Control Variates with Quasi-Monte Carlo

Fred Hickernell
Hong Kong Baptist University

Christiane Lemieux
University of Calgary

Art B. Owen
Stanford University

Thanks

Fred and Christiane

Pierre L'Ecuyer

Louis Pelletier

Université de Montréal CRM

QMC + CV

$$I = \int_{(0,1)^s} f(x)dx, \quad s \text{ large}$$

Simple MC most widely used

$$\hat{I} = \frac{1}{n}\sum_{i=1}^{n} f(x_i), \quad x_i \sim U(0,1)^s$$

Improvements

1. Quasi-Monte Carlo, replacing x_i
2. Control variates, using related known integrals

Try both

Obvious to do this, but some surprises:

1. CV coefficient can change
2. good variates also change

QMC

Use low discrepancy x_i instead of random

Main methods:

nets Niederreiter, Sobol', Faure, Xing . . .

lattices Sloan + Joe, Korobov, Hua, Wang, Fang . . .

Randomized QMC: allows error estimation, and some errors cancel Lemieux + L'Ecuyer, Hong + Hickernell, Cranley + Patterson, Owen, Matousek . . .

Sampling Methods

1. Monte Carlo: $n^{-1/2}$
2. Quasi-Monte Carlo: $n^{-1}(\log n)^{d-1}$, but no practical error estimate
3. Randomized Quasi-Monte Carlo: replication based error estimates, and $n^{-3/2}(\log n)^{(d-1)/2}$ (for nets)

Rates are asymptotic under mild conditions on f

Control Variates

Know $\int h_j(x)dx = \theta_j$ for $j = 1, \ldots, J$

For $\beta = (\beta_1, \ldots, \beta_J)^T$, let:

$$\hat{I}_\beta = \frac{1}{n}\sum_{i=1}^{n}\Big(f(x_i) - \beta^T(h(x_i) - \theta)\Big)$$

Easily $E(\hat{I}_\beta) = I$ any β

Simple example (more later)

$J = 1$,

$f =$ Arithmetic mean option value, e.g. Asian option

$h = h_1 =$ Geometric mean option value

Get $\theta = \int h(x)dx$ by Black Scholes

$\beta = \beta_1 = 1$ works well because $f - h$ has small variance

CV variance

$$\text{Var}(\hat{I}_\beta) = \sigma^2_\beta / n$$

$$\sigma^2_\beta = E\big([f(x_i) - I - \beta^T(h(x_i) - \theta)]^2\big),$$

Unknown optimal β

$$\beta_{\text{mc}} = \left(\int (h(x) - \theta)(h(x) - \theta)^T dx\right)^{-1} \times \int (h(x) - \theta) f(x) dx.$$

Of course $\sigma^2_{\text{mc}} \equiv \sigma^2_{\beta_{\text{mc}}} \le \sigma^2$ Assume $\sigma^2_{\text{mc}} > 0$

Sample value $\hat{\beta}$

$$\hat{\beta}_{\text{mc}} = \left(\sum_{i=1}^n (h(x_i) - \hat{H})(h(x_i) - \hat{H})^T\right)^{-1} \times \sum_{i=1}^n (h(x_i) - \hat{H}) f(x_i)$$

$$\hat{H} = \frac{1}{n} \sum_{i=1}^n h(x_i)$$

MC and CV

Take $\beta = \hat{\beta}$ (Sample least squares)

$\hat{I}_{\hat{\beta}}$ almost as good as $\hat{I}_{\beta_{\text{mc}}}$

$\hat{I}_{\hat{\beta}}$ has small bias

Control variates forgiving of mild errors $\hat{\beta} - \beta_{\text{mc}}$

Common variance estimate

$$\widehat{\text{Var}}(\hat{I}_{\hat{\beta}_{\text{mc}}}) = \frac{1}{n - J - 1} \sum_{i=1}^{n} \Big(f(x_i) - \hat{I} - \hat{\beta}'(h(x_i) - \hat{H}) \Big)^2$$

or use $1/n$ vs $1/(n - J - 1)$

RQMC and CV

$$\text{Var}_{\text{rqmc}}(\hat{I}_\beta) = \text{Var}_{\text{rqmc}}\left(\hat{I} - \sum_{j=1}^{J} \beta_j \hat{H}_j\right)$$

