Perfect Estimation with Imperfect Samples

Chang-Han Rhee

Centrum Wiskunde & Informatica

Scientific Meeting, CWI, November 27, 2015

Joint work with Peter W. Glynn

Monte Carlo Methods in Lay Terms

Repetitive random experiments

e.g. Coin flip: want to estimate $\mathbf{P}(\text{Head})$

- Flip the coin 100 times
- Count the number of head
- Divide by 100 and report the number

Monte Carlo Methods in Mathematical Terms

<u>Goal:</u> Compute $\mathbf{E}Y$

<u>Method:</u> Generate n iid copies $Y^{(1)}, \cdots, Y^{(n)}$ of Y and set

$$\overline{Y}(n) = \frac{1}{n}\sum_{i=1}^{n} Y^{(i)}.$$

By Central Limit Theorem (roughly speaking)

$$\overline{Y}(n) \overset{\mathcal{D}}{\approx} \mathbf{E}Y + \frac{\sigma_Y}{\sqrt{n}} N(0,1).$$

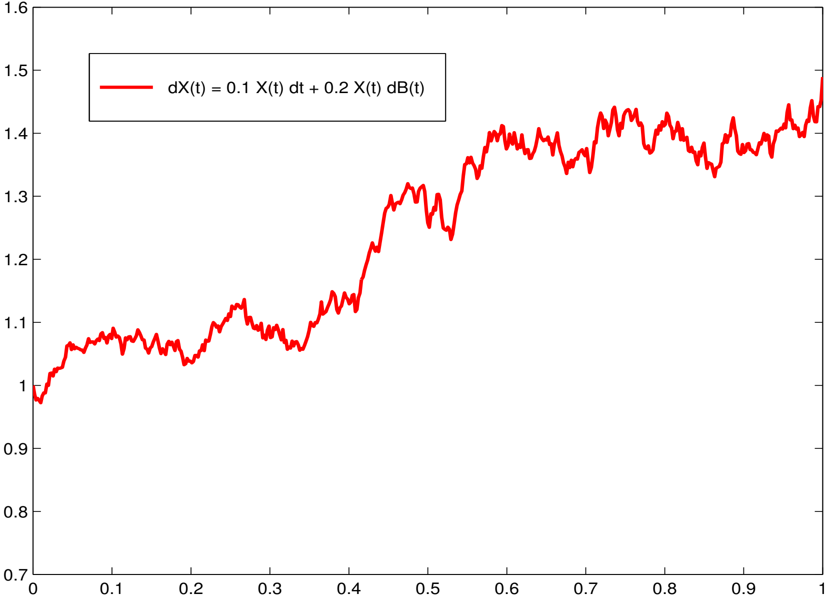
Example: Stochastic Differential Equations

Given

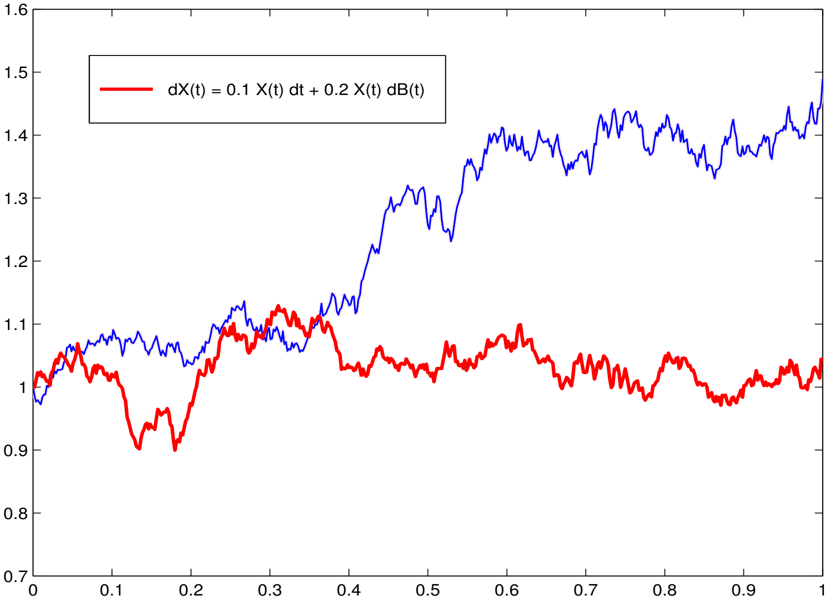
$$dX(t) = \mu(X(t))\,dt + \sigma(X(t))\,dB(t),$$

compute $\mathbf{E}Y$ ($\triangleq \mathbf{E}\,X(1)$).

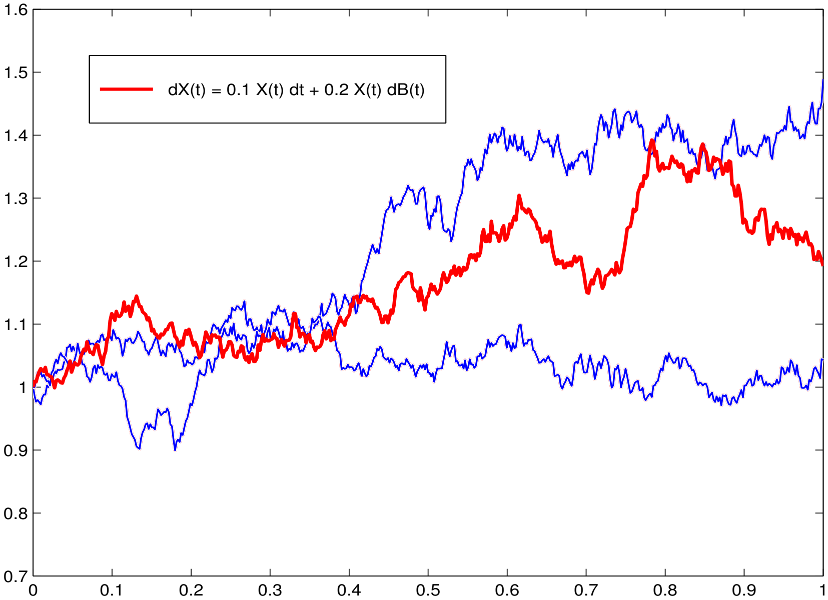
Example: Stochastic Differential Equations



