Control Variates for Markov Chain Monte Carlo

Dellaportas, P.*, Kontoyiannis, I.†, and Tsourti, Z.*
*Dept of Statistics, AUEB
†Dept of Informatics, AUEB

1st Greek Stochastics Meeting
Monte Carlo: Probability and Methods

Lefkada, August 2009

Outline of presentation

1. MCMC and variance considerations
2. Theoretical framework for Control Variates to MCMC
3. Control variates for Random-Walk Metropolis-Hastings algorithms
4. Application of control variates to RW-MH algorithms
5. Summary and Further Research

MCMC framework

Let

- $\pi(\cdot)$ probability measure on $(\mathbb{X}, \mathbb{B})$
- $F : \mathbb{X} \to \mathbb{R}$ function of interest
- $\pi(F) := E_\pi[F(X)] := \int F d\pi$ the quantity to be estimated

In the MCMC setting, we have

- (X_n) Markov chain on $\mathbb{X}$, with:
  - P transition kernel
  - π stationary probability measure
- $\pi(F)$ is estimated by the ergodic average:

$$\mu_n(F) = \frac{1}{n} \sum_{i=1}^{n} F(X_i)$$

Main Theorems for Markov Chains

- **Ergodic Theorem**, for *Ergodic Markov Chains* and appropriate F, $(\pi(|F|) < \infty)$,

$$\lim_{n\to\infty} \mu_n(F) = \pi(F), \qquad \text{a.s.}$$

- **Central Limit Theorem** (CLT) for *Markov Chains*
Under appropriate additional conditions:

$$\sqrt{n}[\mu_n(F) - \pi(F)] = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [F(X_i) - \pi(F)] \xrightarrow{\mathcal{D}} N(0, \sigma_F^2)$$

where σ_F^2, the asymptotic variance of F, is:

$$\sigma_F^2 := \lim_{n\to\infty} Var_\pi[\sqrt{n}\mu_n(F)] = \sum_{n=-\infty}^{\infty} Cov_\pi[F(X_0), F(X_n)]$$

Approaches to variance reduction

- Importance sampling
- Antithetic sampling/variates
- **Control variates**
- Rao-Blackwellization

Introduction of control variates to Markov chains

Let

- $\pi(\cdot)$ target probability measure on $(\mathbb{X}, \mathbb{B})$
- $F : \mathbb{X} \to \mathbb{R}$ function of interest
- $\pi(F) := E_\pi[F(X)] := \int F d\pi$ the quantity to be estimated
- (X_n) Markov chain on $\mathbb{X}$, with:
  - P transition kernel
  - π stationary probability measure

Introduction of control variates to Markov chains

Let

- $\pi(\cdot)$ target probability measure on $(\mathbb{X}, \mathbb{B})$
- $F : \mathbb{X} \to \mathbb{R}$ function of interest
- $\pi(F) := E_\pi[F(X)] := \int F d\pi$ the quantity to be estimated
- (X_n) Markov chain on $\mathbb{X}$, with:
  - P transition kernel
  - π stationary probability measure

further:

- function $U : \mathbb{X} \to \mathbb{R}$, with $\pi(U) = 0$
- modified function $F_\theta = F - \theta U$
- modified estimator $\mu_n(F_\theta) = \mu_n(F) - \theta \mu_n(U)$

Introduction of control variates to Markov chains

All the "regularity" properties of F also hold for F_θ:

$$\lim_{n\to\infty}\mu_n(F_\theta) := \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}F_\theta(X_i) = \pi(F_\theta) = \pi(F), \qquad \text{a.s.}$$

and

$$\sqrt{n}[\mu_n(F_\theta) - \pi(F)] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[F_\theta(X_i) - \pi(F)] \xrightarrow{\mathcal{D}} N(0, \sigma^2_{F_\theta})$$

Introduction of control variates to Markov chains

All the "regularity" properties of F also hold for F_θ:

$$\lim_{n\to\infty} \mu_n(F_\theta) := \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} F_\theta(X_i) = \pi(F_\theta) = \pi(F), \qquad \text{a.s.}$$

and

$$\sqrt{n}[\mu_n(F_\theta) - \pi(F)] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[F_\theta(X_i) - \pi(F)] \xrightarrow{\mathcal{D}} N(0, \sigma^2_{F_\theta})$$

