

QMC for MCMC: Background and recent results

Art B. Owen

Su Chen

Stanford University

Stanford University

With contributions from:

Josef Dick, Makoto Matsumoto, Takuji Nishimura

Simple Monte Carlo

Used in virtually all sciences

$$\mu = \mathbb{E}(f(x)), \quad x \sim p$$

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n f(x_i), \quad x_i \text{ IID } p$$

Recall

$\mathbb{P}(\hat{\mu}_n \rightarrow \mu) = 1$ by Strong Law Large Numbers

If $\mathbb{E}(f(x)^2) < \infty$ then $\text{RMSE} = O(n^{-1/2})$

If $\mathbb{E}(f(x)^2) < \infty$ then Central Limit Theorem

Unfortunately:

MC is **SLOW**: one more digit accuracy \equiv 100 fold more work

MC is **HARD**: getting $x_i \sim p$ is challenging (for Boltzmann, Bayes, \dots)

But there's hope:

QMC improves **accuracy** from $O(n^{-1/2})$ to $O(n^{-1+\epsilon})$

MCMC broadens **applicability**

Talk in one slide

- 1) We want to combine the benefits of QMC and MCMC.
- 2) Sometimes we can, using QMC points that are “completely uniformly distributed” (CUD)
- 3) Previously:
 - (a) convergence required a finite state space, but
 - (b) the largest improvements were for continuous examples
- 4) Now:
 - (a) proven consistency on continuous state spaces
 - (b) more examples (some good, some disappointing)

Markov chain Monte Carlo

Let $\mathbf{x}_i = \phi(\mathbf{x}_{i-1}, \mathbf{v}_i)$ $\mathbf{v}_i \sim \mathbf{U}(0, 1)^d$ (Markov property)

Design $\phi(\cdot, \cdot)$ so that $\text{distn}(\mathbf{x}_i) \rightarrow p$

LLN for reasonable conditions

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n f(\mathbf{x}_i) \rightarrow \int f(\mathbf{x})p(\mathbf{x}) \, d\mathbf{x} \equiv \mu$$

Main MCMC algorithms

- Metropolis-Hastings
 - 1) While at \mathbf{x}_i **propose** move to $\mathbf{y}_{i+1} \sim Q(\mathbf{x}_i \rightarrow \mathbf{y}_{i+1})$
 - 2) **Accept** with probability $A(\mathbf{x}_i \rightarrow \mathbf{y}_{i+1})$
 - 3) If accepted, then $\mathbf{x}_{i+1} = \mathbf{y}_{i+1}$ else $\mathbf{x}_{i+1} = \mathbf{x}_i$
- Gibbs sampler
 - 1) Sample component j from $p(x_j \mid x_{ik}, k \neq j)$
 - 2) Cycle through j 's (sequentially, or randomly)

Recipe for QMC in MCMC

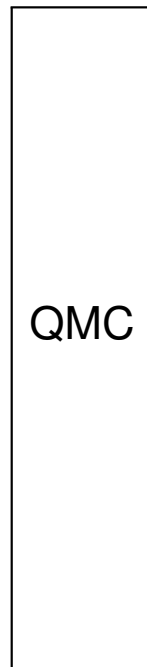
- 1) Each step takes d numbers in $(0, 1)$.
- 2) n steps require $u_1, \dots, u_{nd} \in (0, 1)$
- 3) MCMC uses $u_i \sim \mathbf{U}(0, 1)$
- 4) Replace IID by balanced points

Reasons for caution

- 1) We're using 1 point in $[0, 1]^{nd}$ with $n \rightarrow \infty$
- 2) Our sequence won't be Markovian

MCMC looks like QMC^T

Method	Rows	Columns	
QMC	n points	d variables	$1 \leq d \ll n \rightarrow \infty$
MCMC	r replicates	n steps	$1 \leq r \ll n \rightarrow \infty$



