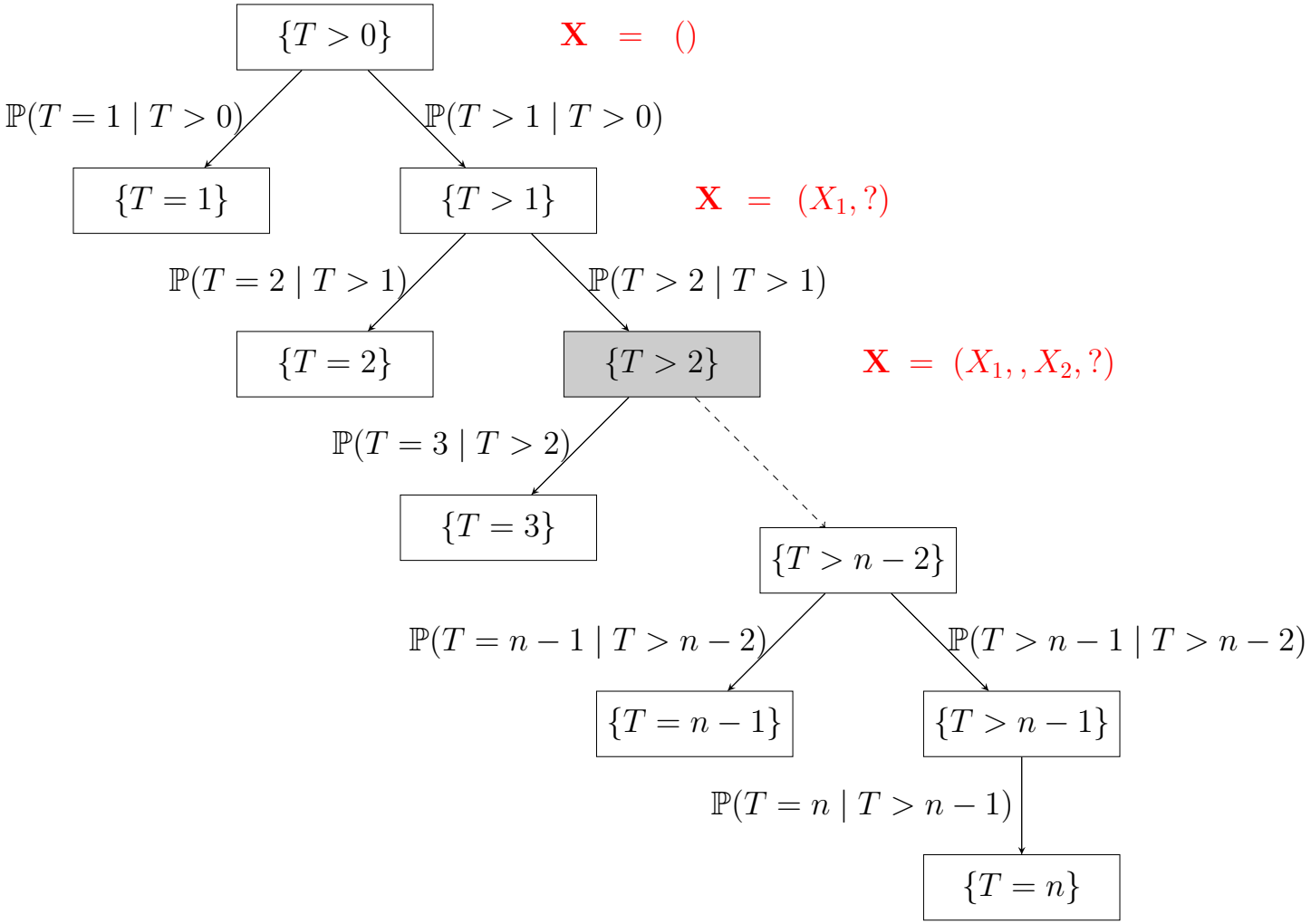
CSMC – **X** Generation



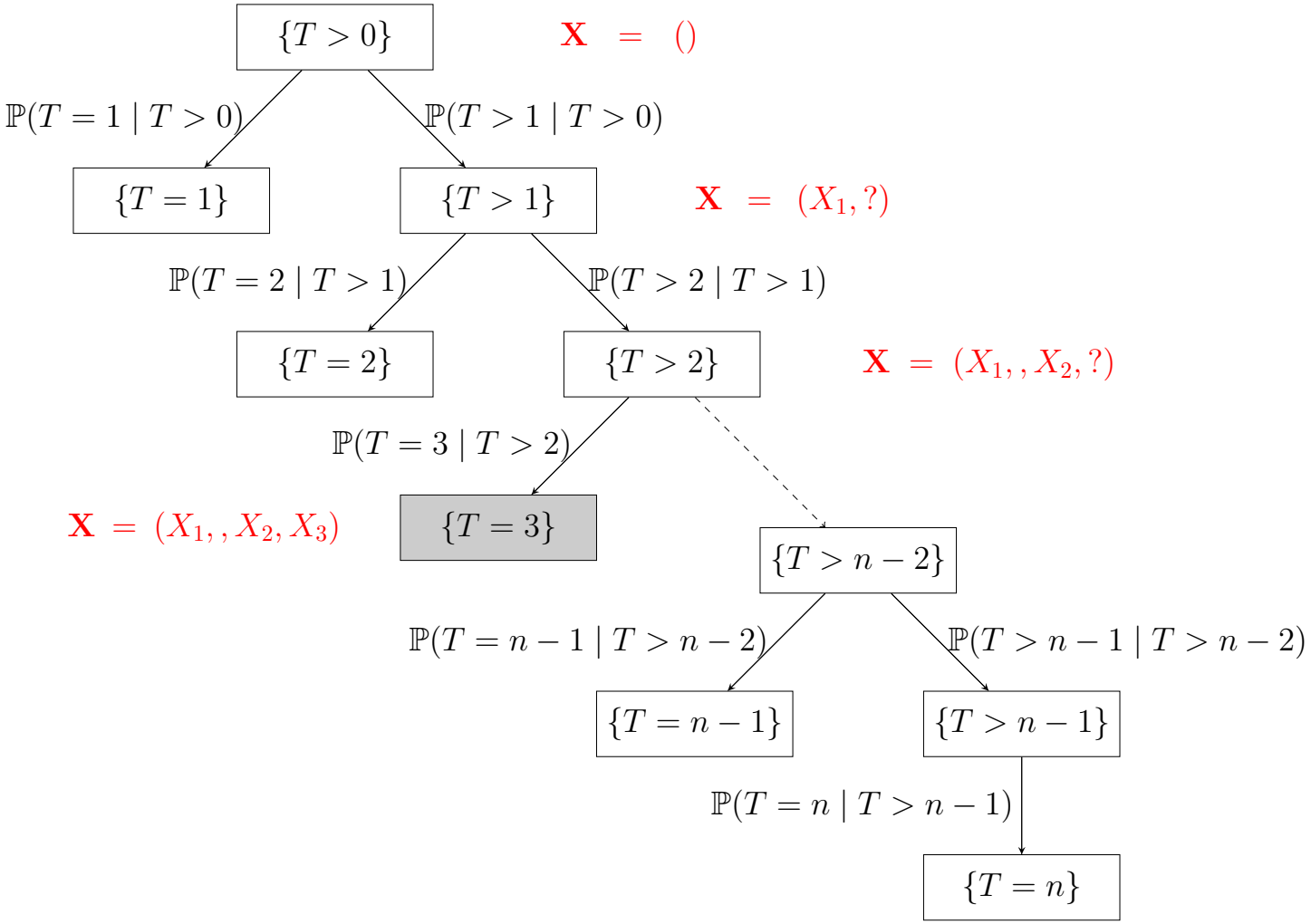
CSMC – **X** Generation



CSMC – **X** Generation



CSMC – **X** Generation



CSMC and Rare Events

Consider the SMC process tree and suppose for simplicity that

$$\mathbb{P}(T = t \mid T > t-1) = 1 - \mathbb{P}(T > t \mid T > t-1) = 1/2 \quad \text{for } 1 \leqslant t \leqslant n-1,$$

and

$$\mathbb{P}(T = n \mid T > n-1) = 1.$$

We start a single walk from the root ($\{T > 0\}$) which ends at some leaf $\{T = t\}$. Then

$$\mathbb{P}(\text{A single random walk reaches the } \{T = n\} \text{ leaf}) = \mathbf{2^{-(n-1)}}.$$

CSMC and Rare Events

Consider the SMC process tree and suppose for simplicity that

$$\mathbb{P}(T = t \mid T > t-1) = 1 - \mathbb{P}(T > t \mid T > t-1) = 1/2 \quad \text{for } 1 \leqslant t \leqslant n-1,$$

and

$$\mathbb{P}(T = n \mid T > n-1) = 1.$$

We start a single walk from the root ($\{T > 0\}$) which ends at some leaf $\{T = t\}$. Then

$$\mathbb{P}(\text{A single random walk reaches the } \{T = n\} \text{ leaf}) = \mathbf{2^{-(n-1)}}.$$

Do we have a problem?

CSMC and Rare Events