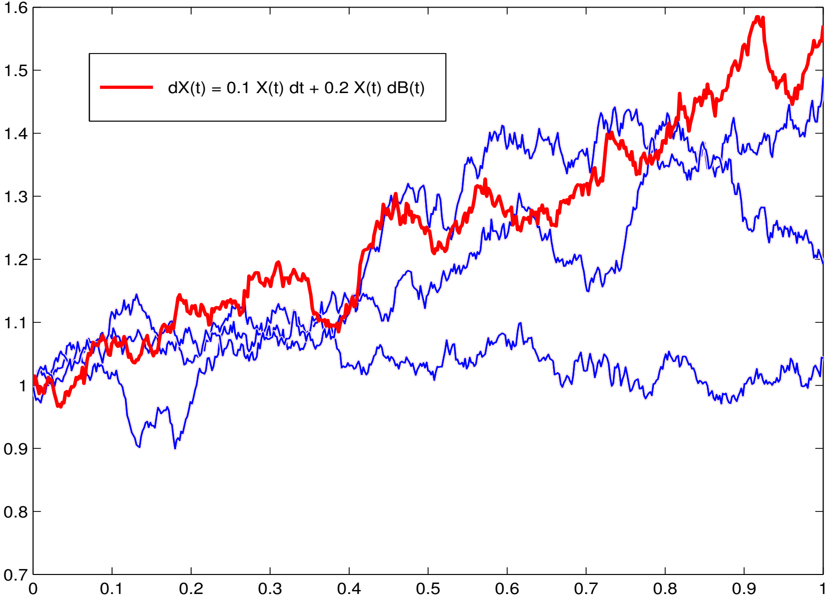
Example: Stochastic Differential Equations



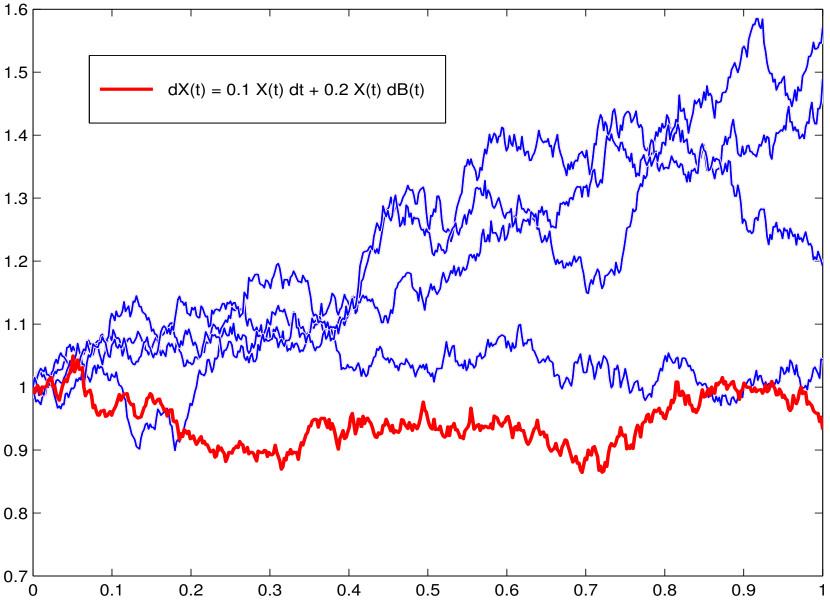
Example: Stochastic Differential Equations



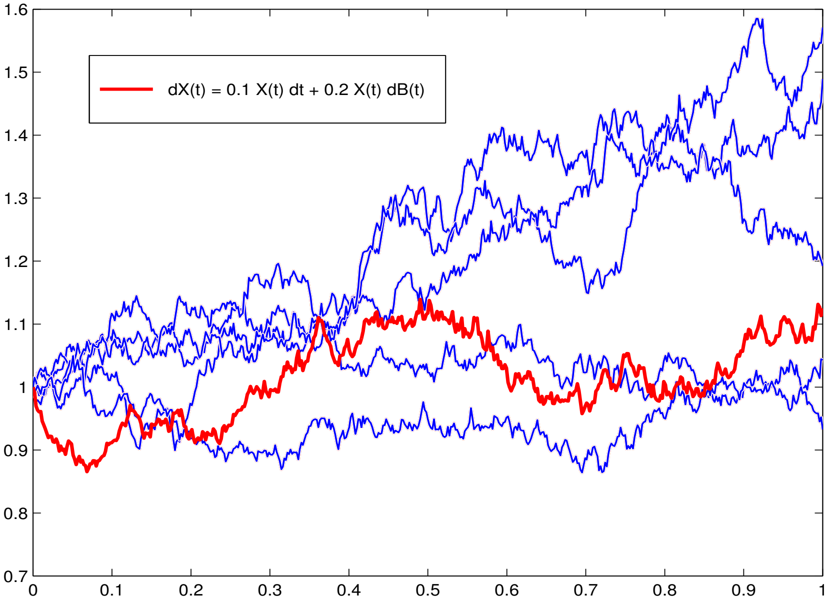
Example: Stochastic Differential Equations



