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

1. Sequential Monte Carlo

2. Splitting Sequential Monte Carlo

3. Analysis

4. Efficient Unreliability Estimation of Highly Reliable Networks
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
Consider a random variable $X \sim f(x)$ taking values in $\mathcal{X}$.

A general objective of Monte Carlo simulation is to calculate

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

where $H : \mathcal{X} \to \mathbb{R}$ is a real-valued function.
The Crude Monte Carlo (CMC)

- Consider a random variable $X \sim f(x)$ taking values in $\mathcal{X}$.
- A general objective of Monte Carlo simulation is to calculate
  
  $$\ell = \mathbb{E}_f (H(X)),$$
  
  where $H : \mathcal{X} \rightarrow \mathbb{R}$ is a real-valued function.
- The Crude Monte Carlo (CMC) estimator of $\ell$ is given by
  
  $$\hat{\ell} = \frac{1}{N} \sum_{i=1}^{N} H(X_i),$$
  
  where $X_i$ for $i = 1, \ldots, N$, are independent copies of random variable generated from $f(x)$. 
Suppose that the vector $\mathbf{X} \in \mathcal{X}$ is decomposable

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

where for each $t = 1, \ldots, T$, $X_t$ can be multidimensional.
Suppose that the vector $\mathbf{X} \in \mathcal{X}$ is decomposable

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

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

Suppose that $f(\mathbf{x}_{1:t})$ is decomposable:

$$f(\mathbf{x}_{1:t}) = f_1(x_1)f_2(x_2 | x_1) \cdots f_t(x_t | x_1, \ldots, x_{t-1}), \quad 1 \leq t \leq n.$$
The Crude Sequential Monte Carlo (CSMC)

- Suppose that the vector $\mathbf{X} \in \mathcal{X}$ is decomposable
  \[ \mathbf{X} = (X_1, X_2, \ldots, X_T), \quad 1 \leq T \leq n, \]
  where for each $t = 1, \ldots, T$, $X_t$ can be multidimensional.

- Suppose that $f(x_{1:t})$ is decomposable:
  \[ f(x_{1:t}) = f_1(x_1)f_2(x_2 | x_1) \cdots f_t(x_t | x_1, \ldots, x_{t-1}), \quad 1 \leq t \leq n. \]

- **Sequential Generation of $\mathbf{X}$:**
  \[ \mathbf{X} = (). \]
The Crude Sequential Monte Carlo (CSMC)

- Suppose that the vector $\mathbf{X} \in \mathcal{X}$ is decomposable
  \[ \mathbf{X} = (X_1, X_2, \ldots, X_T), \quad 1 \leq T \leq n, \]
  where for each $t = 1, \ldots, T$, $X_t$ can be multidimensional.
- Suppose that $f(x_{1:t})$ is decomposable:
  \[ f(x_{1:t}) = f_1(x_1)f_2(x_2 | x_1) \cdots f_t(x_t | x_1, \ldots, x_{t-1}), \quad 1 \leq t \leq n. \]

- **Sequential Generation of $\mathbf{X}$:**
  \[ \mathbf{X} = (x_1). \]
Suppose that the vector $\mathbf{X} \in \mathcal{X}$ is decomposable
\[ \mathbf{X} = (X_1, X_2, \ldots, X_T), \quad 1 \leq T \leq n, \]
where for each $t = 1, \ldots, T$, $X_t$ can be multidimensional.

Suppose that $f(x_{1:t})$ is decomposable:
\[ f(x_{1:t}) = f_1(x_1)f_2(x_2 | x_1) \cdots f_t(x_t | x_1, \ldots, x_{t-1}), \quad 1 \leq t \leq n. \]

**Sequential Generation of $\mathbf{X}$:**
\[ \mathbf{X} = (x_1, x_2). \]
Suppose that the vector $X \in \mathcal{X}$ is decomposable

$$X = (X_1, X_2, \ldots, X_T), \quad 1 \leq T \leq n,$$

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

Suppose that $f(x_{1:t})$ is decomposable:

$$f(x_{1:t}) = f_1(x_1)f_2(x_2 | x_1) \cdots f_t(x_t | x_1, \ldots, x_{t-1}), \quad 1 \leq t \leq n.$$

**Sequential Generation of $X$:**

$$X = (x_1, x_2, x_3).$$
Suppose that the vector $X \in \mathcal{X}$ is decomposable

$$X = (X_1, X_2, \ldots, X_T), \quad 1 \leq T \leq n,$$

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

Suppose that $f(x_{1:t})$ is decomposable:

$$f(x_{1:t}) = f_1(x_1)f_2(x_2 | x_1) \cdots f_t(x_t | x_1, \ldots, x_{t-1}), \quad 1 \leq t \leq n.$$

Sequential Generation of $X$:

$$X = (x_1, x_2, x_3, x_4, \cdots).$$
The Crude Sequential Monte Carlo (CSMC)

- Suppose that the vector $\mathbf{X} \in \mathcal{X}$ is decomposable
  $$\mathbf{X} = (X_1, X_2, \ldots, X_T), \quad 1 \leq T \leq n,$$
  where for each $t = 1, \ldots, T$, $X_t$ can be multidimensional.