Consider the SMC process tree and suppose for simplicity that

$$\mathbb{P}(T = t \mid T > t-1) = 1 - \mathbb{P}(T > t \mid T > t-1) = 1/2 \quad \text{for } 1 \leqslant t \leqslant n-1,$$

and

$$\mathbb{P}(T = n \mid T > n-1) = 1.$$

We start a single walk from the root ($\{T > 0\}$) which ends at some leaf $\{T = t\}$. Then

$$\mathbb{P}(\text{A single random walk reaches the } \{T = n\} \text{ leaf}) = \mathbf{2^{-(n-1)}}.$$

Do we have a problem?

Rare-Event

$$H(\mathbf{x}) = \begin{cases} 2^{n-1}\left(n^{n^n}\right) & \text{if } |\mathbf{x}| = n \\ 0 & \text{otherwise} \end{cases} \Rightarrow \mathbb{E}(H(\mathbf{X})) = ?$$

CSMC and Rare Events

Consider the SMC process tree and suppose for simplicity that

$$\mathbb{P}(T = t \mid T > t-1) = 1 - \mathbb{P}(T > t \mid T > t-1) = 1/2 \quad \text{for } 1 \leqslant t \leqslant n-1,$$

and

$$\mathbb{P}(T = n \mid T > n-1) = 1.$$

We start a single walk from the root ($\{T > 0\}$) which ends at some leaf $\{T = t\}$. Then

$$\mathbb{P}(\text{A single random walk reaches the } \{T = n\} \text{ leaf}) = \mathbf{2^{-(n-1)}}.$$

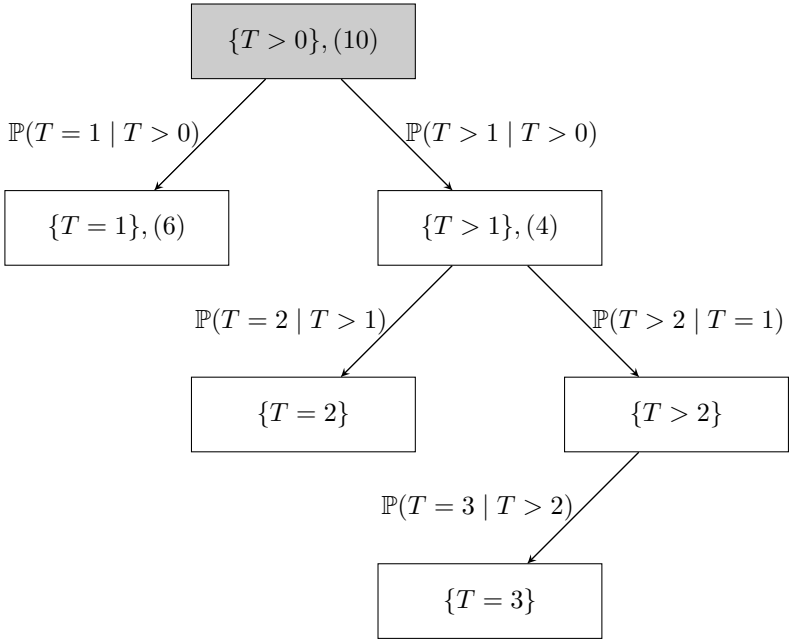
Do we have a problem?