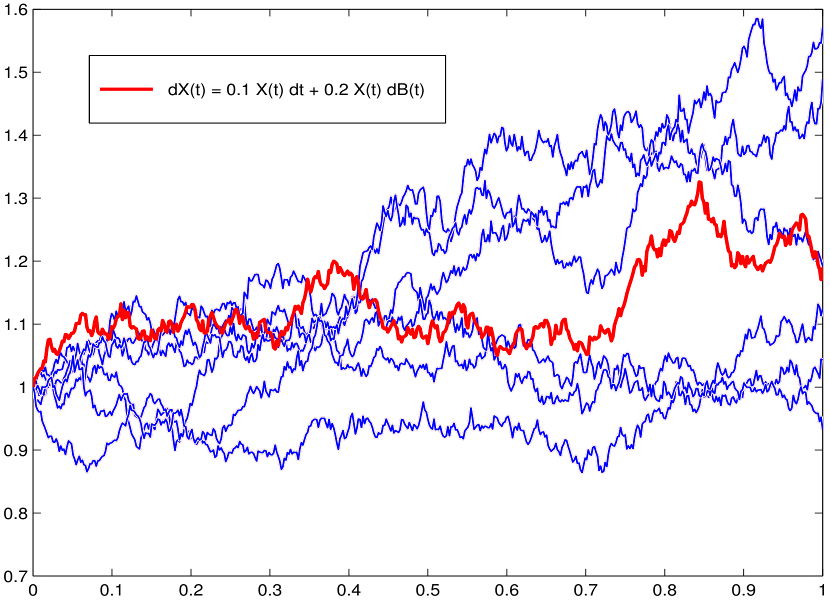
Example: Stochastic Differential Equations



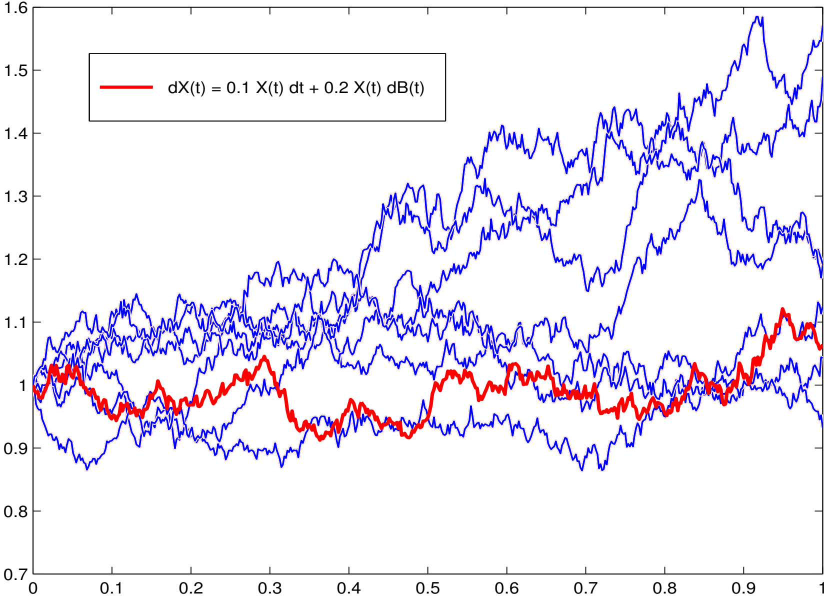
Example: Stochastic Differential Equations



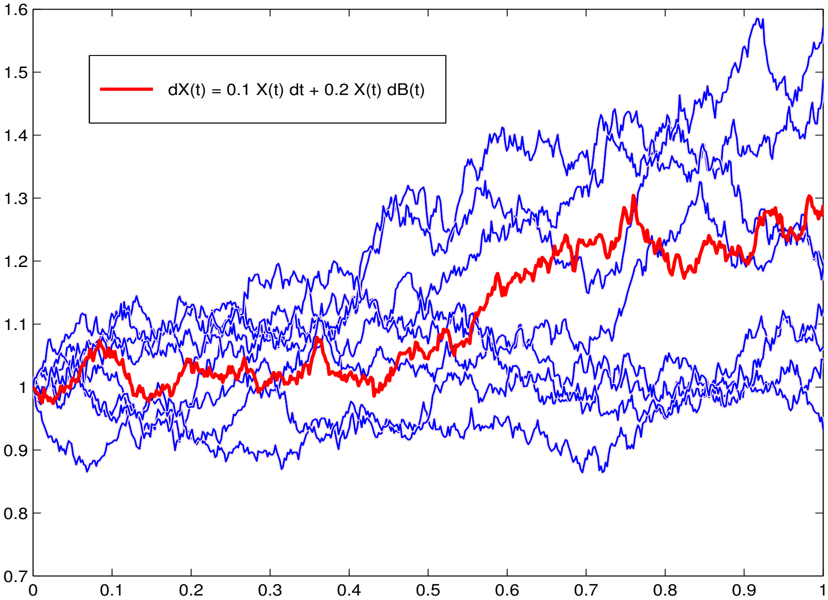
Example: Stochastic Differential Equations