Research interest:

Introduction of control variates to Markov chains

All the "regularity" properties of F also hold for F_θ:

$$\lim_{n\to\infty} \mu_n(F_\theta) := \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} F_\theta(X_i) = \pi(F_\theta) = \pi(F), \qquad \text{a.s.}$$

and

$$\sqrt{n}[\mu_n(F_\theta) - \pi(F)] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[F_\theta(X_i) - \pi(F)] \xrightarrow{\mathcal{D}} N(0, \sigma^2_{F_\theta})$$

Research interest: Find appropriate

- function U
- parameter θ

so as to significantly reduce $(\sigma^2_{F_\theta})$ compared to (σ^2_F).

Choice of function U

In this setting we use:

$$U = G - PG, \qquad \text{for arbitrary } G, \text{ with } \pi(|G|) < \infty$$

Choice of function U

In this setting we use:

$$U = G - PG, \qquad \text{for arbitrary } G, \text{ with } \pi(|G|) < \infty$$

$$\boxed{PG(x) = E\left[G(X_1) | X_0 = x\right]}$$

Choice of function U

In this setting we use:

$$U = G - PG, \qquad \text{for arbitrary } G, \text{ with } \pi(|G|) < \infty$$

$$\boxed{PG(x) = E\left[G(X_1)|X_0 = x\right]}$$

Then,

Which U? $\Rightarrow$ Which G?

Elaboration on variance of the modified estimator ($\sigma^2_{F_\theta}$)

An alternative expression of the variance in the CLT is:

$$\sigma_F^2 = \pi\left(\hat{F}^2 - (P\hat{F})^2\right)$$

$\hat{F}$: the solution of Poisson's equation:

$$P\hat{F} - \hat{F} = -F + \pi(F)$$

Analogously

$$\sigma_\theta^2 := \sigma_{F_\theta}^2 = \pi\left(\hat{F}_\theta^2 - (P\hat{F}_\theta)^2\right)$$

Choice of function U

In this setting we use:

$$U = G - PG, \qquad \text{for arbitrary } G, \text{ with } \pi(|G|) < \infty$$

$$PG(x) = E\left[G(X_1)|X_0 = x\right]$$

Then,

$$\text{Which } U? \Rightarrow \text{ Which } G?$$

Findings from σ_θ^2 elaboration

Choice of function U

In this setting we use:

$$U = G - PG, \qquad \text{for arbitrary } G, \text{ with } \pi(|G|) < \infty$$

$$PG(x) = E\left[G(X_1)|X_0 = x\right]$$

Then,

$$\text{Which } U? \Rightarrow \text{ Which } G?$$

Findings from σ_θ^2 elaboration

- If $G = \hat{F} \Rightarrow (\sigma^2_{F_\theta}) = 0$

Choice of function U

In this setting we use:

$$U = G - PG, \qquad \text{for arbitrary } G, \text{ with } \pi(|G|) < \infty$$

$$PG(x) = E\left[G(X_1)|X_0 = x\right]$$

Then,

$$\text{Which } U? \Rightarrow \text{ Which } G?$$

Findings from σ_θ^2 elaboration

- If $G = \hat{F} \Rightarrow (\sigma^2_{F_\theta}) = 0$

Guidelines for choosing G

- as close as possible to $\hat{F}$

Choice of function U

In this setting we use:

$$U = G - PG, \qquad \text{for arbitrary } G, \text{ with } \pi(|G|) < \infty$$

$$PG(x) = E\left[G(X_1)|X_0 = x\right]$$

Then,

$$\text{Which } U? \Rightarrow \text{ Which } G?$$

Findings from σ_θ^2 elaboration

- If $G = \hat{F} \Rightarrow (\sigma_{F_\theta}^2) = 0$
- The higher the correlation between F, U the smaller the $(\sigma_{F_\theta}^2)$