Rare-Event

$$H(\mathbf{x}) = \begin{cases} 2^{n-1}\left(n^{n^n}\right) & \text{if } |\mathbf{x}| = n \\ 0 & \text{otherwise} \end{cases} \Rightarrow \mathbb{E}(H(\mathbf{X})) = n^{n^n}.$$

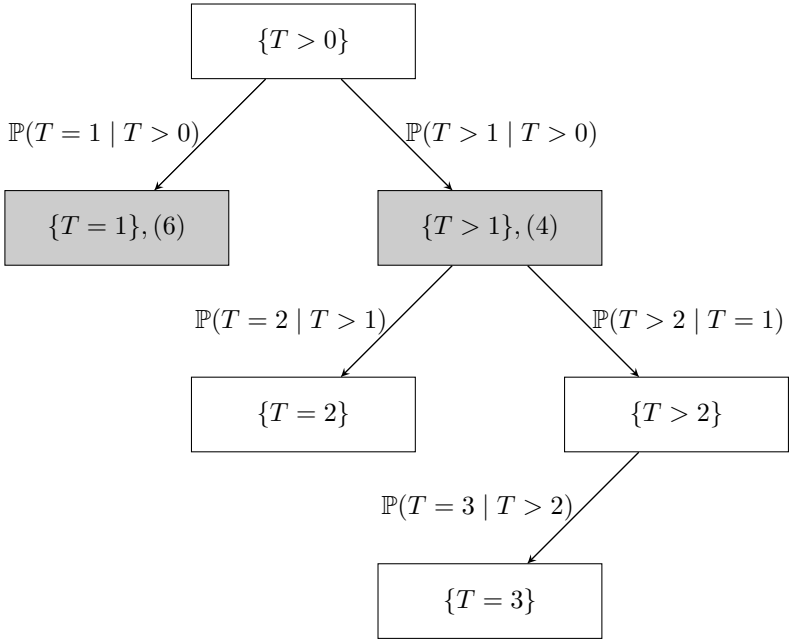
The Splitting Idea

- Define some budget $B \in \mathbb{N} \setminus \{0\}$ and launch B parallel walks (trajectories), from the tree root.
- At each step $t = 1, \ldots, n-1$, we detect the trajectories that "finished" their execution, that is, they are at state $\{T = t\}$. Clearly, there is no point to run them further, so, we better reallocate them to the $\{T > t\}$ state and continue the execution (again, with B trajectories).



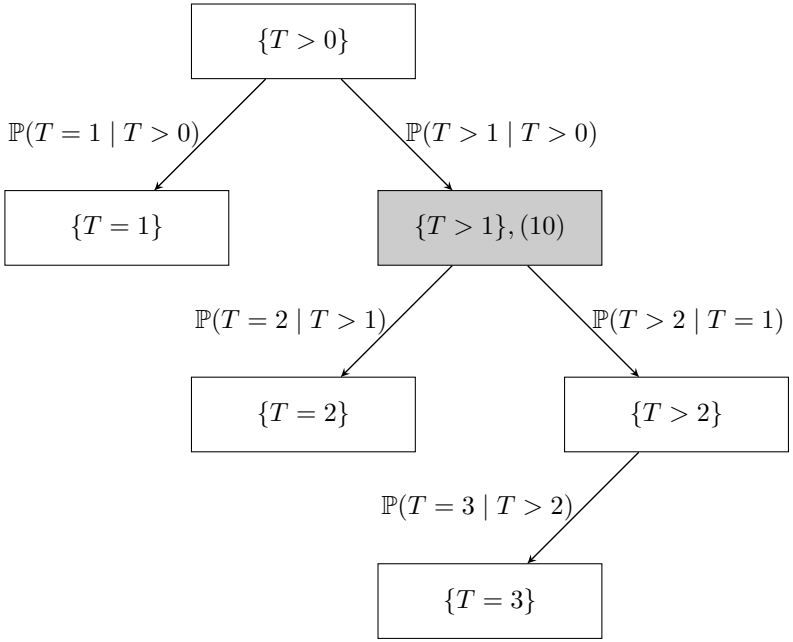
The Splitting Idea

- Define some budget $B \in \mathbb{N} \setminus \{0\}$ and launch B parallel walks (trajectories), from the tree root.
- At each step $t = 1, \ldots, n-1$, we detect the trajectories that "finished" their execution, that is, they are at state $\{T = t\}$. Clearly, there is no point to run them further, so, we better reallocate them to the $\{T > t\}$ state and continue the execution (again, with B trajectories).

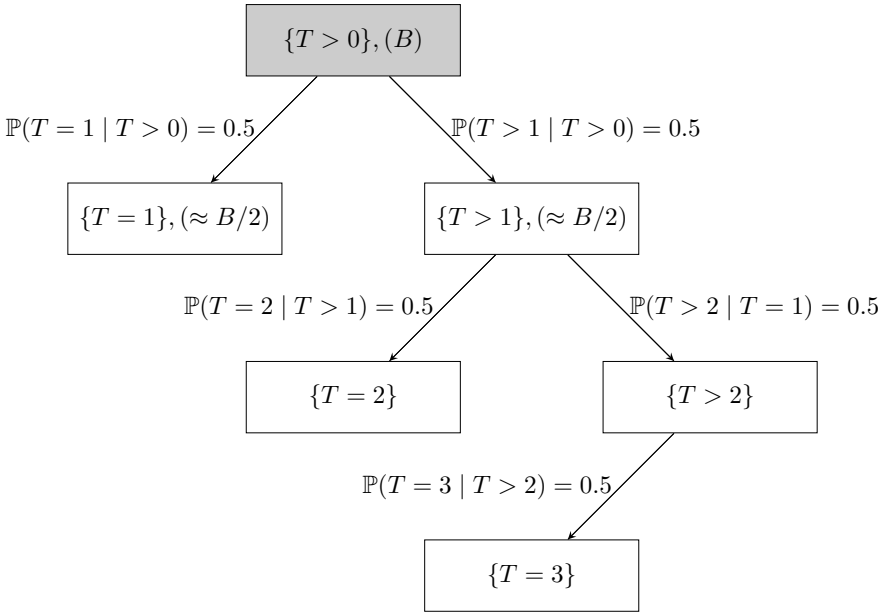


The Splitting Idea

- Define some budget $B \in \mathbb{N} \setminus \{0\}$ and launch B parallel walks (trajectories), from the tree root.
- At each step $t = 1, \ldots, n-1$, we detect the trajectories that "finished" their execution, that is, they are at state $\{T = t\}$. Clearly, there is no point to run them further, so, we better reallocate them to the $\{T > t\}$ state and continue the execution (again, with B trajectories).