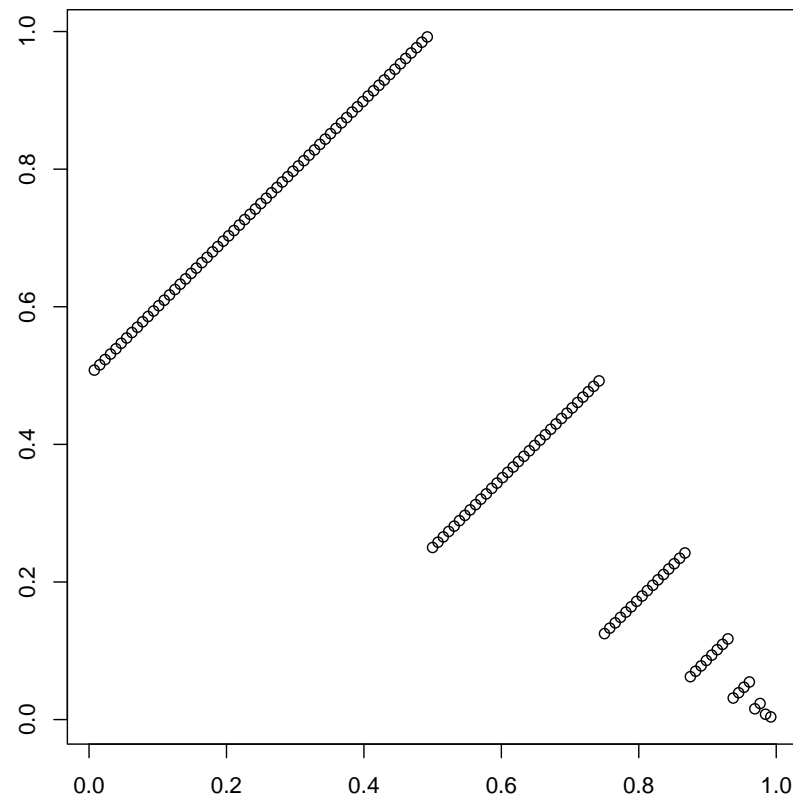
QMC based on equidistribution

MCMC based on ergodicity

Severe failure is possible

van der Corput $u_i \in [0, 1/2) \iff u_{i+1} \in [1/2, 1)$

u_{i+1} VS u_i



High proposal \iff low acceptance and vice versa

Morokoff and Caflisch (1993) describe heat particle going left only

Completely uniformly distributed

$u_1, u_2, \dots \in [0, 1]$ are CUD if

$D_n^*(z_1, \dots, z_n) \rightarrow 0$, where

$z_i = (u_i, \dots, u_{i+d-1})$

For all $d \geq 1$

Overlapping blocks

$$z_1 = (u_1, \dots, u_d)$$

$$z_2 = (u_2, \dots, u_{d+1})$$

$$\vdots \quad \vdots$$

$$z_n = (u_n, \dots, u_{n+d-1})$$

Chentsov (1967) shows we can use **non-overlapping blocks**

$$v_i = (u_{d(i-1)+1}, \dots, u_{di}) \quad \forall d$$

CUD ctd

CUD \equiv one of Knuth's definitions of randomness

Recommendations

- 1) Use all the d -tuples from your RNG
- 2) Be sure to pick a small RNG

As considered in

Niederreiter (1986)

Entacher, Hellekalek, and L'Ecuyer (1999)

L'Ecuyer and Lemieux (1999)

QMC \cap MCMC

Early references

Chentsov (1967)

Plugs in CUD points.

Samples in finite state space by inversion.

Shows consistency.

Uses very nice coupling argument.

Sobol' (1974)

Has $n \times \infty$ points $x_{ij} \in [0, 1]$

Samples from a row until a return to start state

Gets rate $O(1/n) \dots$ if transition probabilities are $a/2^b$ for integers a, b

Recent QMC \cap MCMC

- | | |
|---------------------------|--|
| Liao (1998) | reorders QMC points |
| Chaudary (2004) | QMC wts on rejected proposals |
| ✓ O & Tribble (2005) | CUD pts in Metropolis, finite state space |
| Tribble & O (2008) | constructions for weakly CUD pts |
| ✓ Tribble (2007) | theory and examples |
| Craiu & Lemieux (2007) | QMC in multiple-try Metropolis |
| Lemieux & Sidorsky (2006) | QMC in exact sampling |
| ✓ Chen, Dick & O (2010) | CUD pts in cts state spaces |

Related ideas

Reordering heat particles	Lécot (1989)
Manually adaptive QMC	Ostland, Yu (1997)
QMC for particle filters	Lemieux, Ormoneit, Fleet (2001), UAI
MCMC \cap antithetics	Frigessi, Gäsemyr, Rue (2000)
MCMC \cap Latin hypercubes	Craiu, Meng (2004)
array-RQMC	L'Ecuyer, Lécot, Tuffin (2004)
Rotor-Router	Propp (2004)
Quasi-random walks on balls	Karaivanova, Chi, Gurov (2007)
array-RQMC	L'Ecuyer, Lécot, L'Archevêque-Gaudet (2008)
Rao-Blackwellized MH	Douc, Robert (2009)

Results from Tribble

Data sets	$n = 2^{10}$		$n = 2^{12}$		$n = 2^{14}$	
	min	max	min	max	min	max
Pumps ($d = 11$)	286	1543	304	5003	1186	16089
Vasorestriction ($d = 42$)	14	15	56	76	108	124

Variance reduction factors from Tribble (2007) for two Gibbs sampling problems. For the pumps data, the greatest and least variance reduction for a randomized CUD sequence versus IID sampling is shown. For the vasorestriction data, greatest and least variance reductions for the three regression parameters are shown. See Tribble (2007) for simulation details.

Targets are posterior means of parameters.

CUD points were LFSR,

with Cranley-Patterson rotations.

Continuous state spaces

Tribble's best results were for a smooth setting: continuous state space and the Gibbs sampler, which has no accept-reject component.

This makes sense: QMC wins its biggest improvements on smooth functions

But the only consistency results $\hat{\mu}_n \rightarrow \mu$ were for discrete state spaces, where only small improvements are seen empirically.

Chen, Dick & O Annals Stat.

Chen, Dick & O extend consistency to continuous state spaces.

MCMC remains consistent when driven by u_1, u_2, \dots , if

- 1) u_i are CUD (or CUD in probability)
- 2) m -step transitions are **Riemann** integrable $\forall m \geq 1$, and
- 3)
 - for Metropolis-Hastings: there is a **coupling** region
(Independence sampler can have one)
 - for Gibbs: there is a **contraction** property
(Gibbs for probit model proven to contract)

Josef Dick's talk will show more

Mini-twisters

Matsumoto & Nishimura sent us some small RNGs based on the same principles as the Mersenne twister.

They come in sizes $M = 2^m - 1$ for $10 \leq m \leq 32$.

$$u_1, u_2, \dots, u_M$$

We explore them for some simulations.

Prepend one or more 0s:

$$0, \dots, 0, u_1, \dots, u_M$$

put into a matrix and apply Cranley-Patterson rotations

Summary

Bivariate Gaussian	apparent better convergence rate for mean
Bivariate Gaussian	not much improvement for discrepancy
Hit and run, volume estimator	no improvement
M/M/1 queue, average wait	mixed results
Garch	some big improvements
Heston stochastic volatility	big improvements for in the money case

Synopsis

The smoother the problem, the more CUD points can improve.

Same as for finite dimensional QMC.