Unknown optimal β

$$\beta_{\text{rqmc}} = \text{Cov}_{\text{rqmc}}(\hat{H}, \hat{H})^{-1}\text{Cov}_{\text{rqmc}}(\hat{H}, \hat{I})$$

$$\hat{I} = \frac{1}{n}\sum_{i=1}^{n} f(x_i) \quad \hat{H} = \frac{1}{n}\sum_{i=1}^{n} h(x_i)$$

Note:

β_{rqmc} need not be close to β_{mc}

β_{rqmc} depends on n

Sample least squares estimates β_{mc} not β_{rqmc}

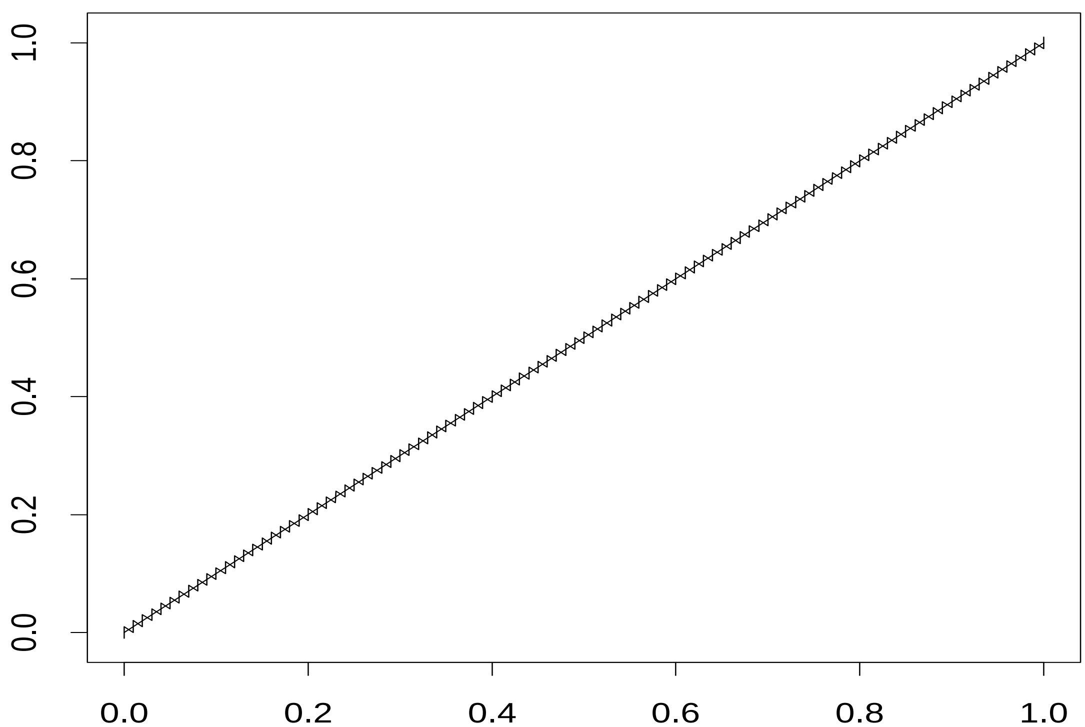
Good CV h should be correlated with f under RQMC, not necessarily under MC

Cautionary example

$$f(x) = (1 + 2\lfloor Mx \rfloor - Mx)/M$$

$$h(x) = x$$

$$M = 50$$



Cautionary ex. results

MC

$$\beta_{\text{mc}} = 1 - 2M^{-2} \doteq 1$$

$$\sigma^2_{\text{mc}} = 4\sigma^2(M^{-2} - M^{-4})$$

$$= \sigma^2/625.25, \qquad \text{for } M = 50$$

RQMC

Randomized $(0,1,1)$-net base $b = n$

IE n strata, $x_i \sim U[(i-1)/n, i/n)$ with $n = M$

$$\text{Var}_{\text{rqmc}}(\hat{I}) \doteq \frac{1}{12n^3}$$

$$\text{Var}_{\text{rqmc}}(\hat{I}_{\beta_{\text{mc}}}) \doteq \frac{2}{12n^3}$$

$$\beta_{\text{rqmc}} = -1$$

$$\text{Var}_{\text{rqmc}}(\hat{I}_{-1}) = 0$$

Stratification and CV

RQMC is something like stratification

For stratified sampling of $(0,1)$

$$\mathrm{Var}_{\text{strat}}\left(\frac{1}{n}\sum_{i=1}^{n} g(X_i)\right) \doteq \frac{1}{12n^3}\int_0^1 g'(x)^2 dx$$