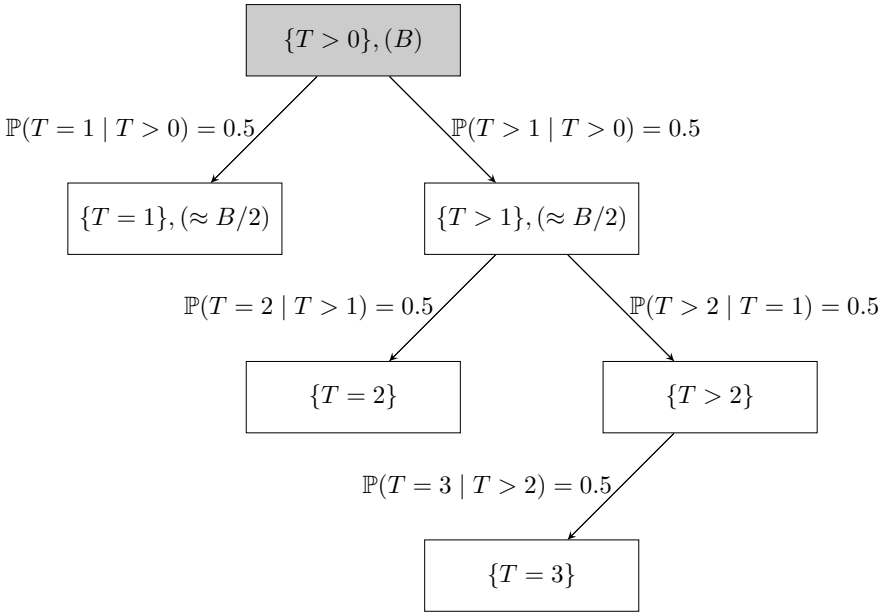


The Splitting Gain



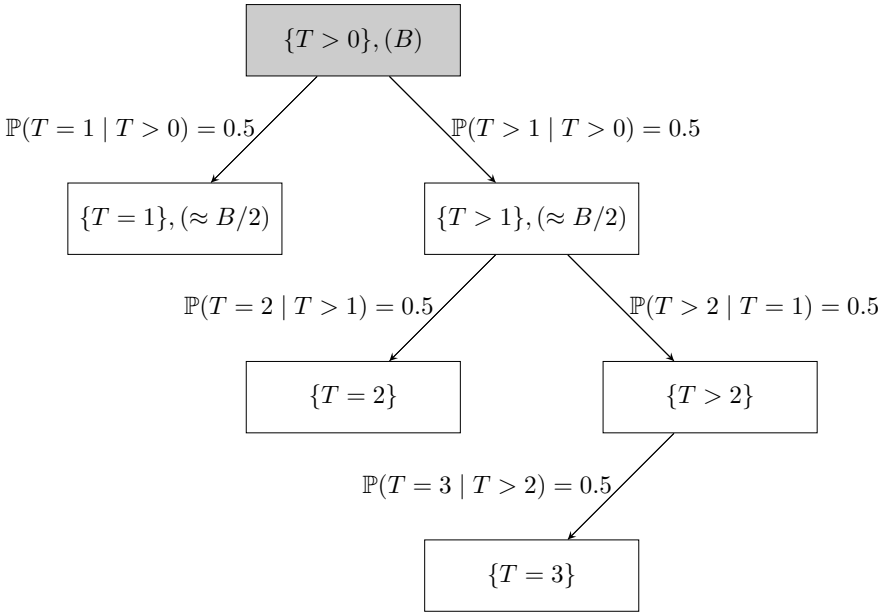
1. $\mathbb{P}(T > t \mid T > t - 1) = 1 - 1/2^B$.

The Splitting Gain



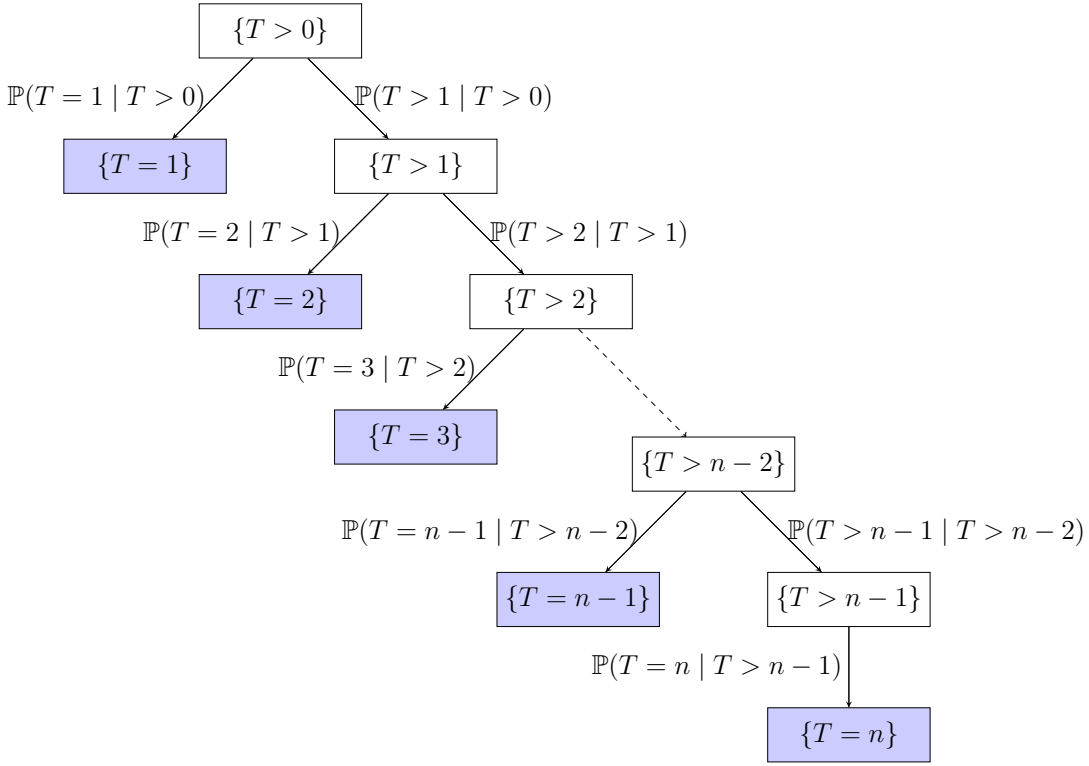
1. $\mathbb{P}(T > t \mid T > t - 1) = 1 - 1/2^B$.
2. $\mathbb{P}(T = n) = \left(1 - 1/2^B\right)^{n-1}$.