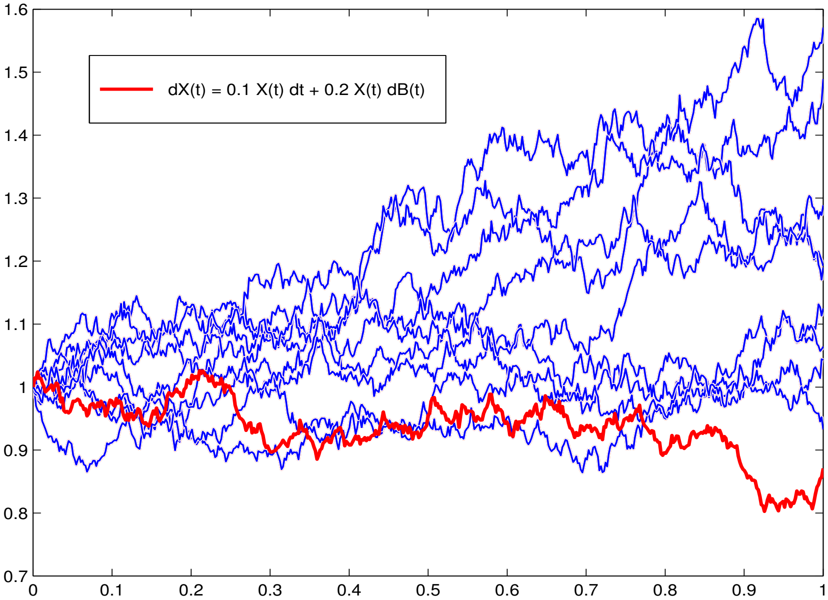
Example: Stochastic Differential Equations



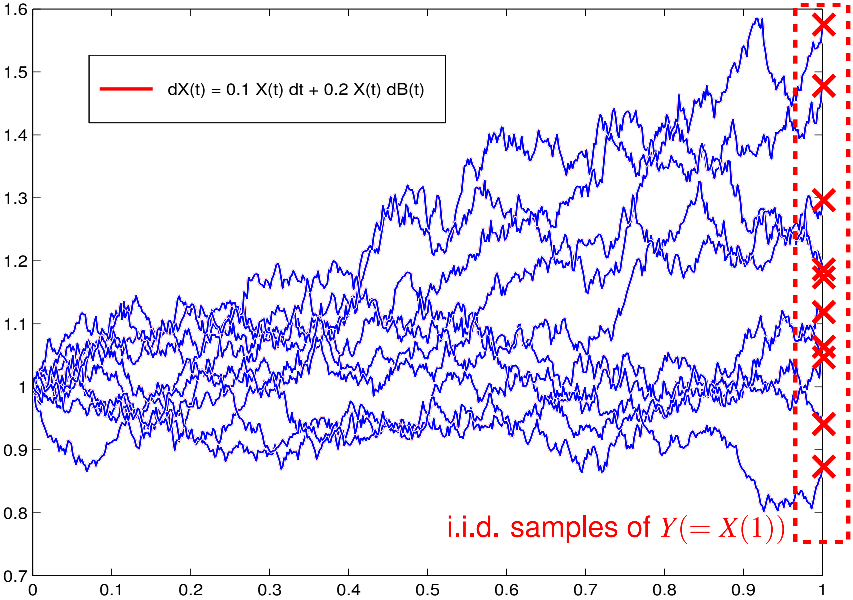
Example: Stochastic Differential Equations



Example: Stochastic Differential Equations



Example: Stochastic Differential Equations



Problem: we don’t know how to generate sample paths exactly

Instead, work with Discrete Approximation

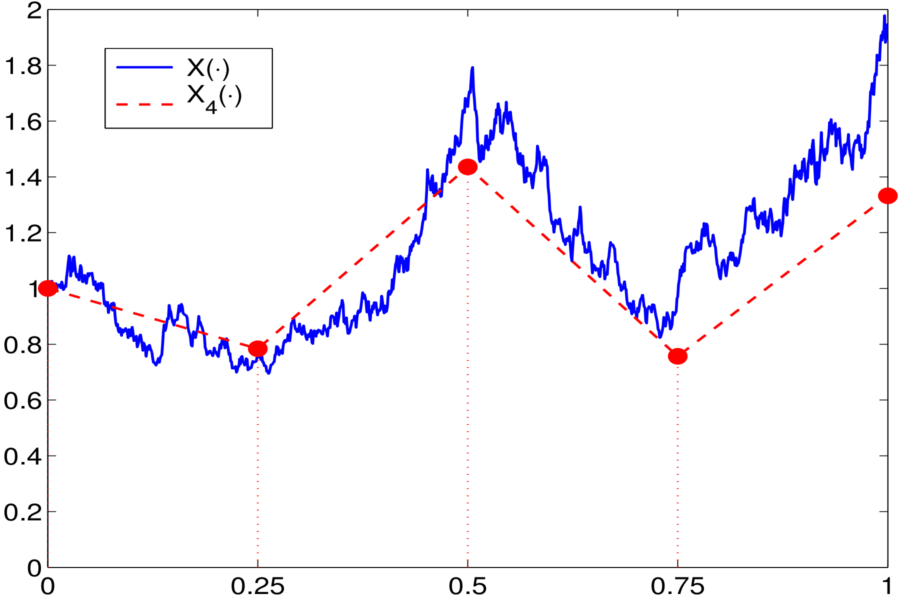
Original Equation:

$$dX(t) = \mu(X(t))\,dt + \sigma(X(t))\,dB(t)$$

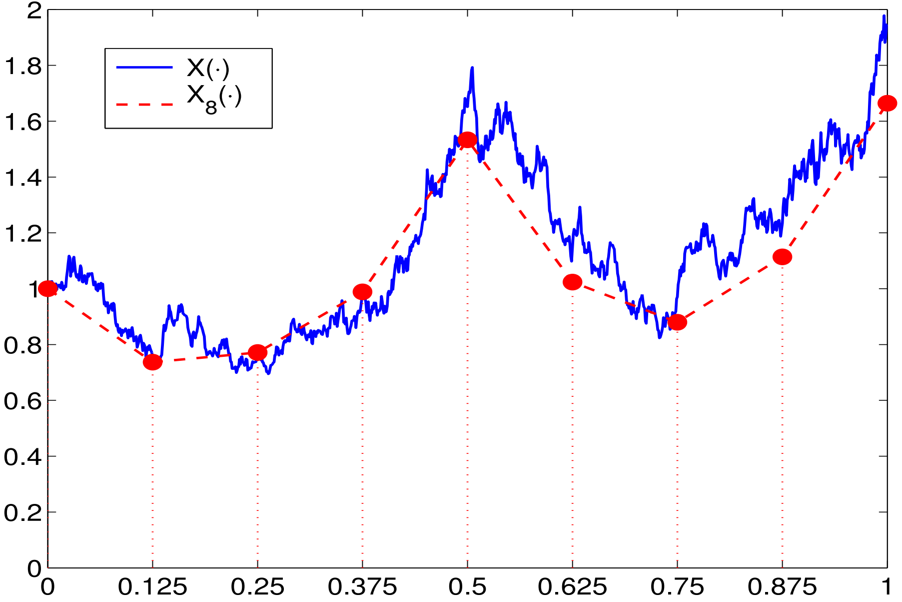
Discrete Approximation (Euler scheme):

$$X_m\left(\tfrac{k+1}{m}\right) - X_m\left(\tfrac{k}{m}\right) \;=\; \mu\left(X_m\left(\tfrac{k}{m}\right)\right)\tfrac{1}{m} \;+\; \sigma\left(X_m\left(\tfrac{k}{m}\right)\right)\left(B\left(\tfrac{k+1}{m}\right) - B\left(\tfrac{k}{m}\right)\right)$$

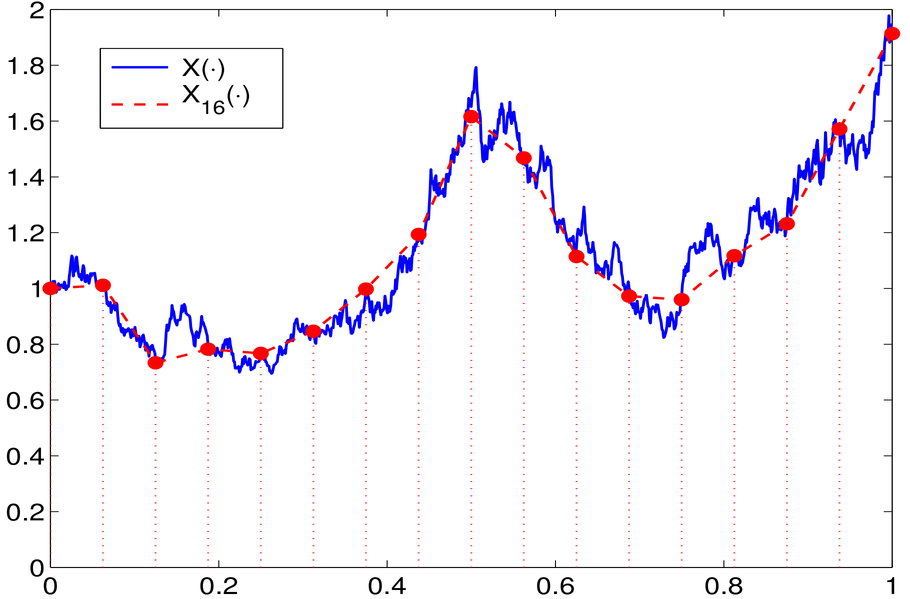
Instead, Work with Discrete Approximation (4 steps)



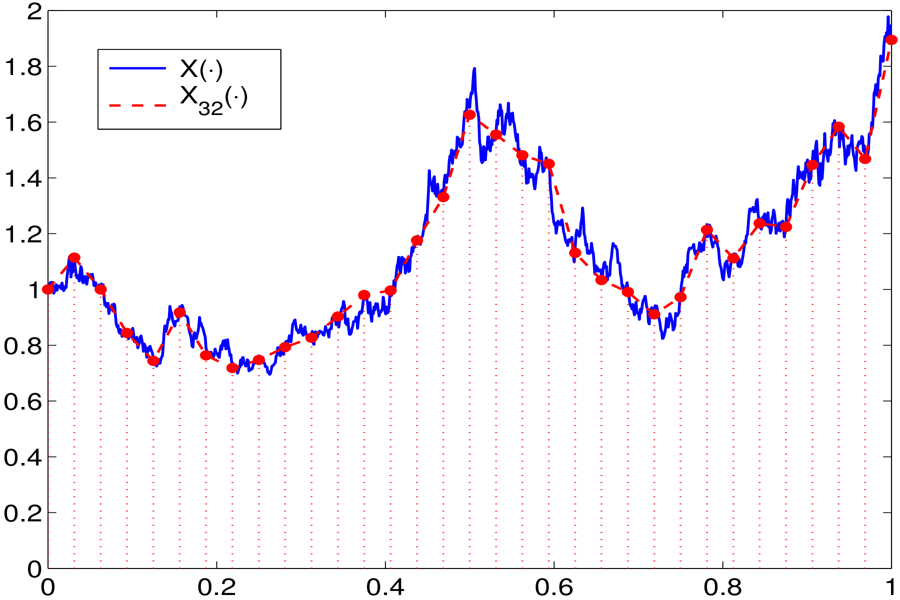
Instead, Work with Discrete Approximation (8 steps)