Guidelines for choosing G

- as close as possible to $\hat{F}$

Choice of function U

In this setting we use:

$$U = G - PG, \qquad \text{for arbitrary } G, \text{ with } \pi(|G|) < \infty$$

$$PG(x) = E\big[G(X_1)|X_0 = x\big]$$

Then,

$$\text{Which } U? \Rightarrow \text{ Which } G?$$

Findings from σ_θ^2 elaboration

- If $G = \hat{F} \Rightarrow (\sigma^2_{F_\theta}) = 0$
- The higher the correlation between F, U the smaller the $(\sigma^2_{F_\theta})$

Guidelines for choosing G

- as close as possible to $\hat{F}$
- leading to $U = G - PG$ highly correlated to F

Optimal value of parameter θ

It can be derived that:

$$\sigma_\theta^2 = \ldots$$

quadratic function with respect to θ.

Optimal value of parameter θ

It can be derived that:

$$\sigma_\theta^2 = \dots$$

quadratic function with respect to θ.

Thus,

$$\theta^* = \frac{\pi\left(\hat{F}G - (P\hat{F})(PG)\right)}{\pi\left(G^2 - (PG)^2\right)}$$

Alternative expression (proved by Dellaportas and Kontoyiannis, 2008) with practical usefulness:

$$\theta^* = \frac{\pi\left(\hat{F}G - (P\hat{F})(PG)\right)}{E_\pi\left[\left(G(X_1) - PG(X_0)\right)^2\right]}$$

Optimal empirical estimate of θ^* for reversible chains

If the chain (X_n) is reversible, Dellaportas and Kontoyiannis (2008) prove:

Theorem

$$\pi\left(\hat{F}G - (P\hat{F})(PG)\right) = \pi\left((F - \pi(F))(G + PG)\right)$$

Optimal empirical estimate of θ^* for reversible chains

If the chain (X_n) is reversible, Dellaportas and Kontoyiannis (2008) prove:

Theorem

$$\pi\left(\hat{F}G - (P\hat{F})(PG)\right) = \pi\left((F - \pi(F))(G + PG)\right)$$

So, the **optimal value** of θ (for reversible chains) can be expressed as

$$\theta^*_{rev} = \frac{\pi\left((F - \pi(F))(G + PG)\right)}{E_\pi\left[(G(X_1) - PG(X_0))^2\right]}$$

Optimal empirical estimate of θ^* for reversible chains

If the chain (X_n) is reversible, Dellaportas and Kontoyiannis (2008) prove:

Theorem

$$\pi\left(\hat{F}G - (P\hat{F})(PG)\right) = \pi\left((F - \pi(F))(G + PG)\right)$$

So, the **optimal value** of θ (for reversible chains) can be expressed as

$$\theta^*_{rev} = \frac{\pi\left((F - \pi(F))(G + PG)\right)}{E_\pi\left[(G(X_1) - PG(X_0))^2\right]}$$

In this case, θ can be adaptively estimated as

$$\hat{\theta}_n = \frac{\mu_n(F(G + PG)) - \mu_n(F)\mu_n(G + PG)}{\frac{1}{n}\sum_{i=1}^n [G(X_i) - PG(X_{i-1})]^2}$$

Extension to multiple control variates

Let's further assume:

- k functions $U_j (= G_j - PG_j) : \mathbb{X} \to \mathbb{R}$, with $\pi(U_j) = 0$
- Notation with vectors:

$$G = \begin{pmatrix} G_1 \\ G_2 \\ \dots \\ G_k \end{pmatrix} : \mathbb{X} \to \mathbb{R}^k \qquad U = \begin{pmatrix} U_1 \\ U_2 \\ \dots \\ U_k \end{pmatrix} : \mathbb{X} \to \mathbb{R}^k \qquad \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \dots \\ \theta_k \end{pmatrix} \in \mathbb{R}^k$$

- modified function
  $F_\theta = F - \langle \theta, U \rangle = F - \sum_{j=1}^k \theta_j U_j$

Extension to multiple control variates

Analogously to the one-dimensional case, we have that:

$$\sigma^2_{F_\theta} = \sigma^2_F - 2\pi\big(\hat{F}\langle\theta, G\rangle - P\hat{F}\langle\theta, PG\rangle\big) + \pi\big(\langle\theta, G\rangle^2 - \langle\theta, PG\rangle^2\big)$$

$$\Downarrow$$

$$\theta^* = K(G)^{-1}\pi\left(\hat{F}G - (P\hat{F})(PG)\right)$$

where

$K(G)_{ij} = E_\pi\big[\big(G_i(X_1) - PG_i(X_0)\big)\big(G_j(X_1) - PG_i(X_0)\big)\big]$