The Splitting Gain



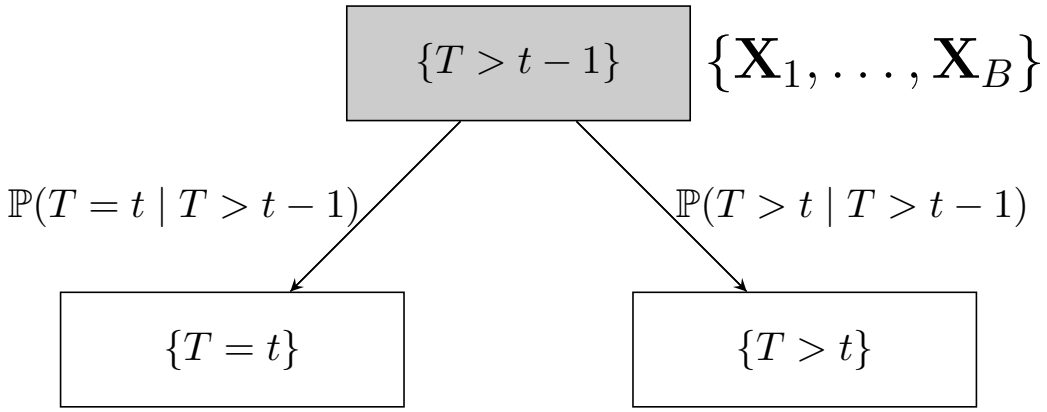
1. $\mathbb{P}(T > t \mid T > t-1) = 1 - 1/2^B$.
2. $\mathbb{P}(T = n) = \left(1 - 1/2^B\right)^{n-1}$.
3. For $B = \lceil \log_2(n-1) \rceil$, it holds that
$$\mathbb{P}(T = n) \geqslant (1 - 1/2^{\log_2(n-1)})^{n-1} \to \mathrm{e}^{-1}, \quad n \to \infty.$$

The SSMC Algorithm (1)

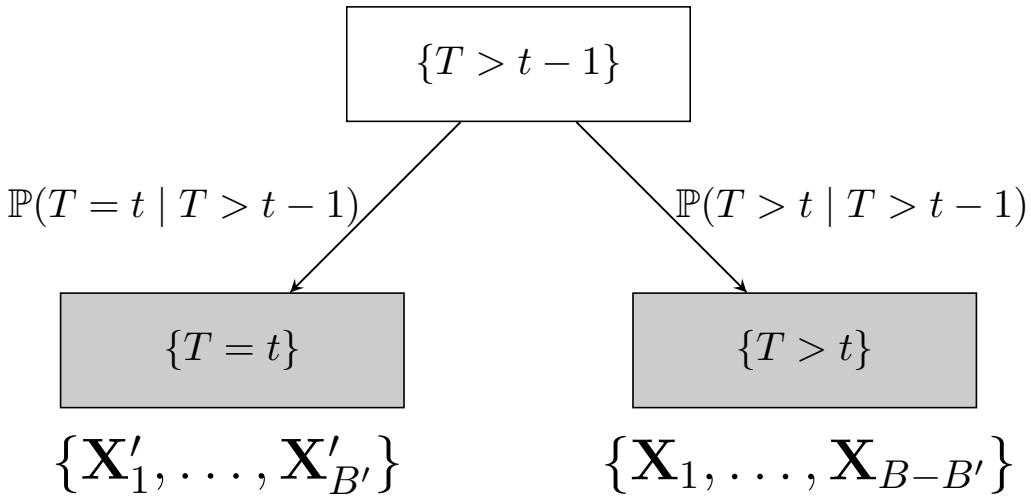


1. Develop efficient estimators (P_t) for $\mathbb{P}(T = t),\ t = 1, \ldots, n$.
2. Develop efficient estimators (H_t) for $\mathbb{E}(H(\mathbf{X}) \mid |\mathbf{X}| = t),\ t = 1, \ldots, n$.
3. Deliver $C = \sum_{t=1}^{n} H_t \cdot P_t$, as an estimator of $\mathbb{E}(H(\mathbf{X}))$ (by the law of total expectation).