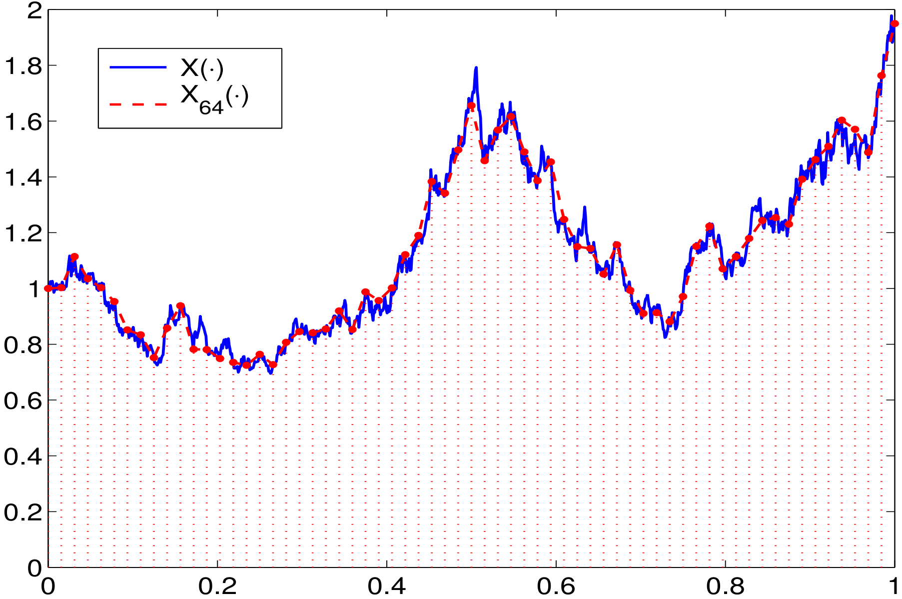
Instead, Work with Discrete Approximation (16 steps)



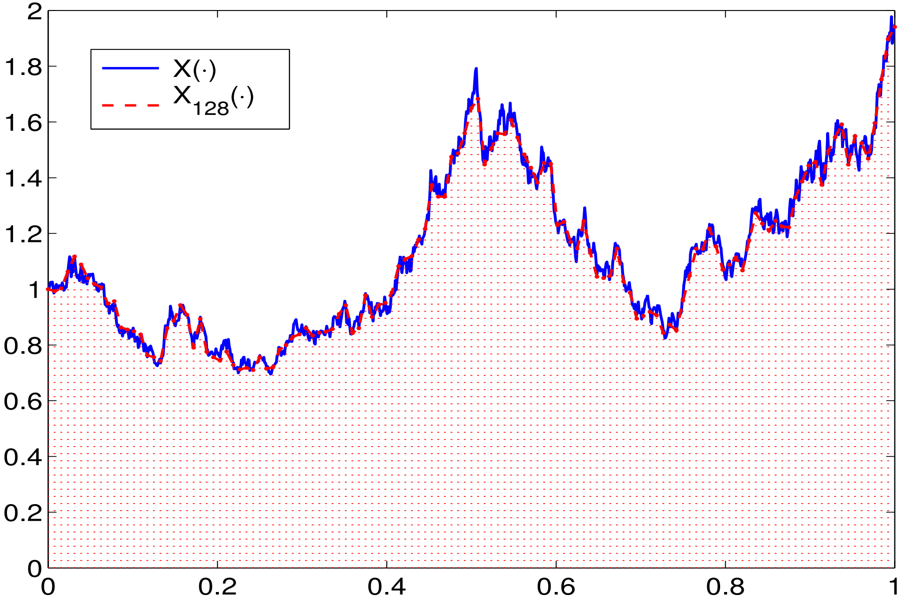
Instead, Work with Discrete Approximation (32 steps)



Instead, Work with Discrete Approximation (64 steps)



Instead, Work with Discrete Approximation (128 steps)



Consequence of Approximation

Now, error has an extra term due to approximation error — bias

$$Error \stackrel{\mathcal{D}}{\approx} \frac{\sigma_Y}{\sqrt{n}} N(0,1) + \mathcal{O}\left(\frac{1}{m}\right),$$

n: # samples
m: # time-steps

Total computation $c = \mathcal{O}(mn)$

- 1000 times more computation for 1 more significant digit

Note that if there were no bias, computation $c = \mathcal{O}(n)$

- 100 times more computation for 1 more significant digit

Moreover, error from bias is difficult to estimate

Consequence of Approximation

Now, error has an extra term due to approximation error — bias

$$Error \overset{\mathcal{D}}{\approx} \frac{\sigma_Y}{\sqrt{n}} N(0,1) + \mathcal{O}\left(\frac{1}{m}\right),$$

n: # samples
m: # time-steps

Total computation $c = \mathcal{O}(mn)$

- 1000 times more computation for 1 more significant digit

Note that if there were no bias, computation $c = \mathcal{O}(n)$

- 100 times more computation for 1 more significant digit

Moreover, error from bias is difficult to estimate

☹

Setup

<u>Goal:</u> Compute $\mathbf{E}Y$, where Y is difficult / impossible to generate exactly

Suppose that we have a sequence of approximations $(Y_m : m \geq 0)$:

- Y_m can be generated exactly
- $Y_m \stackrel{L^2}{\to} Y$ as $m \to \infty$

Setup

Goal: Compute $\mathbf{E}Y$, where Y is difficult / impossible to generate exactly

Suppose that we have a sequence of approximations $(Y_m : m \geq 0)$:

- Y_m can be generated exactly
- $Y_m \stackrel{L^2}{\to} Y$ as $m \to \infty$

Plan: Construct an easy-to-generate random variable Z such that $\mathbf{E}Z = \mathbf{E}Y$

Random Truncation Idea

- Think of Y as a sum of correction terms:

$$Y = \lim_{m\to\infty} Y_m = \lim_{m\to\infty} \Big(\underbrace{Y_0}_{\Delta_0} + \sum_{i=1}^{m} (\underbrace{Y_i - Y_{i-1}}_{\Delta_i})\Big) = \sum_{i=0}^{\infty} \Delta_i.$$