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

**Sequential Generation of $\mathbf{X}$:**

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

**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,

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

X = ()

\[ X = () \]

\[ \mathbb{P}(T = 1 \mid T > 0) \]
\[ \mathbb{P}(T > 1 \mid T > 0) \]

\[ \{T = 1\} \]
\[ \{T > 1\} \]

\[ \mathbb{P}(T = 2 \mid T > 1) \]
\[ \mathbb{P}(T > 2 \mid T > 1) \]

\[ \{T = 2\} \]
\[ \{T > 2\} \]

\[ \mathbb{P}(T = 3 \mid T > 2) \]

\[ \{T = 3\} \]
\[ \{T > n - 2\} \]

\[ \mathbb{P}(T = n - 1 \mid T > n - 2) \]
\[ \mathbb{P}(T > n - 1 \mid T > n - 2) \]

\[ \{T = n - 1\} \]
\[ \{T > n - 1\} \]

\[ \mathbb{P}(T = n \mid T > n - 1) \]

\[ \{T = n\} \]
CSMC – $X$ Generation

$X = ()$

$P(T = 1 | T > 0)$

$X = (X_1, ?)$

$P(T > 1 | T > 0)$

$P(T = 1 | T > 0)$

$P(T > 1 | T > 0)$

$P(T = 2 | T > 1)$

$P(T > 2 | T > 1)$

$P(T > 1 | T > 0)$

$P(T = 2 | T > 1)$

$P(T > 2 | T > 1)$

$P(T = 3 | T > 2)$

$P(T > 3 | T > 2)$

$P(T = 3 | T > 2)$

$P(T > 3 | T > 2)$

$P(T = n - 1 | T > n - 2)$

$P(T > n - 1 | T > n - 2)$

$P(T > n - 2 | T > n - 2)$

$P(T = n - 1 | T > n - 2)$

$P(T > n - 1 | T > n - 2)$

$P(T = n | T > n - 1)$

$P(T > n | T > n - 1)$

$P(T = n | T > n - 1)$

$P(T > n | T > n - 1)$

$P(T = n | T > n - 1)$

$P(T > n | T > n - 1)$

$P(T = n | T > n - 1)$
CSMC – X Generation

\[
X = \emptyset
\]

\[
X = (X_1,?)
\]

\[
X = (X_1, , X_2,?)
\]
CSMC – \(X\) Generation

\[\begin{align*}
\{T > 0\} & \quad X = () \\
\{T = 1\} & \quad \mathbb{P}(T = 1 | T > 0) \\
\{T > 1\} & \quad \mathbb{P}(T > 1 | T > 0) \\
\{T = 2\} & \quad \mathbb{P}(T = 2 | T > 1) \\
\{T > 2\} & \quad \mathbb{P}(T > 2 | T > 1) \\
\{T = 3\} & \quad X = (X_1, \ldots) \\
\{T > n - 2\} & \quad \mathbb{P}(T = 3 | T > 2) \\
\{T = n - 1\} & \quad \mathbb{P}(T = n - 1 | T > n - 2) \\
\{T > n - 1\} & \quad \mathbb{P}(T > n - 1 | T > n - 2) \\
\{T = n\} & \quad \mathbb{P}(T = n | T > n - 1)
\end{align*}\]
Consider the SMC process tree and suppose for simplicity that

\[ P(T = t \mid T > t - 1) = 1 - P(T > t \mid T > t - 1) = 1/2 \quad \text{for } 1 \leq t \leq n-1, \]

and

\[ P(T = n \mid T > n - 1) = 1. \]

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

\[ P(\text{A single random walk reaches the } \{ T = n \} \text{ leaf}) = 2^{-(n-1)}. \]
Consider the SMC process tree and suppose for simplicity that
\[ \mathbb{P}(T = t \mid T > t - 1) = 1 - \mathbb{P}(T > t \mid T > t - 1) = 1/2 \quad \text{for } 1 \leq t \leq n-1, \]
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
\[ \mathbb{P}(\text{A single random walk reaches the } \{T = n\} \text{ leaf}) = 2^{-(n-1)}. \]

Do we have a problem?
Consider the SMC process tree and suppose for simplicity that
\[ P(T = t \mid T > t - 1) = 1 - P(T > t \mid T > t - 1) = 1/2 \]
for \( 1 \leq t \leq n-1 \), and
\[ P(T = n \mid T > n - 1) = 1. \]
We start a single walk from the root (\( \{ T > 0 \} \)) which ends at some leaf \( \{ T = t \} \). Then
\[ P(\text{A single random walk reaches the } \{ T = n \} \text{ leaf}) = 2^{-(n-1)}. \]

Do we have a problem?