Gaussian Gibbs sampler

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right) \in \mathbb{R}^2$$

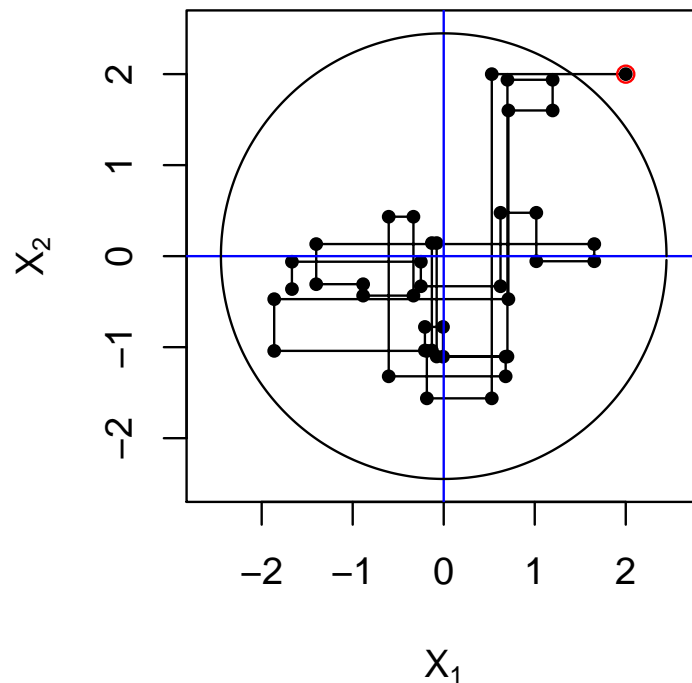
Alternate

$$X_1 \sim \text{DIST}(X_1 \mid X_2 = x_2) = \mathcal{N}(\rho x_2, 1 - \rho^2)$$

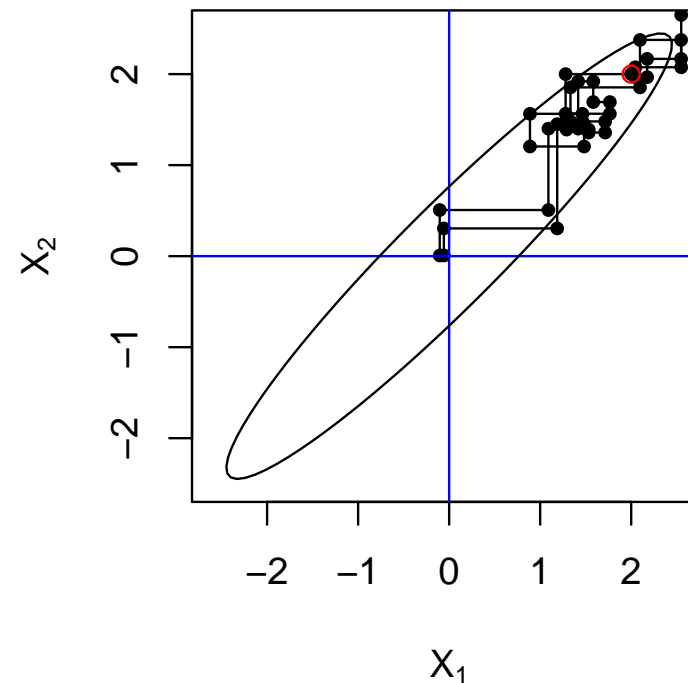
$$X_2 \sim \text{DIST}(X_2 \mid X_1 = x_1) = \mathcal{N}(\rho x_1, 1 - \rho^2)$$