Random Truncation Idea

- Think of Y as a sum of correction terms:

$$Y = \lim_{m\to\infty} Y_m = \lim_{m\to\infty} \Big(\underbrace{Y_0}_{\Delta_0} + \sum_{i=1}^{m} \underbrace{(Y_i - Y_{i-1})}_{\Delta_i}\Big) = \sum_{i=0}^{\infty} \Delta_i.$$

- Not useful for computation yet because of infinite sum

Random Truncation Idea

- Think of Y as a sum of correction terms:

$$Y = \lim_{m\to\infty} Y_m = \lim_{m\to\infty} \Big(\underbrace{Y_0}_{\Delta_0} + \sum_{i=1}^{m} \underbrace{(Y_i - Y_{i-1})}_{\Delta_i} \Big) = \sum_{i=0}^{\infty} \Delta_i.$$

- Not useful for computation yet because of infinite sum

- What if truncate at random N; compensate the bias with weights?

$$\sum_{i=0}^{N} w_i \Delta_i, \qquad\qquad (N < \infty)$$

Random Truncation Idea

- Think of Y as a sum of correction terms:

$$Y = \lim_{m\to\infty} Y_m = \lim_{m\to\infty} \big(\underbrace{Y_0}_{\Delta_0} + \sum_{i=1}^{m} (\underbrace{Y_i - Y_{i-1}}_{\Delta_i})\big) = \sum_{i=0}^{\infty} \Delta_i .$$

- Not useful for computation yet because of infinite sum

- What if truncate at random N; compensate the bias with weights?

$$\sum_{i=0}^{N} w_i \Delta_i, \qquad (N < \infty)$$

- What is the right choice of w_i's?

Random Truncation Idea

- Think of Y as a sum of correction terms:

$$Y = \lim_{m\to\infty} Y_m = \lim_{m\to\infty} \Big(\underbrace{Y_0}_{\Delta_0} + \sum_{i=1}^{m} \underbrace{(Y_i - Y_{i-1})}_{\Delta_i} \Big) = \sum_{i=0}^{\infty} \Delta_i .$$

- Not useful for computation yet because of infinite sum

- What if truncate at random N; compensate the bias with weights?

$$\sum_{i=0}^{N} w_i \Delta_i, \qquad (N < \infty)$$

- What is the right choice of w_i's?

$$w_i = \frac{1}{\mathbf{P}(N \geq i)}$$

Perfect Estimation Possible with Imperfect Samplers!

We can prove that

$$\mathbf{E}\sum_{i=0}^{N} w_i \Delta_i = \mathbf{E}Y$$

i.e., $Z \triangleq \sum_{i=0}^{N} w_i \Delta_i \left(= \sum_{i=0}^{N} \frac{Y_i - Y_{i-1}}{\mathbf{P}(N \geq i)} \right)$ is an unbiased estimator of $\mathbf{E}Y$.

Implications

Efficient and perfectly unbiased estimators for

- Solutions of stochastic differential equations

  Rhee & Glynn (2012, 2015a)

- Stationary expectations of Markov chains

  Glynn & Rhee (2014)

- Sensitivity of intractible performance measures of Markov chains

  Rhee & Glynn (2015b, 2015c)

- Many more

Concluding Remarks

- Working with biased samples is often difficult
- A random truncation idea that can turn biased samples into perfect (i.e., unbiased) estimators
- A comprehensive theory is developed
- Extremely general—countless potential applications

Supplements

Proof of Unbiasedness

Recall:

$$Y = \lim_{m\to\infty} Y_m = \lim_{m\to\infty} \Big(\underbrace{Y_0}_{\Delta_0} + \sum_{i=1}^{m} (\underbrace{Y_i - Y_{i-1}}_{\Delta_i})\Big) = \sum_{i=0}^{\infty} \Delta_i.$$

Proof of Unbiasedness

Recall:

$$Y = \lim_{m\to\infty} Y_m = \lim_{m\to\infty} \Big(\underbrace{Y_0}_{\Delta_0} + \sum_{i=1}^{m} (\underbrace{Y_i - Y_{i-1}}_{\Delta_i})\Big) = \sum_{i=0}^{\infty} \Delta_i.$$

$$\mathbf{E} \sum_{i=0}^{N} w_i \Delta_i$$

Proof of Unbiasedness

Recall:

$$Y = \lim_{m\to\infty} Y_m = \lim_{m\to\infty} \Big(\underbrace{Y_0}_{\Delta_0} + \sum_{i=1}^{m} (\underbrace{Y_i - Y_{i-1}}_{\Delta_i})\Big) = \sum_{i=0}^{\infty} \Delta_i.$$

$$\mathbf{E} \sum_{i=0}^{N} w_i \Delta_i = \mathbf{E} \sum_{i=0}^{N} \frac{\Delta_i}{\mathbf{P}(N \geq i)}$$

Proof of Unbiasedness

Recall:

$$Y = \lim_{m\to\infty} Y_m = \lim_{m\to\infty} \Big(\underbrace{Y_0}_{\Delta_0} + \sum_{i=1}^{m} \underbrace{(Y_i - Y_{i-1})}_{\Delta_i}\Big) = \sum_{i=0}^{\infty} \Delta_i.$$

$$\mathbf{E} \sum_{i=0}^{N} w_i \Delta_i = \mathbf{E} \sum_{i=0}^{N} \frac{\Delta_i}{\mathbf{P}(N \geq i)} = \mathbf{E} \sum_{i=0}^{\infty} \frac{\Delta_i \mathbb{I}(N \geq i)}{\mathbf{P}(N \geq i)}$$