Taking $g = f - h^T\beta$

$$\beta_{\text{strat}} \doteq \left(\int_0^1 h'(x)h'(x)^T dx\right)^{-1} \int h'(x)f'(x)dx$$

Correlation within strata counts more than correlation between strata

In the limit, a good CV has h' correlated with f' (uncentered)

Stratification of $(0,1)^s$

Asymptotic variance:

$$\frac{1}{12n^{1+2/s}}\int_{[0,1)^s}\Bigl\|\nabla\Bigl(f(x)-\beta^T h(x)\Bigr)\Bigr\|_2^2 dx$$

Asymptotic coefficient:

$$\Bigl(\int_{[0,1)^s}\nabla h\nabla h^T dx\Bigr)^{-1}\int_{[0,1)^s}\nabla h\nabla f\, dx$$

∇f has s components, one for each dimension of stratification

A good CV has gradient correlated with that of f

RQMC

$f(x) = f_G(x) + f_B(x)$

$\int f_G(x) f_B(x) dx = 0 \quad \int f_B(x) dx = 0$

RQMC “good” for f_G “bad” for f_B

Nets: f_B linear combination of indicators of small elementary intervals

Lattices: f_B part of f in dual lattice

Bad parts of h should correlate with those of f

$$\begin{aligned}
\hat{I}_\beta &= \frac{1}{n}\sum_{i=1}^{n} f(x_i) - \beta^T(h(x_i) - \theta) \\
&= \frac{1}{n}\sum_{i=1}^{n} f_G(x_i) - \beta^T(h_G(x_i) - \theta) \\
&+ \frac{1}{n}\sum_{i=1}^{n} f_B(x_i) - \beta^T(h_B(x_i) - \theta) \\
&\doteq I + \frac{1}{n}\sum_{i=1}^{n} f_B(x_i) - \beta^T(h_B(x_i) - \theta)
\end{aligned}$$

RQMC

nets

For nets we can almost replace ∇ by s-fold partial derivative

$$\frac{\partial^s}{\partial x^{(1)} \partial x^{(2)} \dots \partial x^{(s)}}$$

lattices

No analogue yet. Depends on limiting behavior of dual lattice.

Small example

$f(x) = \sin(\pi(x^{(1)} + x^{(2)}))$ on $(0,1)^2$

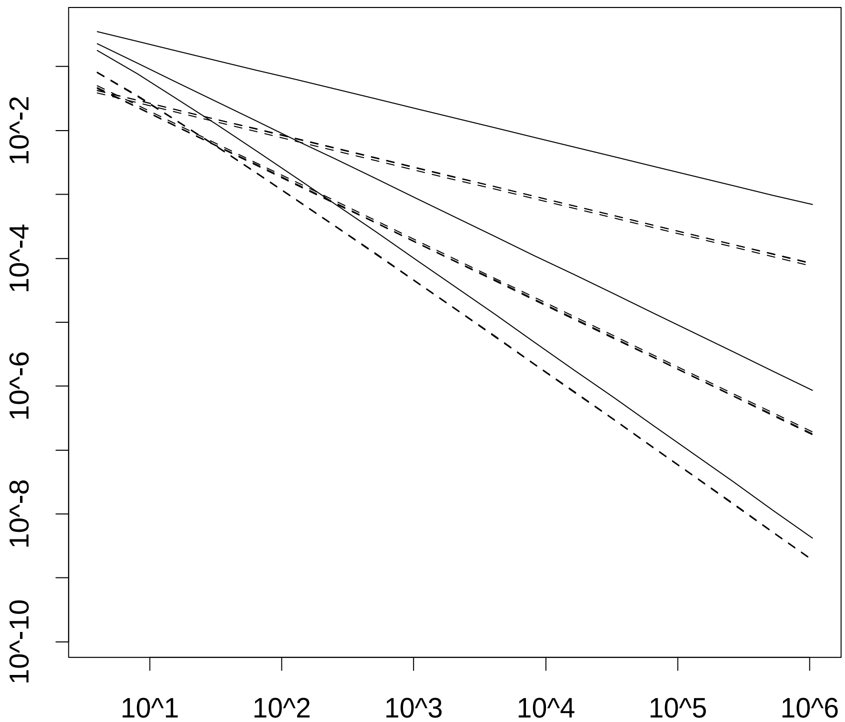
$h(x) = (x^{(1)} + x^{(2)} - 1)^3 - (x^{(1)} + x^{(2)} - 1)$