Gaussian Gibbs sampler

Correlation 0
40 steps



Correlation 0.95
40 steps

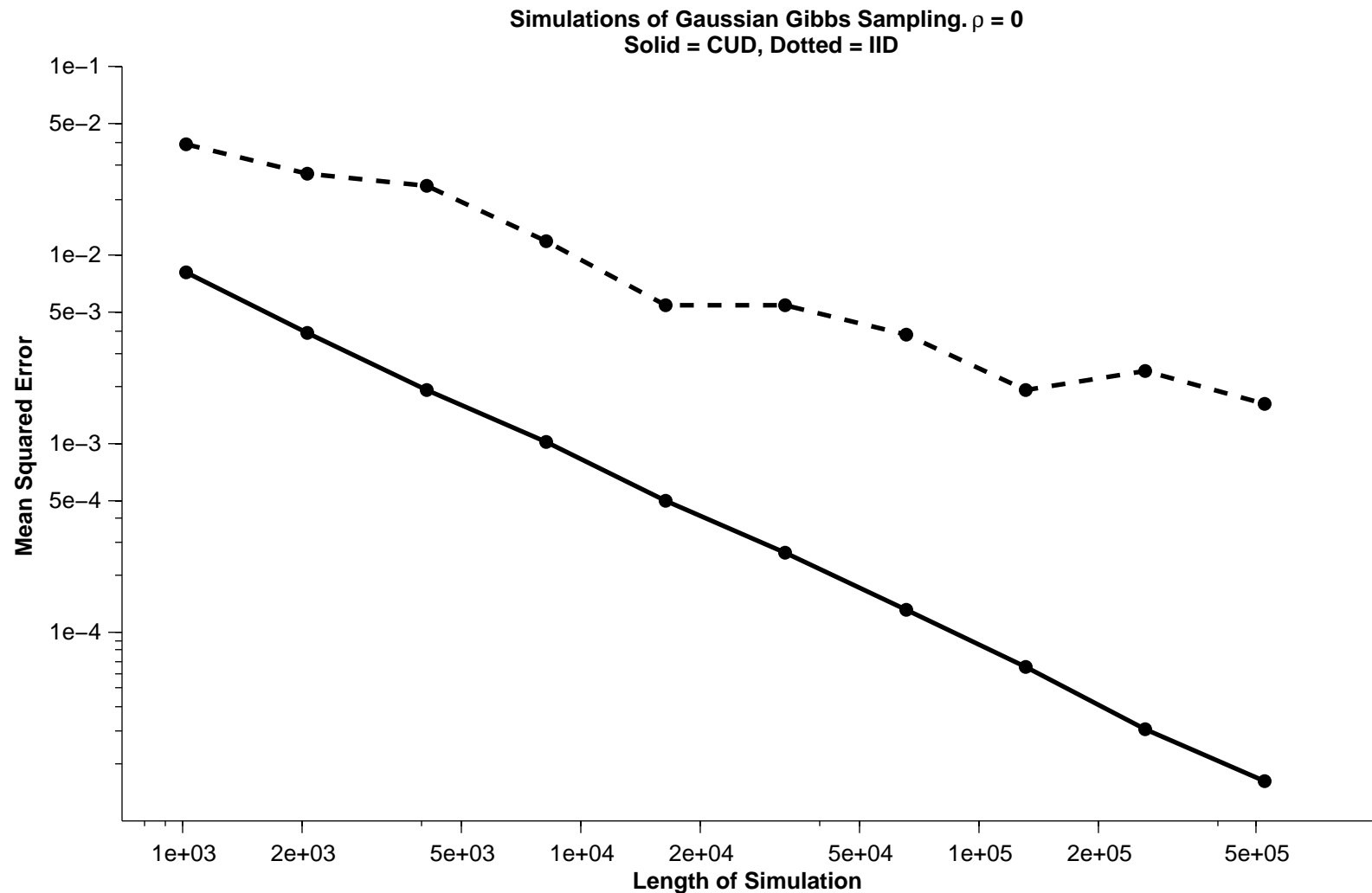


Sampling, $i = 1, \dots, n$

$$X_{i1} \leftarrow \rho X_{i-1,2} + \sqrt{1 - \rho^2} \Phi^{-1}(u_{2i-1})$$

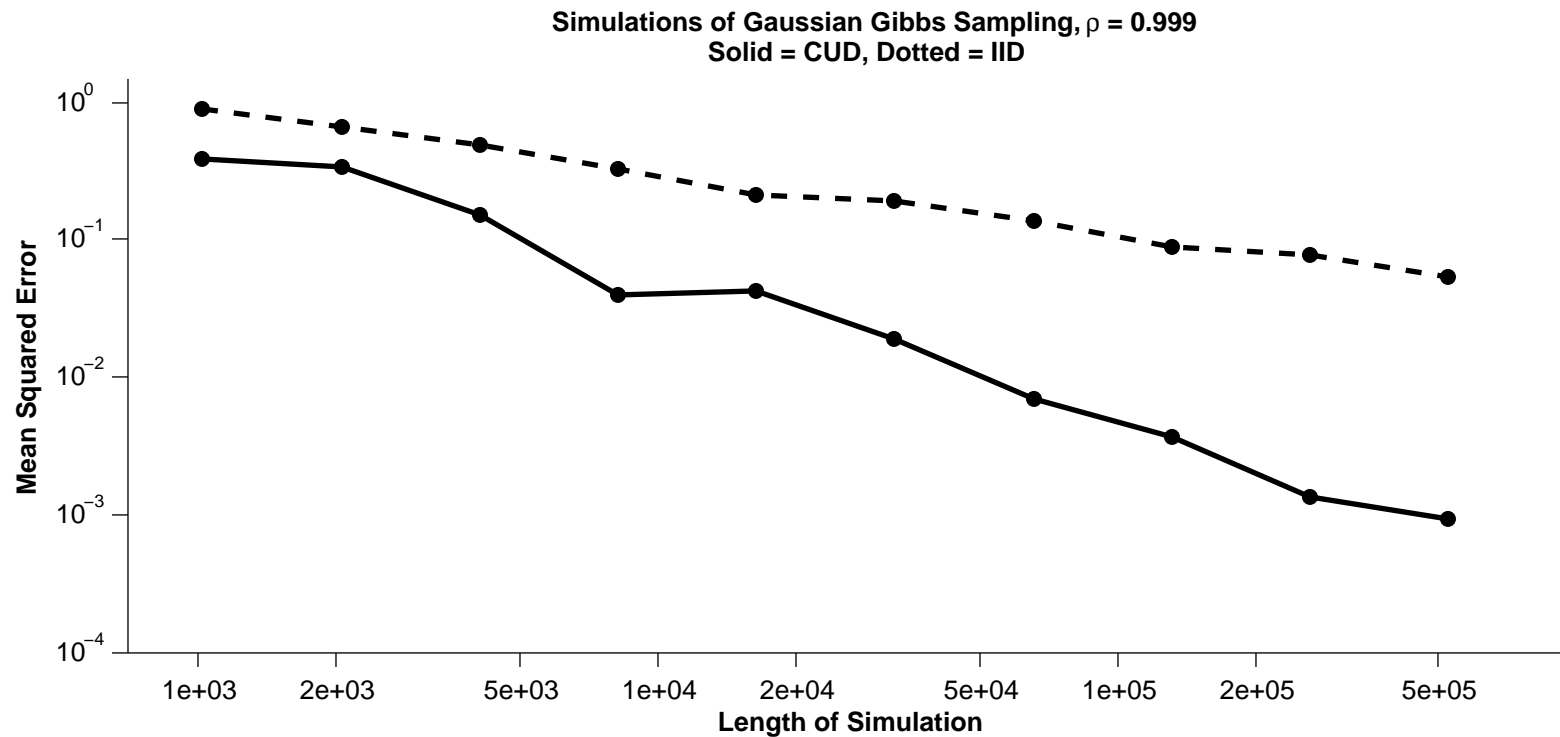
$$X_{i2} \leftarrow \rho X_{i1} + \sqrt{1 - \rho^2} \Phi^{-1}(u_{2i})$$

Gaussian Gibbs $\rho = 0$



Estimate $\mathbb{E}(X)$ start at $(1, 1)$

Gaussian Gibbs $\rho = 0.999$



Estimate $\mathbb{E}(X)$ start at $(1, 1)$

\therefore models like AR(1) are promising

Hit and run MCMC

The hit and run sampler generates points uniformly inside a convex region R .