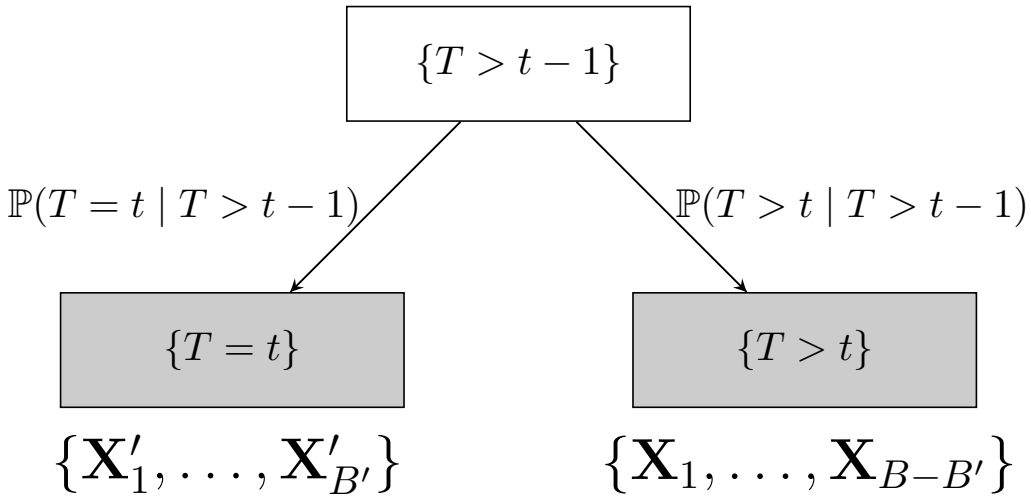
The SSMC Algorithm (2)



The SSMC Algorithm (2)



The SSMC Algorithm (2)



Estimators

- $P_t = (1 - P_{t-1}) \cdot \frac{B'}{B}, \quad t = 1, \ldots, n.$
- $H_t = \frac{1}{B} \cdot \sum_{i=1}^{B'} H(\mathbf{X}'_i), \quad t = 1, \ldots, n.$
- $C = \sum_{t=1}^{n} H_t \cdot P_t.$

The SSMC Analysis

Suppose that SSMC Algorithm outputs a random variable C. Then, the following holds.

Theorem (Slava, Dirk and Ilya: Unbiased estimator)

The SSMC Algorithm outputs an unbiased estimator, that is, it holds that

$$\mathbb{E}(C) = \mathbb{E}_f\left(H(\mathbf{X})\right).$$

The SSMC Analysis

Suppose that SSMC Algorithm outputs a random variable C. Then, the following holds.

Theorem (Slava, Dirk and Ilya: Unbiased estimator)

The SSMC Algorithm outputs an unbiased estimator, that is, it holds that

$$\mathbb{E}(C) = \mathbb{E}_f\left(H(\mathbf{X})\right).$$

Theorem (Slava, Dirk and Ilya: Efficiency of SSMC Algorithm)

Suppose that the following holds for all $i = 1, \ldots, n$.

1. $f_i(x_i \mid x_1, \ldots, x_{i-1}) = p_i = \mathcal{O}(1/\mathcal{P})$ *where $\mathcal{P}$ is a polynomial in n.*
2. $H(\mathbf{x}\ ;\ \mathbf{x} \in \mathcal{X}_i) = H_i$ *where H_i is a constant.*

Then, under above assumptions, the SSMC Algorithm is efficient, that is, it holds that

$$\mathrm{CV} = \sqrt{\mathrm{Var}\,(C)}/\mathbb{E}\,(C)$$

is upper-bounded by a polynomial in n.

Application: Terminal Network Reliability Estimation

- The task of calculating the terminal network reliability belongs to the #P complexity class.
- One of the well-known algorithms is the SMC Lomonosov's Turnip (LT).

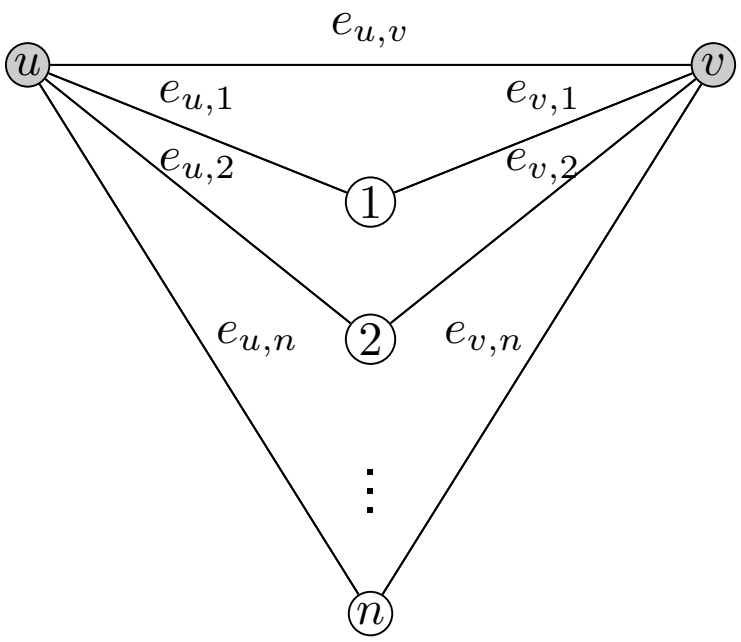


Figure: Simple graph $\mathfrak{S}(n)$.

