Variance reduction[a]

Idea: Since the Monte Carlo Error is

$$E[Y] - \frac{1}{M}\sum_{j=1}^{M} Y(\omega_j) \approx C_\alpha \sqrt{\frac{\mathrm{Var}[Y]}{M}}$$

then we introduce techniques to ”reduce” $\mathrm{Var}[Y]$ while keeping $E[Y]$ unchanged.

For these techniques to be efficient, we need to use particular features of Y ...

---

[a] See for instance *Monte Carlo methods in financial engineering*, by P. Glasserman

Control Variates Suppose that we want to compute $E[Y]$ and instead of sampling just Y_j we also sample an auxiliary r.v., X_j for which we know $E[X]$.

Then, for a given β, we consider the unbiased estimator

$$V_M = \frac{1}{M}\sum_{j=1}^{M} Y_j - \beta\frac{1}{M}\sum_{j=1}^{M}(X_j - E[X])$$

Questions:

– Is there a way to choose optimally β to minimize $\mathrm{Var}[V_M]$?

– Does the strategy reduce the computational effort?

Now we compute

$$
\begin{aligned}
\mathrm{Var}[V_M](\beta) &= \frac{1}{M}\mathrm{Var}[Y - \beta X] \\
&= \frac{1}{M}\{\mathrm{Var}[Y] + \beta^2 \mathrm{Var}[X] - 2\beta \mathrm{Cov}[Y, X]\}
\end{aligned}
$$

and we minimize over β, yielding

$$\beta^* = \frac{\mathrm{Cov}[Y, X]}{\mathrm{Var}[X]}$$

and

$$\mathrm{Var}[V_M](\beta^*) = \frac{1}{M}\mathrm{Var}[Y]\left(1 - \frac{(\mathrm{Cov}[Y, X])^2}{\mathrm{Var}[X]\mathrm{Var}[Y]}\right) < \frac{1}{M}\mathrm{Var}[Y]$$

Obs:

-As long as X is correlated with Y the above procedure reduces the variance ...

- In practice, we can approximate β^* by using sample covariances and variances (at least for large M)

Does control variates pay in terms of computational work?

Assume that the work to generate the pair (X_j, Y_j) is twice the work to generate Y_j.

Method	# samples	Computat. Work
MC	$\left(\frac{C}{\epsilon}\right)^2 \mathrm{Var}[Y]$	$\left(\frac{C}{\epsilon}\right)^2 \mathrm{Var}[Y]$
MC + Cont.Var.	$\left(\frac{C}{\epsilon}\right)^2 \mathrm{Var}[V(\beta^*)]$	$2\left(\frac{C}{\epsilon}\right)^2 \mathrm{Var}[V(\beta^*)]$

We see that the strategy only pays if

$$\mathrm{Var}[Y] > 2\mathrm{Var}[V(\beta^*)]$$

i.e.

$$1 > 2(1-(\rho_{XY})^2) \text{ iff } (\rho_{XY})^2 > 1/2.$$

Obs: As said before, we need to use some knowledge of Y to find a suff. correlated X!!

Antithetic Variates Let $Y = g(X)$ and s.t. X has a symmetric distribution around its mean. Assume $E[X] = 0$.

Then X and $-X$ are identically distributed, yielding

$$E[g(X)] = E[g(-X)]$$

and

$$E[Y] = E\left[\frac{g(X) + g(-X)}{2}\right].$$

We then use the unbiased estimator

$$V_M = \frac{1}{M}\sum_{j=1}^{M}\frac{g(X_j) + g(-X_j)}{2}$$

Again we may ask if it pays to do so in terms of computational work . . .

We have

$$\mathrm{Var}[V_M] = \frac{\mathrm{Var}[g(X) + g(-X)]}{4M}.$$

Then, assuming that computing the pair $(g(X), g(-X))$ takes double the work for $g(X)$, we need

$$2\mathrm{Var}[V_1] \leq \mathrm{Var}[Y]$$

i.e.

$$\mathrm{Var}[g(X) + g(-X)] \leq 2\mathrm{Var}[g(X)]$$

or just

$$\mathrm{Cov}[g(X), g(-X)] < 0.$$

Obs: If g is linear then $\mathrm{Var}[g(X) + g(-X)] = 0$ so we expect the method to work for functions that are close to linear

On the other hand, for $g(X) = X^2$,

$$\mathrm{Cov}[g(X), g(-X)] = \mathrm{Var}[g(X)] > 0$$

and it is worse to apply antithetic variates than to use standard Monte Carlo!