Given x_i it picks a random direction θ_i and chooses x_{i+1} at random on

$$R \cap \{x_i + r\theta \mid -\infty < r < \infty\}.$$

You can use it to estimate the ratio of two nested convex regions.

The best known way to estimate volume of high dimensional convex regions uses a cascade of nested regions.

Unfortunately

CUD brought no significant improvement for $\text{vol}(\text{triangle})$

M/M/1 queue initial transient

Exponential arrivals at rate $\rho = 0.9$ and service times at rate 1

Customer $i \geq 1$ has **arrival time** A_i , the **service time** S_i , and **waiting time** W_i , where

$$A_0 = 0$$

$$A_i = A_{i-1} - \log(1 - u_{2i-1})/\rho$$

$$S_i = -\log(1 - u_{2i})$$

$$W_1 = 0$$

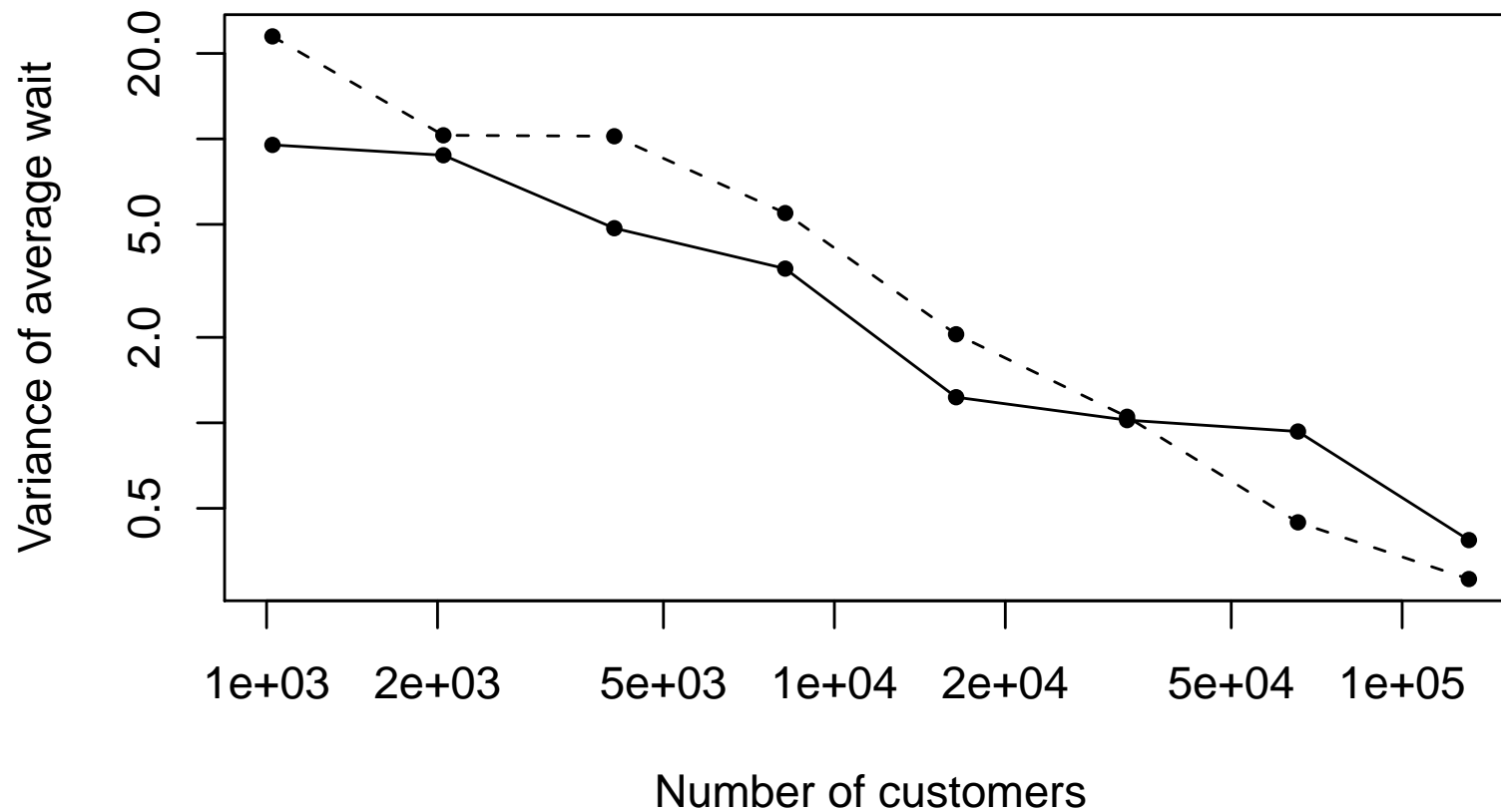
$$W_i = (W_{i-1} + S_{i-1} - A_i)_+$$

Average wait of first n customers is

$$\bar{W}_n = \frac{1}{n} \sum_{i=1}^n W_i \quad \text{we simulate for } \mathbb{E}(\bar{W}_n)$$

Variance of average wait

500 simulations of Lindley's formula
Solid=CUD Dotted=IID



Heston's stochastic volatility

$$dS = rSdt + \sqrt{V}S dW_1, \quad 0 < t < T$$

$$dV = \kappa(\theta - V)dt + \sigma\sqrt{V} dW_2$$

Parameters from J. Zhu (2008):