For **reversible chains**, Dellaportas and Kontoyiannis (2008) prove:

Theorem

$$\theta^* = K(G)^{-1}\pi\big(\big(F - \pi(F)\big)\big(G + PG\big)\big)$$

Use of control variates for the reduction of variance of MCMC algorithms

Control Variate methodology directly applicable to reversible MCMC algorithms

Derivation of $\mu_n(F_\theta)$ straightforward. Quantities needed:

- $F(X_i)$, $G_j(X_i)$ for $i = 1, \ldots, n$, $j = 1, \ldots, k$: OK, readily available
- BUT $PG_j(X_i)$?
  This issue may require further attention...

Most popular MCMC algorithms:

- Gibbs sampler
  Treated in detail in Dellaportas and Kontoyiannis (2008)
- Metropolis-Hastings algorithm
  Random Walk (RW-MH)

Use of control variates for the reduction of variance of MCMC algorithms

Control Variate methodology directly applicable to reversible MCMC algorithms

Derivation of $\mu_n(F_\theta)$ straightforward. Quantities needed:

- $F(X_i)$, $G_j(X_i)$ for $i = 1, \ldots, n$, $j = 1, \ldots, k$: OK, readily available
- BUT $PG_j(X_i)$?
  This issue may require further attention...

Most popular MCMC algorithms:

- Gibbs sampler
  - Treated in detail in Dellaportas and Kontoyiannis (2008)
- Metropolis-Hastings algorithm
  - Random Walk (RW-MH)

Setting of RW-MH

Assume the following:

- Our target distribution is a d-dimensional probability density $\pi(x)$
- The "Random Walk Metropolis-Hastings" (RW-MH) algorithm used for the simulation of π can be described as:
  - Assume initial value X_0 and $Y_0 = X_0 + \Delta_0$, where $\Delta_0 \sim P_\Delta$ (d-variate symmetrical distribution)
  - At step $t+1$, given $X_t = x_t$, $Y_t = y_t = x_t + \Delta_t$, we have that:
    - $X_{t+1} = \begin{cases} x_t + \Delta_t = y_t & \text{w.pr. } \rho(x_t, y_t) \\ x_t & \text{w.pr. } 1 - \rho(x_t, y_t) \end{cases}$

      where

      $$\rho(x_t, y_t) = \min\left\{1, \frac{\pi(y_t)}{\pi(x_t)}\right\}$$
    - we simulate $\Delta_{t+1} \sim P_\Delta$

Use of control variates for RW-MH - $PG(x)$ elaboration

$$PG(x) = E_{X_1}\left[G\left(X_1\right) \mid X_0 = x\right]$$

Use of control variates for RW-MH - $PG(x)$ elaboration

$$
\begin{aligned}
PG(x) &= E_{X_1}\left[G\left(X_1\right) \mid X_0 = x\right] \\
&= E_{Y_0}\left\{E_{X_1}\left[G\left(X_1\right) \mid X_0 = x, Y_0 = y\right]\right\}
\end{aligned}
$$

Use of control variates for RW-MH - $PG(x)$ elaboration

$$
\begin{aligned}
PG(x) &= E_{X_1}\left[G\left(X_1\right) \mid X_0 = x\right] \\
&= E_{Y_0}\left\{E_{X_1}\left[G\left(X_1\right) \mid X_0 = x,\, Y_0 = y\right]\right\} \\
&= E_{Y_0}\left\{\rho\left(x, y\right) \cdot G\left(y\right) + \left(1 - \rho\left(x, y\right)\right) G\left(x\right)\right\}
\end{aligned}
$$