Asymptotic variance constants from low dimensional integrals

MC, Stratification, $(0, m, 2)$–nets

CV Coef $\rightarrow$ / Rate $\downarrow$	None 0	β_{mc} 2.675	β_{strat} 2.809	$\widetilde{\beta}_{\text{rnet}}$ 2.547	Gain
n^{-1}	0.5	0.0059	0.0072	0.0071	84.2
n^{-2}	0.825	0.0351	0.0333	0.0402	24.7
$n^{-3}\log(n)$	1.464	0.297	0.307	0.294	4.98

Asymptotic RMSE vs n



Dark lines for no CV

Dashed lines use CV

Top to bottom: MC, Strat, r-net

Equidistribution trumps CV

in this 2 dimensional example

Diminishing returns

For MC 10-fold variance reduction is a 10-fold time savings

For RQMC with Var$(\hat{I}) = O(n^{-2})$ a 10-fold variance reduction is a $10^{1/2} \doteq 3.16$-fold time saving

For RQMC with Var$(\hat{I}) = O(n^{-3})$ a 10-fold variance reduction is a $10^{1/3} \doteq 2.15$-fold time saving

Replication for β_{rqmc}

Run R independent replicates of RQMC with $\tilde{n}$ points

Get $\hat{I}_r$, $\hat{H}_r$ $r = 1, \dots, R$

RQMC is $R = 1$ with $\tilde{n} = n$

MC is $R = n$ with $\tilde{n} = 1$

Replication estimate

$$\hat{I}_\bullet = \frac{1}{R}\sum_{r=1}^{R}\hat{I}_r, \qquad \hat{H}_\bullet = \frac{1}{R}\sum_{r=1}^{R}\hat{H}_r$$

$$\hat{I}_{\hat{\beta}} = \hat{I}_\bullet - \hat{\beta}^T(\hat{H}_\bullet - \theta)$$

$$\hat{\beta} = \Bigl(\sum_{r=1}^{R}(\hat{H}_r - \hat{H}_\bullet)(\hat{H}_r - \hat{H}_\bullet)^T\Bigr)^{-1} \times \Bigl(\sum_{r=1}^{R}(\hat{H}_r - \hat{H}_\bullet)(\hat{I}_r - \hat{I}_\bullet)\Bigr).$$

Estimates β_{rqmc} for $\tilde{n} < n$ points

Error estimates

Minimize

$$\mathrm{SS}(\beta_0, \beta) = \sum_{r=1}^{R} (\hat{I}_r - \beta_0 - \hat{H}_r^T \beta)^2$$

with $\beta_0 = \hat{I}_{\hat{\beta}} + \hat{\beta}^T \theta$ and $\beta = \hat{\beta}$.

$\widehat{\mathrm{Var}}(\hat{\beta}) = \mathrm{SS}(\hat{I}_{\hat{\beta}}, \hat{\beta}) / (R(R - J - 1))$.

Internal replicates

Split $x_1, \dots, x_n$ into R consecutive blocks of $\tilde{n}$ points

A scrambled (λ, t, m, s)-net in base b splits into λb^r scrambled $(t, m-r, s)$-nets

Random shifted lattice rules with $n = b^m$ points split into $R = b^{m-\tilde{m}}$ such with $b^{\tilde{m}}$ points

Conservativism

Internal replicates tend to be negatively correlated

This deflates true error

and inflates error estimate

Aliasing

complicates internal replication of randomly shifted lattices

Picking R

$$\text{Var}_{\text{rqmc}}(\hat{I}_{\hat{\beta}}) = \frac{1}{R}\text{Var}_{\text{rqmc}}(\hat{I}_{\beta_{\text{rqmc},\tilde{n}}})(1 + O_p(R^{-1}))$$

The $1 + O_p(R^{-1})$ comes from errors in $\hat{\beta}$

Probably best to have bounded (or very slowly growing) R as $n \to \infty$

$$\begin{aligned}\text{Var}_{\text{rqmc}}(\hat{I}_{\beta_{\text{rqmc},\tilde{n}}}) &= O(\tilde{n}^{-2+\epsilon}), \quad \text{implies} \\ \text{Var}(\hat{I}_{\hat{\beta}}) &= O(\tilde{n}^{-2+\epsilon}R^{-1}) \\ &= O(R^{1-\epsilon}n^{-2+\epsilon})\end{aligned}$$