For S : $S(0) = 100$ $r = 0.04$ $T = 6$

For V : $V(0) = 0.025$ $\theta = 0.04$ $\kappa = 2$ $\sigma = 0.3$

$$\text{Corr}(dW_1, dW_2) \equiv \rho = -0.5$$

Estimate European call: $\mathbb{E}((S(T) - K)_+)$

Heston simulations

Split $[0, T]$ into 2^k intervals $k \in \{8, 10\}$

Use 2^{k+1} CUD numbers per simulation (update price and volatility)

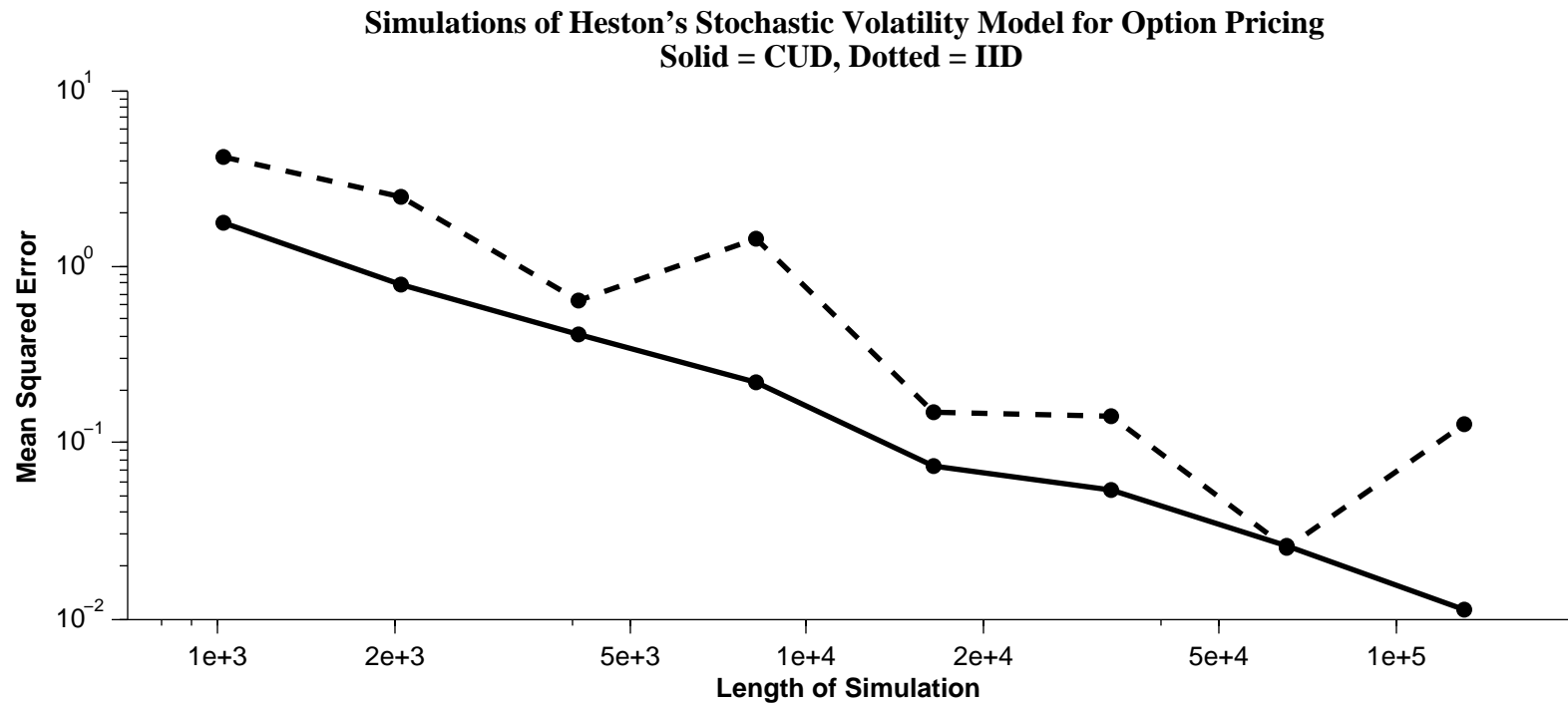
Use 2^{r+k+1} for 2^r simulations $10 \leq r \leq 17$

Update via 'variance form', not 'volatility form'

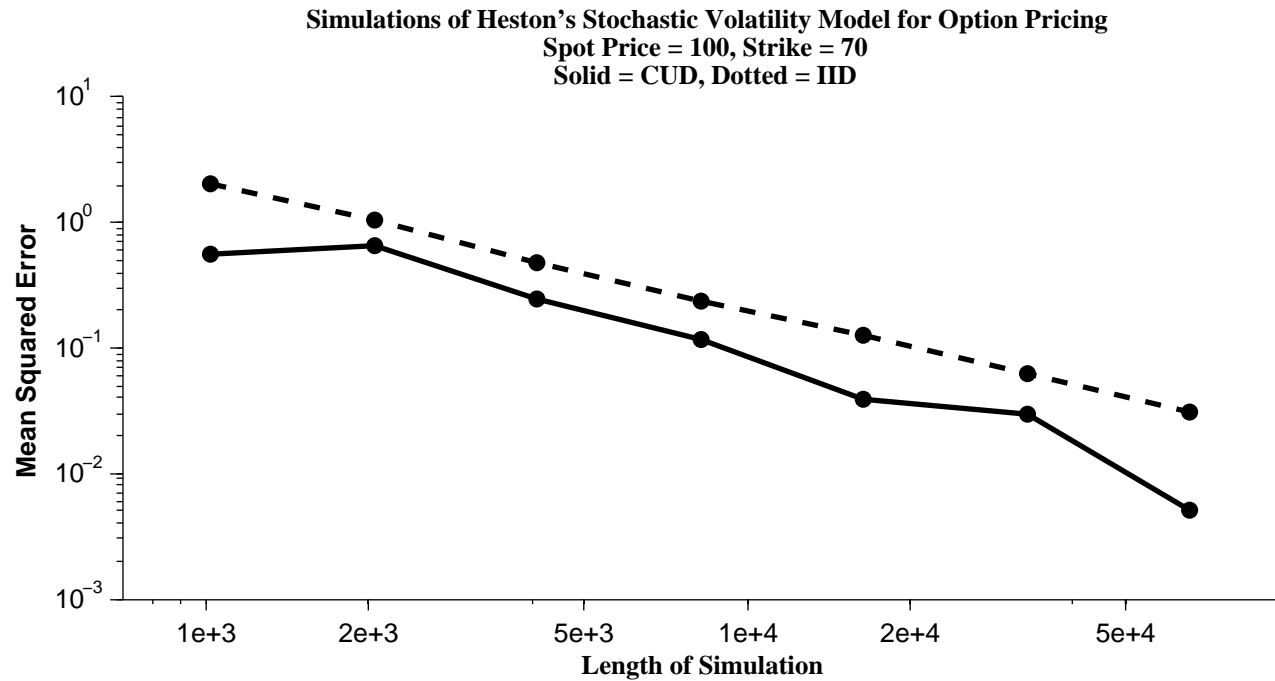
Use 100 rotations (adding $\mathbf{U}(0, 1)^2$)