Use of control variates for RW-MH - $PG(x)$ elaboration

$$
\begin{aligned}
PG(x) &= E_{X_1}\left[G\left(X_1\right) \mid X_0 = x\right] \\
&= E_{Y_0}\left\{E_{X_1}\left[G\left(X_1\right) \mid X_0 = x, Y_0 = y\right]\right\} \\
&= E_{Y_0}\left\{\rho\left(x,y\right) \cdot G\left(y\right) + \left(1 - \rho\left(x,y\right)\right) G\left(x\right)\right\} \\
&= G(x) + E_{Y_0}\left\{\rho(x,y) \cdot (G(y) - G(x))\right\}
\end{aligned}
$$

Use of control variates for RW-MH - $PG(x)$ elaboration

$$
\begin{aligned}
PG(x) &= E_{X_1}\left[G\left(X_1\right) \mid X_0 = x\right] \\
&= E_{Y_0}\left\{E_{X_1}\left[G\left(X_1\right) \mid X_0 = x, Y_0 = y\right]\right\} \\
&= E_{Y_0}\left\{\rho\left(x,y\right) \cdot G\left(y\right) + \left(1-\rho\left(x,y\right)\right) G\left(x\right)\right\} \\
&= G(x) + E_{Y_0}\left\{\rho(x,y) \cdot \left(G(y) - G(x)\right)\right\}
\end{aligned}
$$

i.e.

$$
PG(x) = G(x) + E_{Y_0}\left\{\rho(x,y) \cdot \left(G(y) - G(x)\right)\right\}, \text{ where } Y_0 \sim q(y \mid x)
$$

Use of control variates for RW-MH - $PG(x)$ elaboration

$$
\begin{aligned}
PG(x) &= E_{X_1}\left[G\left(X_1\right) \mid X_0 = x\right] \\
&= E_{Y_0}\left\{E_{X_1}\left[G\left(X_1\right) \mid X_0 = x, Y_0 = y\right]\right\} \\
&= E_{Y_0}\left\{\rho\left(x,y\right)\cdot G\left(y\right) + \left(1-\rho\left(x,y\right)\right)G\left(x\right)\right\} \\
&= G(x) + E_{Y_0}\left\{\rho(x,y)\cdot\left(G(y)-G(x)\right)\right\}
\end{aligned}
$$

i.e.

$$
PG(x) = G(x) + E_{Y_0}\left\{\rho(x,y)\cdot\left(G(y)-G(x)\right)\right\}, \text{ where } Y_0 \sim q(y \mid x)
$$

Several approaches may be considered for the estimation of $PG(x)$.

- Monte Carlo estimation based on P_{Δ}
- Importance sampling based on the proposed values y

Application of control variates to RW-MH algorithms - Univariate Normal

- Target : $N\left(0, \sigma_t^2 = 10\right)$
- Proposal : $N\left(0, \sigma_{pr}^2 = 160\right)$
- We use $F(x) = x$ and $G(x) = x$

Application of control variates to RW-MH algorithms - Univariate Normal

- Target : $N\left(0, \sigma_t^2 = 10\right)$
- Proposal : $N\left(0, \sigma_{pr}^2 = 160\right)$
- We use $F(x) = x$ and $G(x) = x$
- Framework for the evaluation of variance reduction:
  - Optimal value of θ based on Dellaportas and Kontoyiannis (2008) approach for reversible chains
  - Terms of $PG(x)$ assessed using Monte Carlo estimates from the proposal distribution
  - $T = 100$ repetitions for each simulation scenario.
  - "Variance Reduction Factors":

$$\text{VRF} = \frac{S^2_{\mu_n(F)}}{S^2_{\mu_n(F_\theta)}}$$

Application of control variates to RW-MH algorithms - Univariate Normal

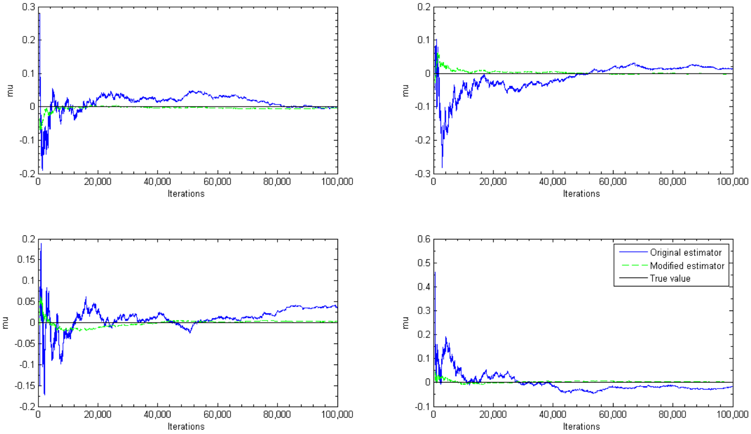

Figure: Ergodic means for different realizations with $n = 100,000$ and $n_{PG} = 50$