Importance Sampling

Idea: Change the probability measure to reduce the variance!!

Let ρ_X, ρ_Z be pdfs s.t. $\frac{\rho_X(x)}{\hat{\rho}_Z(x)} < C$. Then

$$\begin{aligned} E[Y] &= E[g(X)] \\ &= \int_{\mathbb{R}} g(x)\rho_X(x)dx \\ &= \int_{\mathbb{R}} \underbrace{g(x)\frac{\rho_X(x)}{\hat{\rho}_Z(x)}}_{=\hat{g}(x)} \hat{\rho}_Z(x)dx \\ &= E[\hat{g}(Z)] \end{aligned}$$

However, $E[(g(X))^2]$ and $E[(\hat{g}(Z))^2]$ may be different!!

Example: Let $g > 0$ and choose

$$\hat{\rho} = Cg\rho,$$

with the normalizing constant

$$C = \left(\int_{\mathbb{R}} g\rho \right)^{-1}$$

Then

$$E[(\hat{g}(Z))^2] = \int_{\mathbb{R}} \frac{g^2\rho}{Cg\rho} \rho = \left(\int_{\mathbb{R}} g\rho \right)^2 = E[\hat{g}(Z)]^2$$

and we have a zero variance estimator!!

This approach is not practical because we need to know $C = \left(\int_{\mathbb{R}} g\rho\right)^{-1}$ which is equivalent to solve the original problem!

It indicates though that $\hat{\rho}$ has to follow the product $g\rho$ as much as possible.

Think of the case where g is nonzero for $x \in A$. How should you choose $\hat{\rho}$?

Numerical Example, Variance reduction: Ex 5.13

Look at J. Carlsson's implementation

`uppg5_13.m`

Uses antithetic variates and control variates.

Consider the computation of a call option on an index Z,

$$\pi_t = e^{-r(T-t)} E[\max(Z(T) - K, 0)], \tag{37}$$

where Z is the average of d stocks,

$$Z(t) \equiv \frac{1}{d} \sum_{i=1}^{d} S_i(t)$$

and

$$dS_i(t) = rS_i(t)dt + \sigma_i S_i(t) dW_i(t), \quad i = 1, \ldots, d$$

with volatilities

$$\sigma_i \equiv 0.2 * (2 + \sin(i)) \quad i = 1, \ldots, d.$$

The correlation between Wiener processes is given by

$$E[dW_i(t)dW_{i'}(t)] = exp(-2\,|i-i'|/d))dt \quad 1 \le i, i' \le d.$$

The goal of this exercise is to experiment with two different variance reduction techniques, namely the antithetic variates and the control variates.

From now on we take $d = 10$, $r = 0.04$ and $T = 0.5$ in the example above.