Compare to exact answer

$$K = 100 = S(0), dt = T/2^8$$



$$K = 70 \quad dt = T/2^8$$



GARCH(1, 1) model

$$\log\left(\frac{X_t}{X_{t-1}}\right) \sim \mathcal{N}\left(r + \lambda\sqrt{h_t} - \frac{1}{2}h_t + \varepsilon_t, h_t\right) \quad 1 \leq t \leq T$$

$$h_t = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \beta_1h_{t-1}$$

Parameters from J.-C. Duan (1995):

For X_t : $r = 0$ $\lambda = 7.452 \times 10^{-3}$ $T = 30$

For h_t : $\alpha_0 = 1.524 \times 10^{-5}$ $\alpha_1 = 0.1883$ $\beta_1 = 0.7162$

h starts at 0.64×0.2413

0.2413 is the stationary variance

European call, strike $K = 1$

Garch simulations

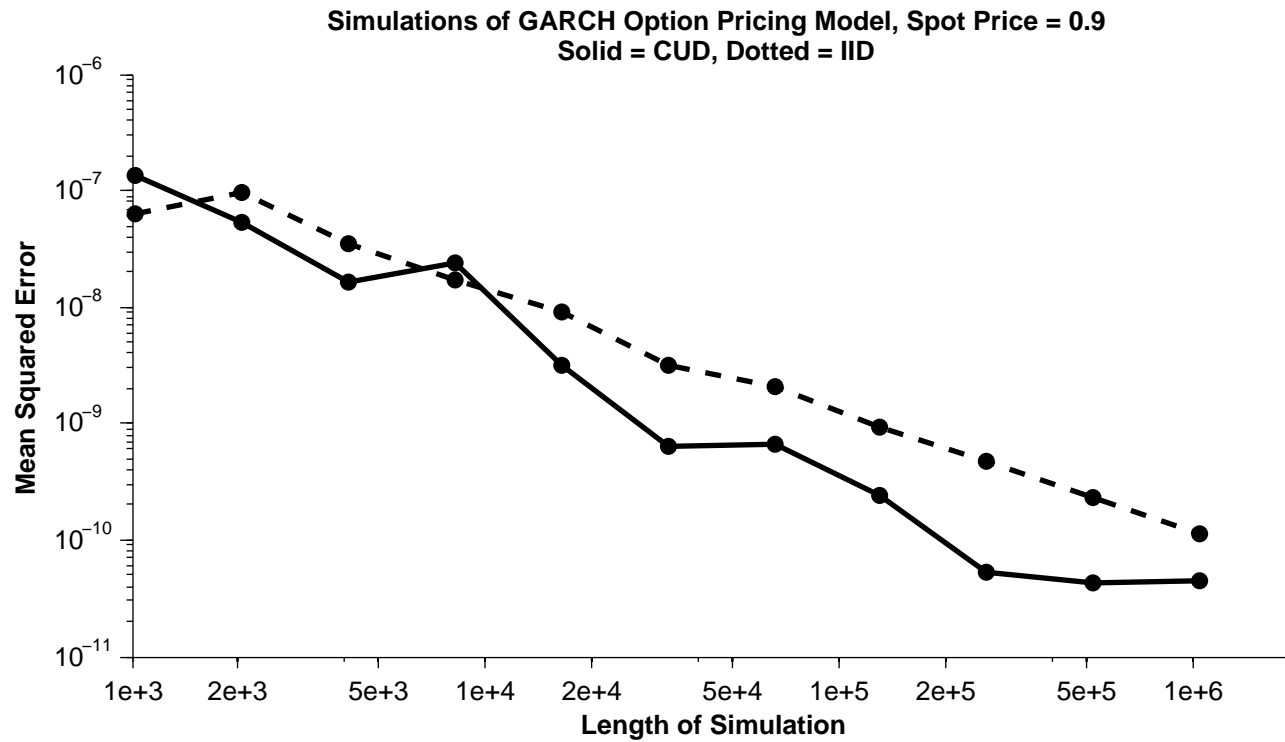
Problem has 30 intervals

Use 30 CUD numbers per simulation (update one price change ε_t)