Proof of Unbiasedness

Recall:

$$Y = \lim_{m\to\infty} Y_m = \lim_{m\to\infty} \Big(\underbrace{Y_0}_{\Delta_0} + \sum_{i=1}^{m} (\underbrace{Y_i - Y_{i-1}}_{\Delta_i})\Big) = \sum_{i=0}^{\infty} \Delta_i.$$

$$\mathbf{E} \sum_{i=0}^{N} w_i \Delta_i = \mathbf{E} \sum_{i=0}^{N} \frac{\Delta_i}{\mathbf{P}(N \geq i)} = \mathbf{E} \sum_{i=0}^{\infty} \frac{\Delta_i \mathbb{I}(N \geq i)}{\mathbf{P}(N \geq i)} = \sum_{i=0}^{\infty} \frac{\mathbf{E}\Delta_i \mathbb{I}(N \geq i)}{\mathbf{P}(N \geq i)}$$

Proof of Unbiasedness

Recall:

$$Y = \lim_{m\to\infty} Y_m = \lim_{m\to\infty} \Big(\underbrace{Y_0}_{\Delta_0} + \sum_{i=1}^{m} (\underbrace{Y_i - Y_{i-1}}_{\Delta_i})\Big) = \sum_{i=0}^{\infty} \Delta_i.$$

$$\mathbf{E} \sum_{i=0}^{N} w_i \Delta_i = \mathbf{E} \sum_{i=0}^{N} \frac{\Delta_i}{\mathbf{P}(N \geq i)} = \mathbf{E} \sum_{i=0}^{\infty} \frac{\Delta_i \mathbb{I}(N \geq i)}{\mathbf{P}(N \geq i)} = \sum_{i=0}^{\infty} \frac{\mathbf{E}\Delta_i \mathbb{I}(N \geq i)}{\mathbf{P}(N \geq i)}$$

$$= \sum_{i=0}^{\infty} \frac{\mathbf{E}\Delta_i \, \mathbf{P}(N \geq i)}{\mathbf{P}(N \geq i)}$$

Proof of Unbiasedness

Recall:

$$Y = \lim_{m\to\infty} Y_m = \lim_{m\to\infty} \Big(\underbrace{Y_0}_{\Delta_0} + \sum_{i=1}^{m} (\underbrace{Y_i - Y_{i-1}}_{\Delta_i})\Big) = \sum_{i=0}^{\infty} \Delta_i.$$

$$\mathbf{E} \sum_{i=0}^{N} w_i \Delta_i = \mathbf{E} \sum_{i=0}^{N} \frac{\Delta_i}{\mathbf{P}(N \geq i)} = \mathbf{E} \sum_{i=0}^{\infty} \frac{\Delta_i \mathbb{I}(N \geq i)}{\mathbf{P}(N \geq i)} = \sum_{i=0}^{\infty} \frac{\mathbf{E}\Delta_i \mathbb{I}(N \geq i)}{\mathbf{P}(N \geq i)}$$

$$= \sum_{i=0}^{\infty} \frac{\mathbf{E}\Delta_i \, \mathbf{P}(N \geq i)}{\mathbf{P}(N \geq i)} = \sum_{i=0}^{\infty} \mathbf{E}\Delta_i$$

Proof of Unbiasedness

Recall:

$$Y = \lim_{m\to\infty} Y_m = \lim_{m\to\infty} \Big(\underbrace{Y_0}_{\Delta_0} + \sum_{i=1}^{m} (\underbrace{Y_i - Y_{i-1}}_{\Delta_i})\Big) = \sum_{i=0}^{\infty} \Delta_i.$$

$$\mathbf{E} \sum_{i=0}^{N} w_i \Delta_i = \mathbf{E} \sum_{i=0}^{N} \frac{\Delta_i}{\mathbf{P}(N \geq i)} = \mathbf{E} \sum_{i=0}^{\infty} \frac{\Delta_i \mathbb{I}(N \geq i)}{\mathbf{P}(N \geq i)} = \sum_{i=0}^{\infty} \frac{\mathbf{E}\Delta_i \mathbb{I}(N \geq i)}{\mathbf{P}(N \geq i)}$$

$$= \sum_{i=0}^{\infty} \frac{\mathbf{E}\Delta_i \, \mathbf{P}(N \geq i)}{\mathbf{P}(N \geq i)} = \sum_{i=0}^{\infty} \mathbf{E}\Delta_i$$

$= \mathbf{E}Y$ if, for example, $\sum_{i=0}^{\infty} \mathbf{E}|\Delta_i| < \infty$

Proof of Unbiasedness

Recall:

$$Y = \lim_{m\to\infty} Y_m = \lim_{m\to\infty} \Big(\underbrace{Y_0}_{\Delta_0} + \sum_{i=1}^{m} \underbrace{(Y_i - Y_{i-1})}_{\Delta_i} \Big) = \sum_{i=0}^{\infty} \Delta_i.$$

$$\mathbf{E} \sum_{i=0}^{N} w_i \Delta_i = \mathbf{E} \sum_{i=0}^{N} \frac{\Delta_i}{\mathbf{P}(N \geq i)} = \mathbf{E} \sum_{i=0}^{\infty} \frac{\Delta_i \mathbb{I}(N \geq i)}{\mathbf{P}(N \geq i)} = \sum_{i=0}^{\infty} \frac{\mathbf{E}\Delta_i \mathbb{I}(N \geq i)}{\mathbf{P}(N \geq i)}$$

$$= \sum_{i=0}^{\infty} \frac{\mathbf{E}\Delta_i \, \mathbf{P}(N \geq i)}{\mathbf{P}(N \geq i)} = \sum_{i=0}^{\infty} \mathbf{E}\Delta_i$$

$= \mathbf{E}Y$ if, for example, $\sum_{i=0}^{\infty} \mathbf{E}|\Delta_i| < \infty$

i.e., $Z \triangleq \sum_{i=0}^{N} \Delta_i / \mathbf{P}(N \geq i)$ is an unbiased estimator of $\mathbf{E}Y$.

(Rhee & Glynn, 2012)