For the application of control variates to (37) use the geometric average

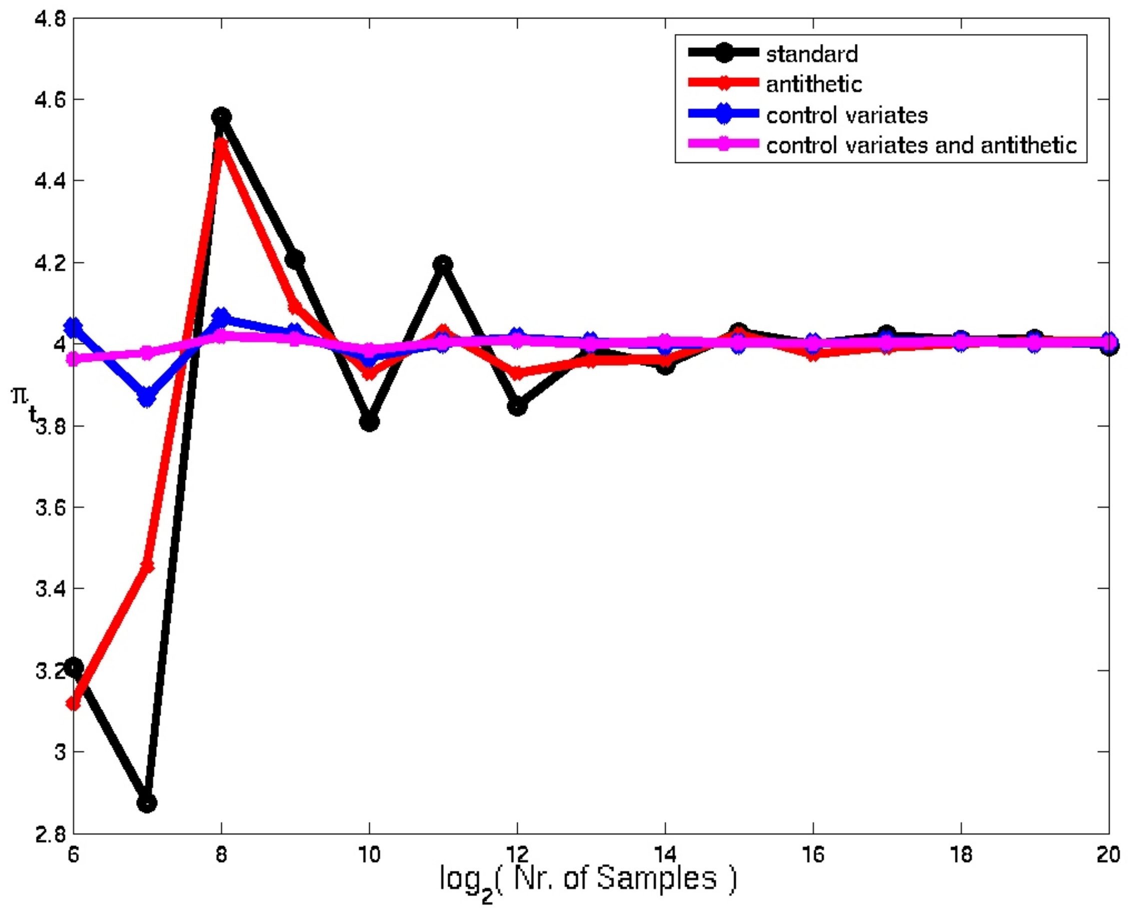
$$\hat{Z}(t) \equiv \{\prod_{i=1}^{d} S_i(t)\}^{\frac{1}{d}},$$