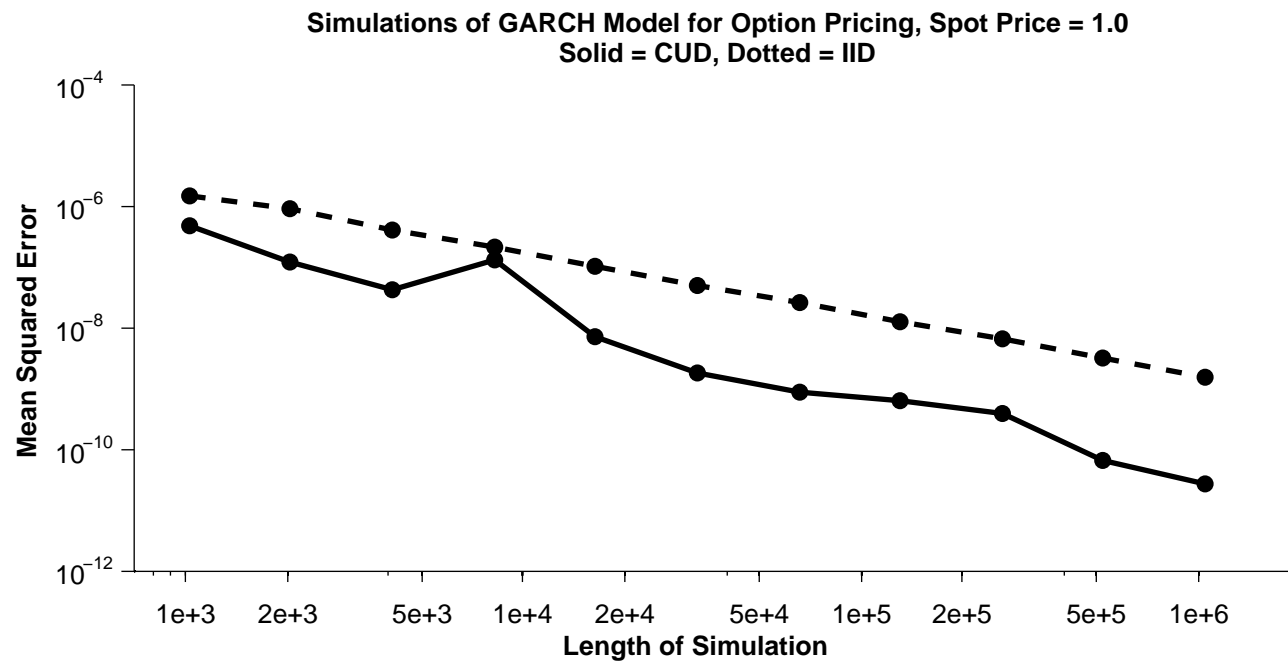
Get $\lfloor 2^m / 30 \rfloor$ prices per CUD sequence

Use 100 rotations (adding $\mathbf{U}(0, 1)^{30}$)

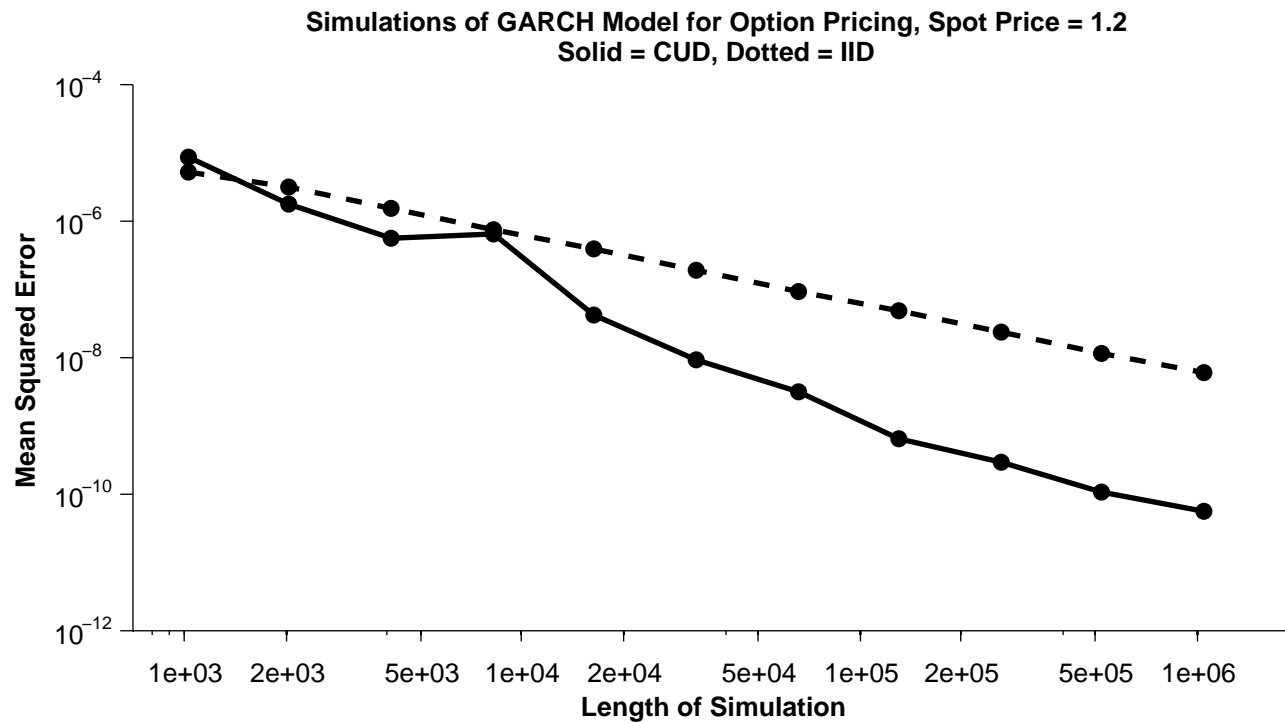
Garch $X_0 = 0.9$



Garch $X_0 = 1.0$



Garch $X_0 = 1.2$



Conclusions

Some QMC works in MCMC

Improvements range from modest to powerful

Just like QMC in MC

We thank

- NSF for funding
- Coworkers: Josef Dick, Makoto Matsumoto, Takuji Nishimura
- Organizers, especially: Henryk Wozniakowski, Leszek Plaskota