**Rare-Event**

\[ H(x) = \begin{cases} 2^{n-1} (n^n) & \text{if } |x| = n \\ 0 & \text{otherwise} \end{cases} \Rightarrow E(H(X)) = ? \]
Consider the SMC process tree and suppose for simplicity that
\[ P(T = t \mid T > t - 1) = 1 - P(T > t \mid T > t - 1) = 1/2 \quad \text{for } 1 \leq t \leq n-1, \]
and
\[ P(T = n \mid T > n - 1) = 1. \]

We start a single walk from the root \( \{ T > 0 \} \) which ends at some leaf \( \{ T = t \} \). Then
\[ P(\text{A single random walk reaches the } \{ T = n \} \text{ leaf}) = 2^{-(n-1)}. \]

Do we have a problem?

**Rare-Event**

\[ H(x) = \begin{cases} 2^{n-1} (n^n) & \text{if } |x| = n \\ 0 & \text{otherwise} \end{cases} \Rightarrow E(H(X)) = n^n. \]
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).

\[
\begin{align*}
P(T = 1 | T > 0) & \quad P(T > 1 | T > 0) \\
\{ T = 1 \} & \quad \{ T > 1 \}, (10) \\
P(T = 2 | T > 1) & \quad P(T > 2 | T = 1) \\
\{ T = 2 \} & \quad \{ T > 2 \} \\
& \quad P(T = 3 | T > 2) \\
\{ T = 3 \} &
\end{align*}
\]
The Splitting Gain

1. \( P(T > t \mid T > t - 1) = 1 - 1/2^B. \)
The Splitting Gain

\[ P(T = 1 \mid T > 0) = 0.5 \]
\[ P(T > 1 \mid T > 0) = 0.5 \]
\[ P(T = 2 \mid T > 1) = 0.5 \]
\[ P(T > 2 \mid T = 1) = 0.5 \]
\[ P(T = 3 \mid T > 2) = 0.5 \]

1. \[ P(T > t \mid T > t - 1) = 1 - 1/2^B. \]
2. \[ P(T = n) = (1 - 1/2^B)^{n-1}. \]
1. \( P(T > t \mid T > t - 1) = 1 - 1/2^B. \)

2. \( P(T = n) = (1 - 1/2^B)^{n-1}. \)

3. For \( B = \lceil \log_2(n - 1) \rceil \), it holds that
   \[ P(T = n) \geq (1 - 1/2^{\log_2(n-1)})^{n-1} \rightarrow e^{-1}, \quad n \rightarrow \infty. \]
1. Develop efficient estimators \((P_t)\) for \(\mathbb{P}(T = t), t = 1, \ldots, n\).
2. Develop efficient estimators \((H_t)\) for \(\mathbb{E}(H(X) \mid |X| = t), t = 1, \ldots, n\).
3. Deliver \(C = \sum_{t=1}^{n} H_t \cdot P_t\), as an estimator of \(\mathbb{E}(H(X))\) (by the law of total expectation).
\[ T > t - 1 \]

\[ \{X_1, \ldots, X_B\} \]

\[ \mathbb{P}(T = t \mid T > t - 1) \]

\[ \mathbb{P}(T > t \mid T > t - 1) \]

\[ \{T = t\} \]

\[ \{T > t\} \]
The SSMC Algorithm (2)

\[ \begin{align*}
\{ T > t - 1 \} & \quad \{ T = t \} \quad \{ T > t \} \\
\mathbb{P}(T = t | T > t - 1) & \quad \mathbb{P}(T > t | T > t - 1) \\
\{ X_1', \ldots, X_{B'}' \} & \quad \{ X_1, \ldots, X_{B-B'} \}
\end{align*} \]
The SSMC Algorithm (2)

\[ \{ T > t \} \]

\[ \{ T = t \} \]

\[ \{ X_1', \ldots, X_{B'} \} \]

\[ \{ X_1, \ldots, X_{B-B'} \} \]

\[ \mathbb{P}(T = t | T > t - 1) \]

\[ \mathbb{P}(T > t | T > t - 1) \]

Estimators

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

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

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

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

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

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

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

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

is upper-bounded by a polynomial in $n$. 
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 $\mathcal{G}(n)$. 
Simple graph’s CV

Figure: Logarithmically scaled CV of LT as a function of $n$ for $\mathfrak{G}(n)$ networks.
Model 1 — Simple graph $n = 50$

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

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>$\hat{K}$</th>
<th>RE</th>
<th>REE</th>
</tr>
</thead>
<tbody>
<tr>
<td>LT</td>
<td>$1.93 \times 10^{-41}$</td>
<td>76.5%</td>
<td>99.7%</td>
</tr>
<tr>
<td>SSMC (ST)</td>
<td>$8.67 \times 10^{-38}$</td>
<td>2.48%</td>
<td>2.41%</td>
</tr>
</tbody>
</table>

**Table:** Performance of LT and ST for $\mathcal{G}(50)$ 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