Simple graph's CV

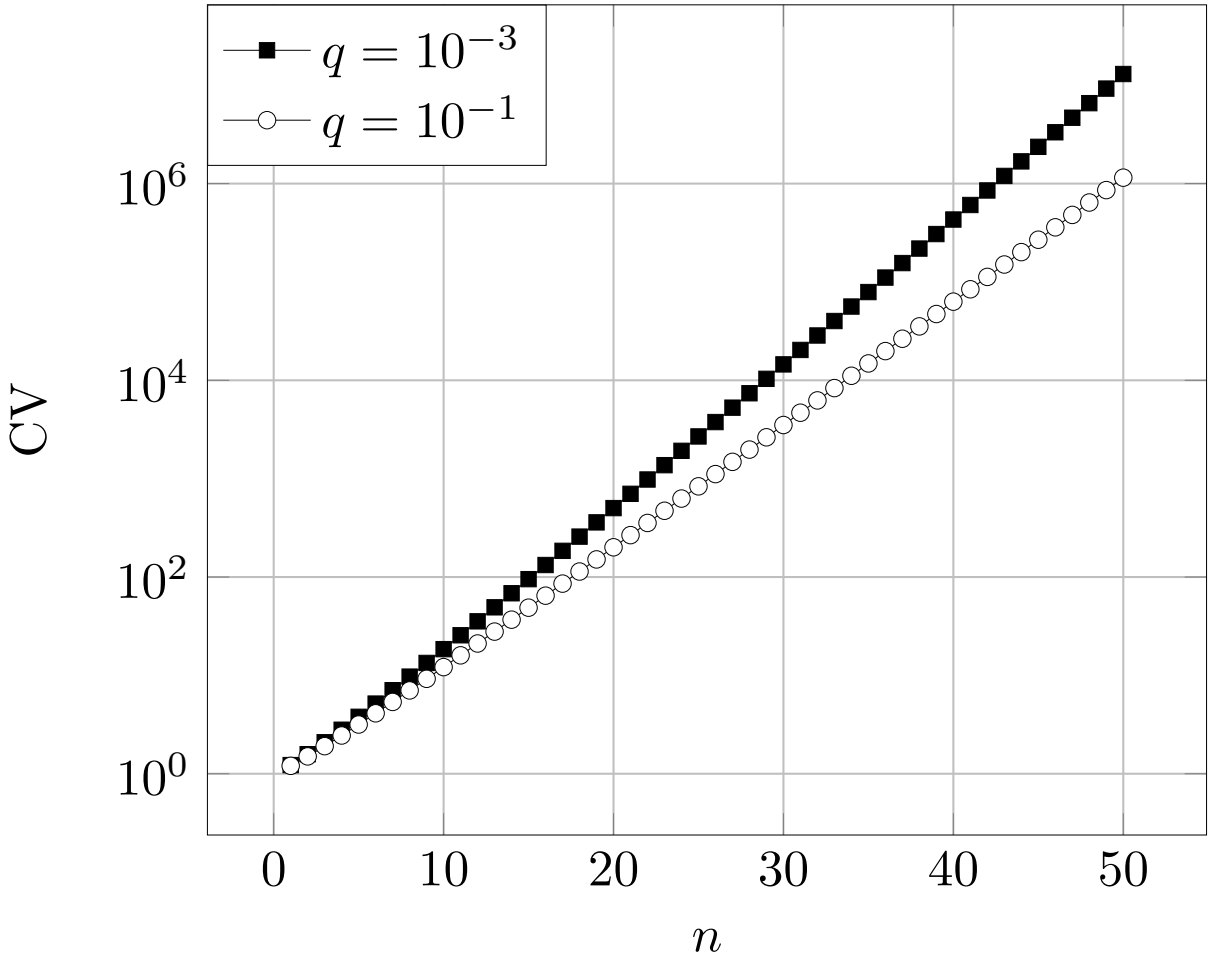


Figure: Logarithmically scaled CV of LT as a function of n for $\mathfrak{S}(n)$ networks.

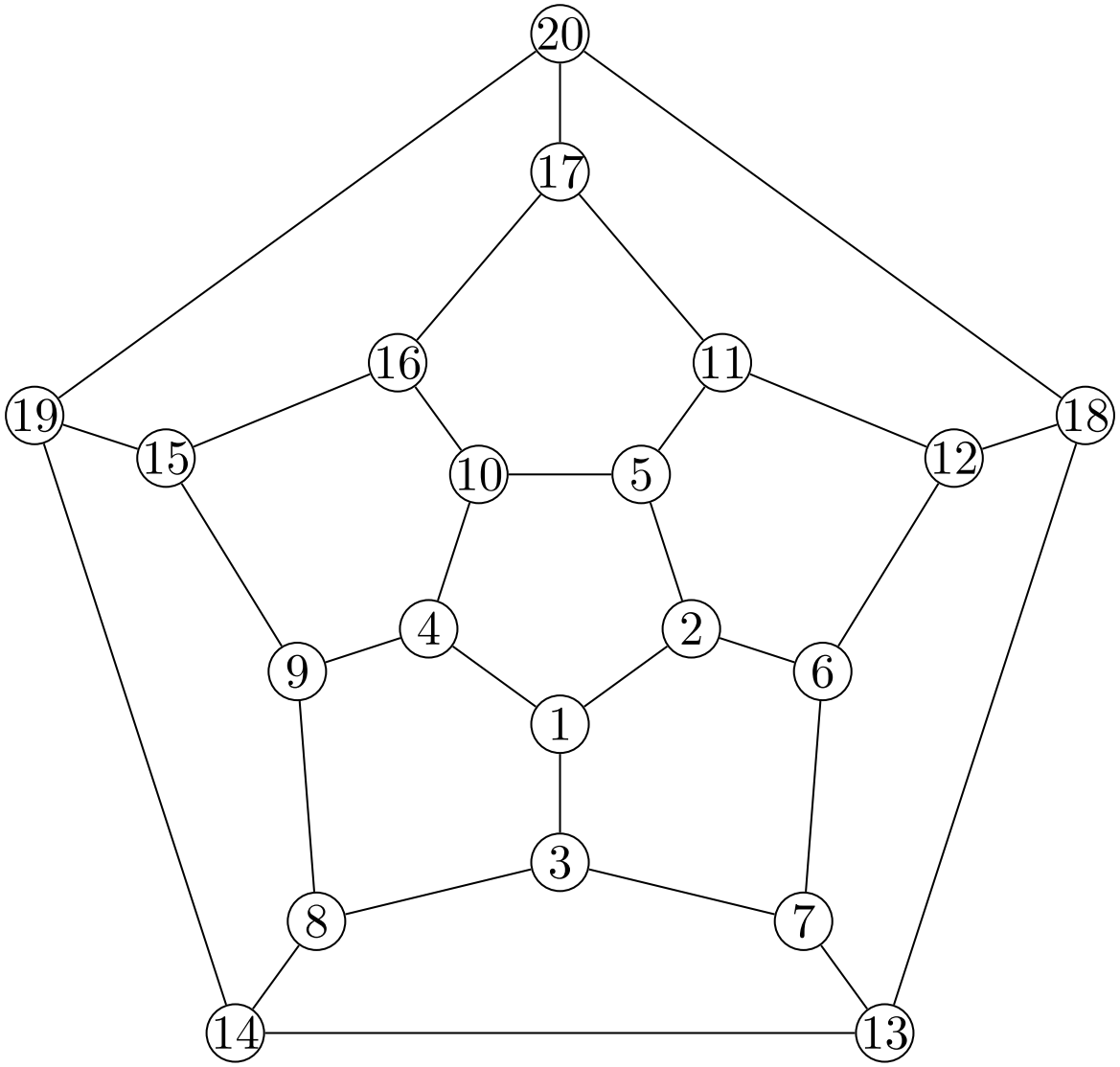
Model 1 — Simple graph $n = 50$

- $\widehat{\mathcal{K}}$ is the estimator of network unreliability.
- $\widehat{\mathrm{RE}}$ is the confidence interval.
- The relative experimental error (REE) is the error % of $\widehat{\mathcal{K}}$ form the exact network unreliability.