We probably need R over 5 or $J + 5$

Asian option example

Asset worth $S(t)$ at time t

Option pays $\max(0,(1/s)\sum_{i=1}^{s} S(t_i)-K)$ at T

$t_i = iT/s$

Geometric B.M. for $S(t_i)$

$$S(0)\exp\left[(r-\sigma^2/2)t_i + \sigma\sqrt{T/s}\sum_{j=1}^{i}\Phi^{-1}(x^{(j)})\right]$$

Discounted value:

$$f(x) = e^{-rT}\max\left(0, \frac{1}{s}\sum_{i=1}^{s} S(t_i, x) - K\right).$$

Control variate:

$$h_1(x) = e^{-rT}\max\left(0, \prod_{i=1}^{s} S(t_i, x)^{1/s} - K\right).$$

Parameters

Initial price	$S(0) = 100$
Strike price	$K = 120$
Risk-free interest	$r = 0.05$
Expiration	$T = 1$ year
Sampling	$s = 16$ times
Volatility	$\sigma = 0.3$

Out of money option.

Has about 0.17 probability of paying.

Truncation for BVHK

$f - A$ and $h - G$ are bounded, where

$$A(x) = e^{-rT}\left(\frac{1}{s}\sum_{i=1}^{s} S(t_i, x) - K\right)$$

$$G(x) = e^{-rT}\left(\prod_{i=1}^{s} S(t_i, x)^{1/s} - K\right)$$

Both A and G have known integrals

Maybe QMC works better for $f - A$

with or without the CV $h - G$

MC could be worse because $f - A$

can have large variance for out-of-money option

CV strategies, for MC

Name	*Estimate*
MC$_0$	$\hat{I}(f)$
MC$_1$	$\hat{I}(f - \beta_1 h_1) + \beta_1 I(h_1)$
MC$_3$	$\hat{I}(f - \beta_2 h_1 - \beta_3 A - \beta_4 G)$
	$+\beta_2 I(h_1) + \beta_3 I(A) + \beta_4 I(G)$
MC$_B$	$\hat{I}(f - A) + I(A)$
MC$_{BB}$	$\hat{I}(f - A - \beta_5(h_1 - G))$
	$+I(A) + \beta_5 I(h_1 - G)$

Same CV strategies used with RQMC

QMC$^{(2)}$: $R = 85$ reps of $(0, 2, 16)$-net in base 17

QMC$^{(3)}$: $R = 5$ reps of $(0, 3, 16)$-net in base 17
uses est of β_{mc}

Of course: MC has $n = 5 \times 17^3 = 24565$

Estimated coefficients

Coef	Of	In	MC	QMC$^{(2)}$	QMC$^{(3)}$
β_1	h	1	1.10	1.08	1.10
β_2	h	3	1.04	1.01	1.04
β_3	A	3	0.534	1.33	0.519
β_4	G	3	-0.525	-1.37	-0.510
β_5	$h - g$	BB	0.988	1.03	0.987

Remarks

1. Arith mean $>$ Geom mean, so $\beta_1 > 1$
2. A–coef plus G–coef $\doteq 0$
3. Only QMC$^{(2)}$ estimates β_{rqmc}

Estimated RMSE

	MC	$\mathrm{QMC}^{(2)}$	$\mathrm{QMC}^{(3)}$
0	4.41e−2	2.05e−2	3.70e−3
1	2.99e−3	2.16e−3	1.34e−3
3	2.08e−3	1.48e−3	1.04e−3
B	9.05e−2	1.69e−2	2.94e−3
BB	2.81e−3	1.52e−3	7.35e−4

Remarks

1. $\mathrm{QMC}^{(3)}$ beats $\mathrm{QMC}^{(2)}$ beats MC (always)
2. Methods that don't use CV do badly
3. MC_3 is best MC (as it should be)
4. $\mathrm{QMC}_3^{(2)}$ is best $\mathrm{QMC}^{(2)}$ (barely)
5. MC_1 almost as good as $\mathrm{QMC}_0^{(3)}$
6. $\mathrm{QMC}_{BB}^{(3)}$ best overall

Efficiency gains

CV gives more improvement than QMC

142 in replacing MC_0 by $\text{QMC}_0^{(3)}$

217 in replacing MC_0 by MC_1

450 in replacing MC_0 by MC_3

Synergy

3600 in replacing MC_0 by $\text{QMC}_{BB}^{(3)}$

BVHK

yields small gains

Conclusions

CV coefficients depend on sampling method

Effective variates can also depend on sampling method

QMC with sub-optimal CV coefficients worked well