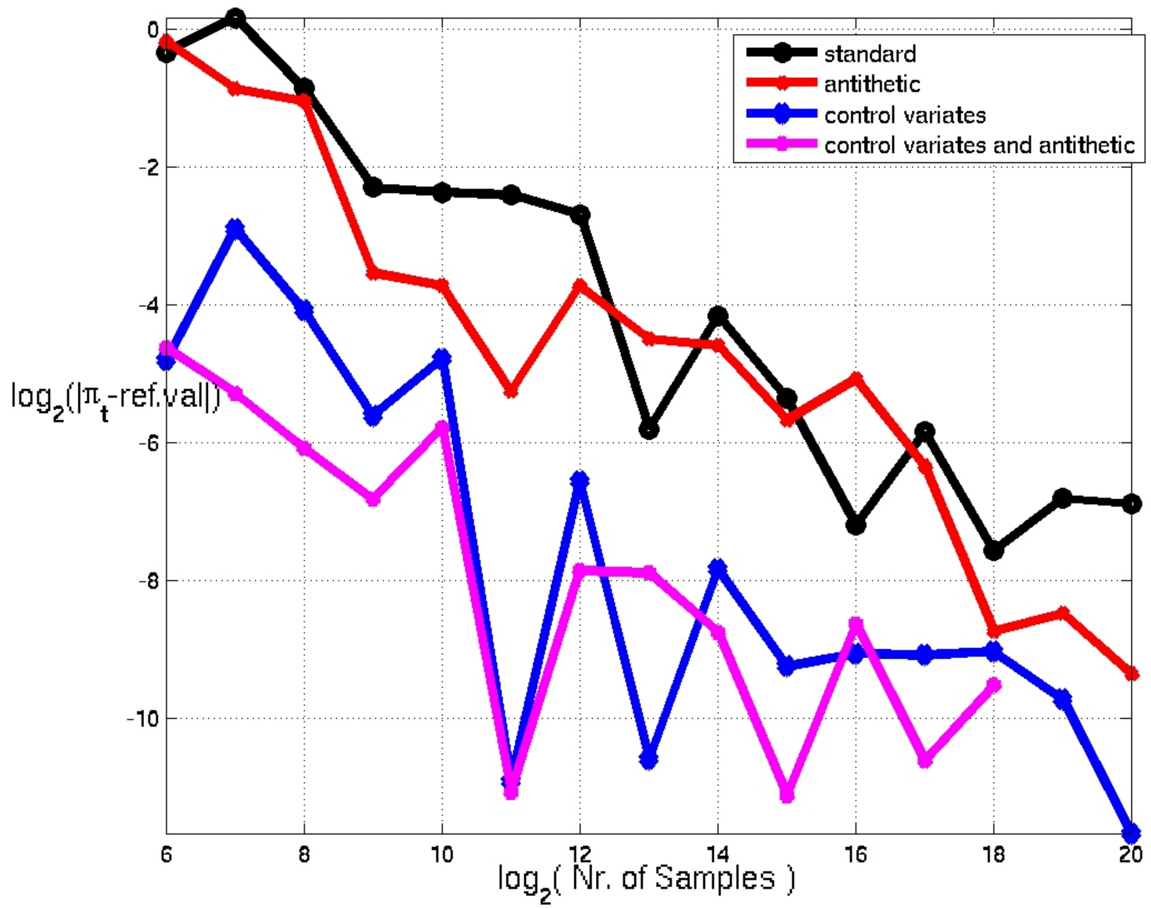
compute

$$\hat{\pi}_t = e^{-r(T-t)} E[\max(\hat{Z}(T) - K, 0)]$$

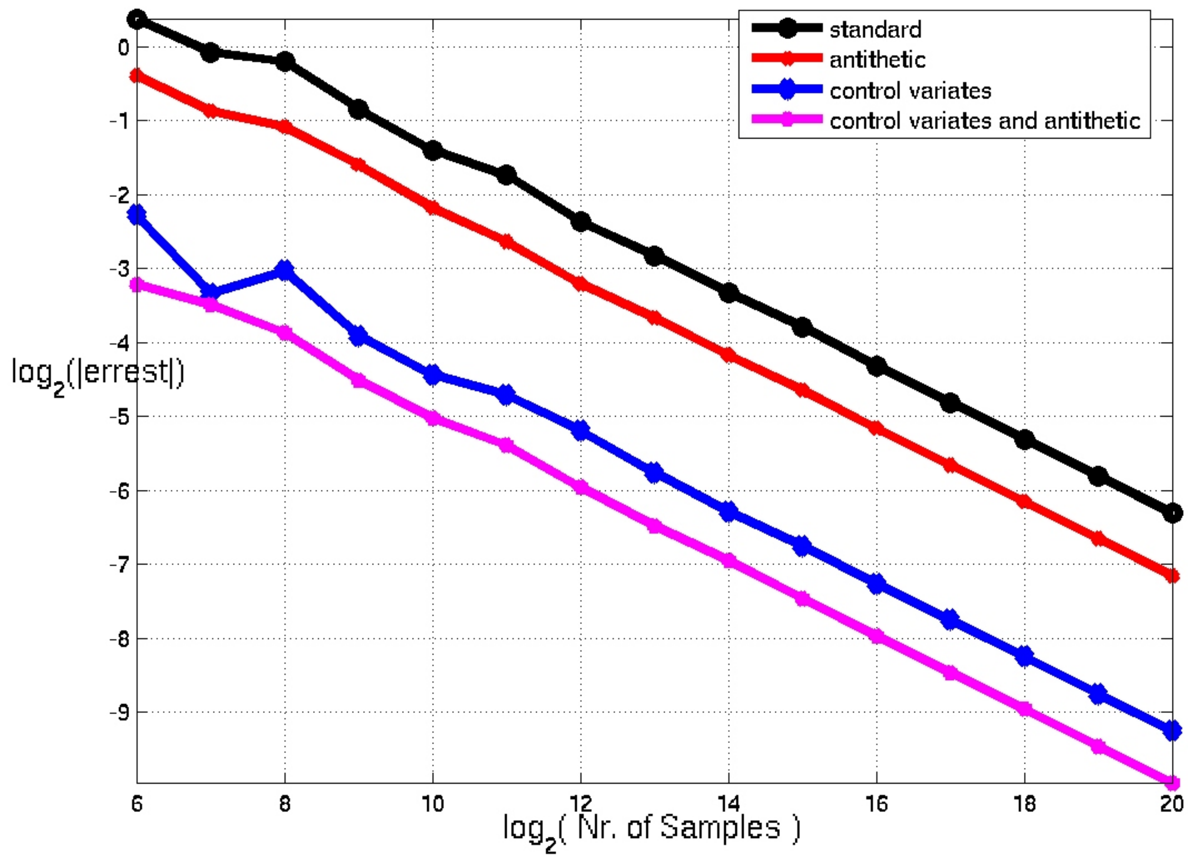
exactly (hint: see which SDE $\hat{Z}$ follows and find a way to apply Black-Scholes formula). Then approximate