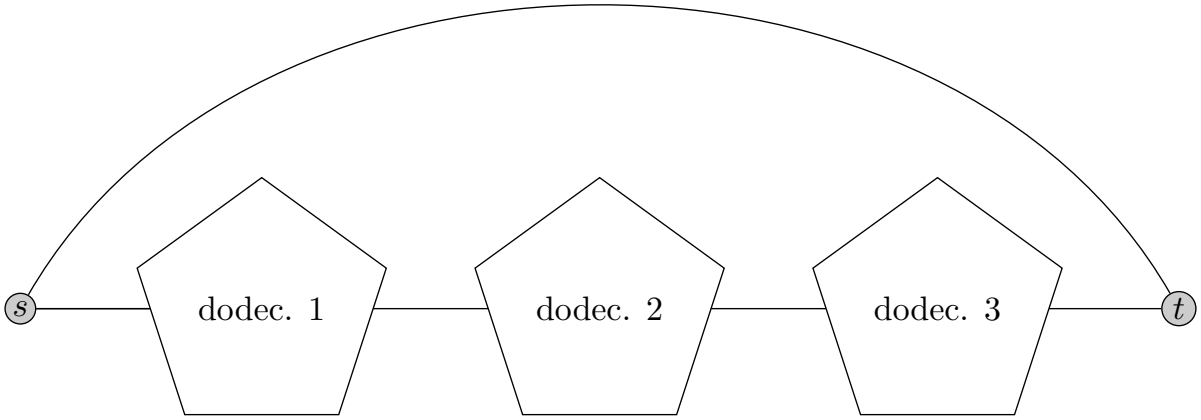
Algorithm	$\widehat{\mathcal{K}}$	RE	REE
LT	1.93×10^{-41}	76.5%	99.7%
SSMC (ST)	8.67×10^{-38}	2.48%	2.41%

Table: Performance of LT and ST for $\mathfrak{S}(50)$ network.

The dodecahedron graph

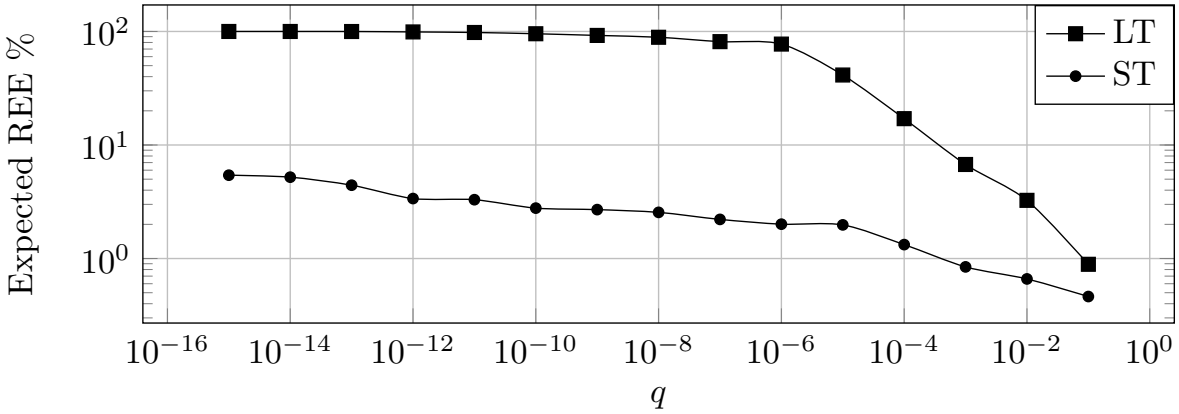


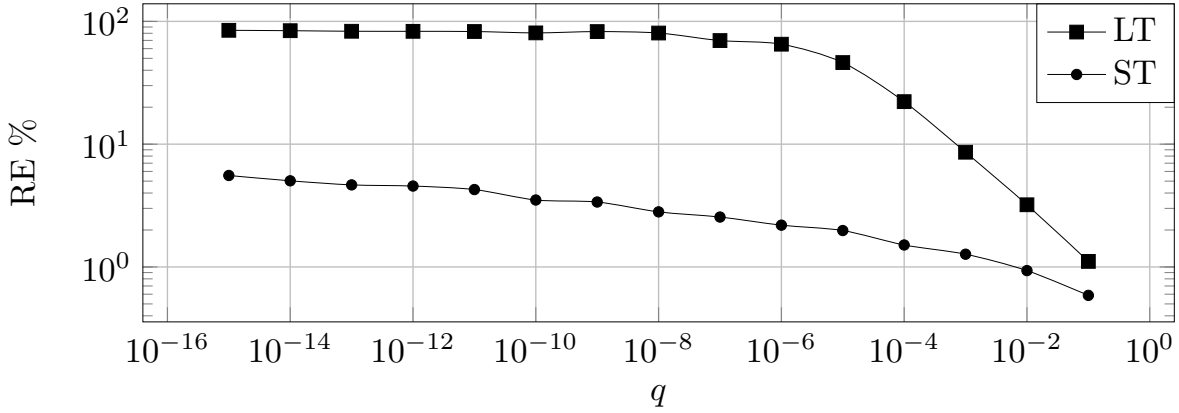
A graph with 90 edges – merging 3 dodecahedrons



Each edge fails with probability 0.4. The s–t edge fails with probability q.

A graph with 90 edges – merging 3 dodecahedrons





The End