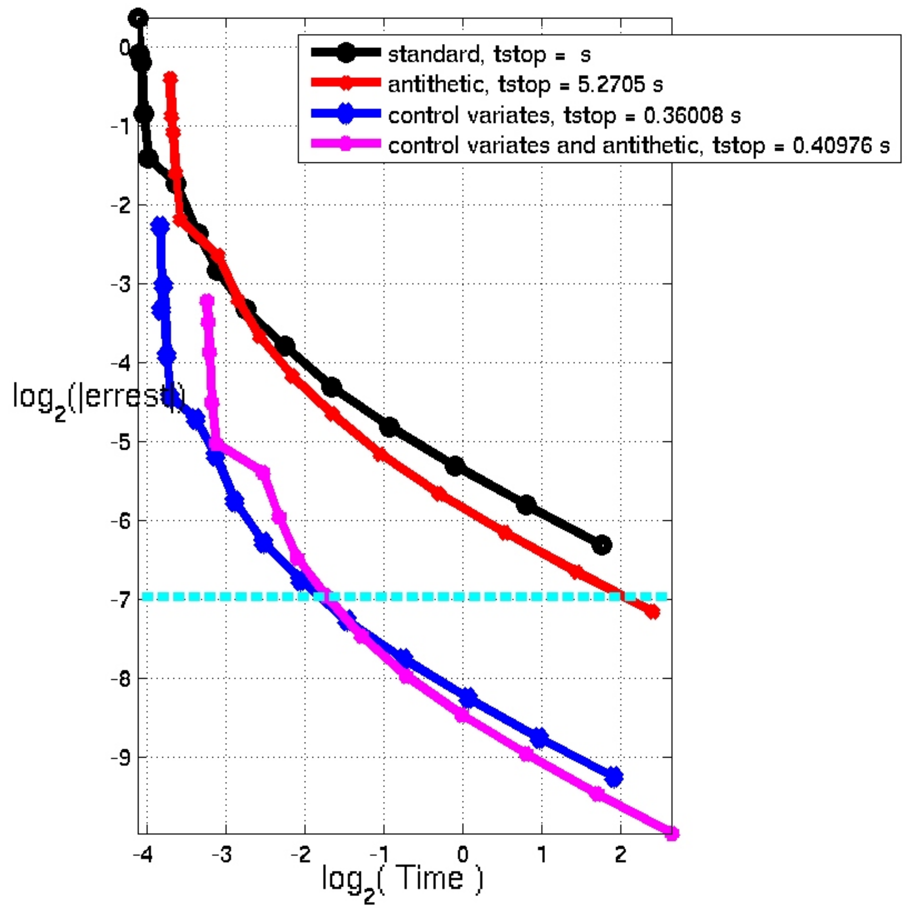
$$\begin{aligned}\pi_t \approx \hat{\pi}_t + \frac{e^{-r(T-t)}}{M}\sum_{j=1}^{M}\Big\{ & \max(Z(W(T,\omega_j)) - K, 0) \\ & - \max(\hat{Z}(W(T,\omega_j)) - K, 0)\Big\}.\end{aligned}$$

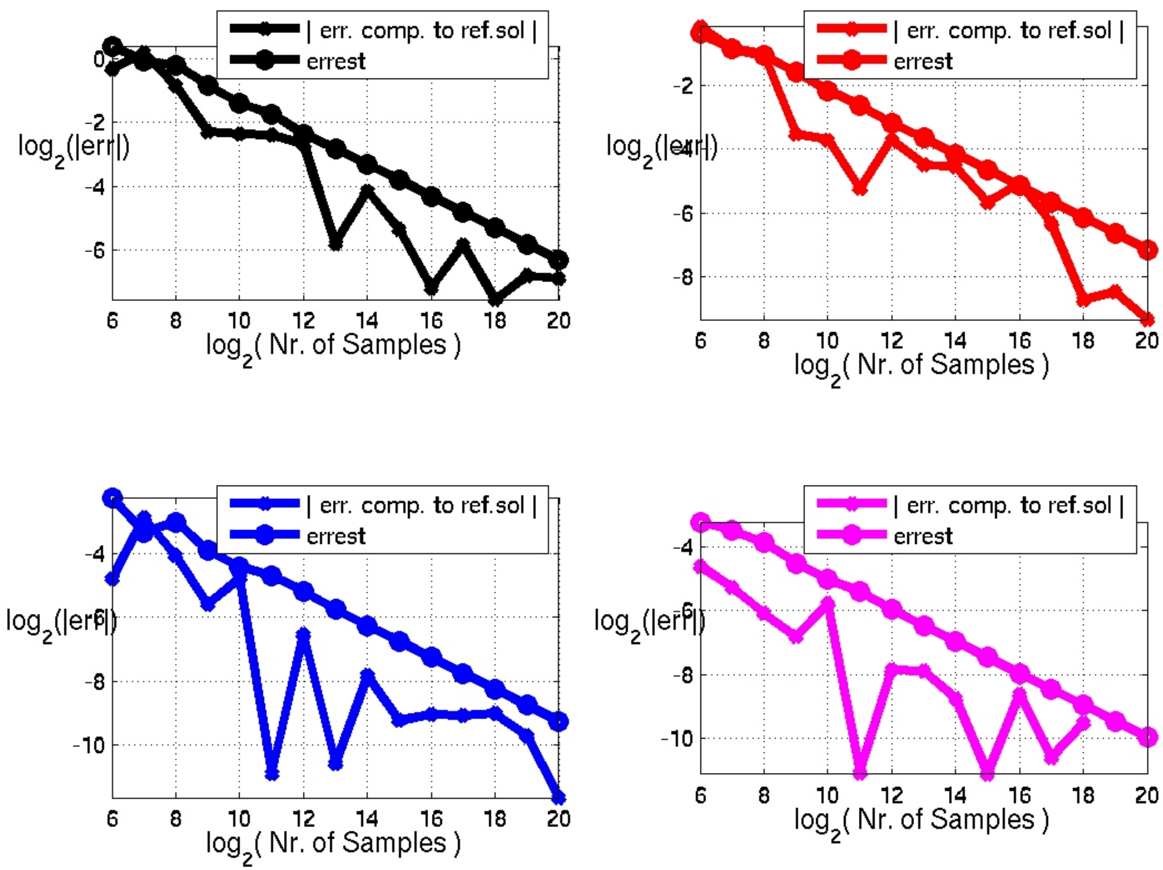

standard
antithetic
control variates
control variates and antithetic
π_t
log_2(Nr. of Samples)
4.8
4.6
4.4
4.2
4
3.8
3.6
3.4
3.2
3
2.8
6
8
10
12
14
16
18
20



standard
antithetic
control variates
control variates and antithetic
log2(|πt-ref.val|)
log2(Nr. of Samples)
0
-2
-4
-6
-8
-10
6
8
10
12
14
16
18
20


standard
antithetic
control variates
control variates and antithetic
log2(|errest|)
log2(Nr. of Samples)
0
-1
-2
-3
-4
-5
-6
-7
-8
-9
6
8
10
12
14
16
18
20


standard, tstop = s
antithetic, tstop = 5.2705 s
control variates, tstop = 0.36008 s
control variates and antithetic, tstop = 0.40976 s
log2(|errest|)
log2(Time)
0
-1
-2
-3
-4
-5
-6
-7
-8
-9
-4
-3
-2
-1
0
1
2


| err. comp. to ref.sol |
errest
log2(|err|)
log2(Nr. of Samples)
| err. comp. to ref.sol |
errest
log2(|err|)
log2(Nr. of Samples)
| err. comp. to ref.sol |
errest
log2(|err|)
log2(Nr. of Samples)
| err. comp. to ref.sol |
errest
log2(|err|)
log2(Nr. of Samples)

Can you improve the use of Control Variates for this example?