Application of control variates to RW-MH algorithms - Univariate Normal

Table: VRF's in Univariate Normal case - Normal proposal - Simplest $G(x) = x$

	Length of Markov chain (number of iterations n)						
n_{PG}	**1,000**	**2,000**	**5,000**	**10,000**	**20,000**	**50,000**	**100,000**
5	7.25	7.06	5.61	9.12	8.96	7.64	8.60
10	11.56	9.54	8.46	14.35	10.79	15.33	12.26
20	13.05	22.19	15.94	30.20	20.64	19.12	26.02
50	23.31	40.45	44.01	35.85	45.65	39.51	34.95
100	36.06	34.68	43.31	60.45	49.34	40.57	54.33
200	41.55	56.49	32.98	42.09	57.14	78.11	61.99
500	36.52	62.54	70.50	50.65	55.28	55.97	50.40
1,000	32.30	41.40	66.05	68.69	50.00	64.52	96.26

Application of control variates to RW-MH algorithms - Univariate Normal

Study of the effect of G functions

1. $G(x) = x$
2. $G_1(x) = x,\ G_2(x) = x^2$

   $\cdots$

k. $G_1(x) = x,\ G_2(x) = x^2,\ \cdots,\ G_k(x) = x^k$

   $\cdots$

Application of control variates to RW-MH algorithms - Univariate Normal

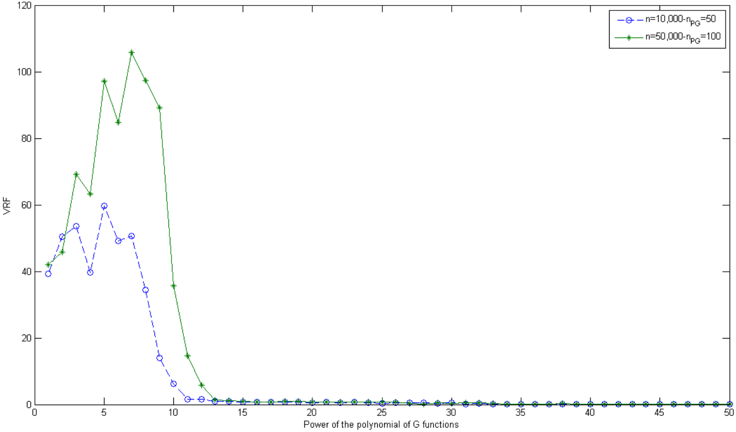

Figure: Plot of VRF by the order of polynomial G functions

Application of control variates to RW-MH algorithms - Univariate Normal

Table: VRF's in Univariate Normal case - Normal proposal - Polynomial G functions of order 4, i.e. $G_1(x) = x,\ G_2(x) = x^2,\ G_3(x) = x^3,\ G_4(x) = x^4$

	Length of Markov chain (number of iterations n)						
n_{PG}	**1,000**	**2,000**	**5,000**	**10,000**	**20,000**	**50,000**	**100,000**
5	6.47	5.94	5.72	8.76	9.34	8.46	8.19
10	10.40	9.33	9.74	14.27	12.18	14.82	11.71
20	9.89	21.51	16.15	38.03	27.55	22.25	30.87
50	19.51	37.58	64.26	29.73	51.58	52.28	49.82
100	25.83	20.75	59.09	68.35	66.50	60.89	71.72
200	15.15	42.49	37.94	69.41	80.34	152.59	106.34
500	17.82	31.58	78.46	78.33	82.35	100.31	112.40
1,000	16.68	23.61	92.50	114.56	109.26	127.89	180.37

Application of control variates to RW-MH algorithms - Univariate Normal

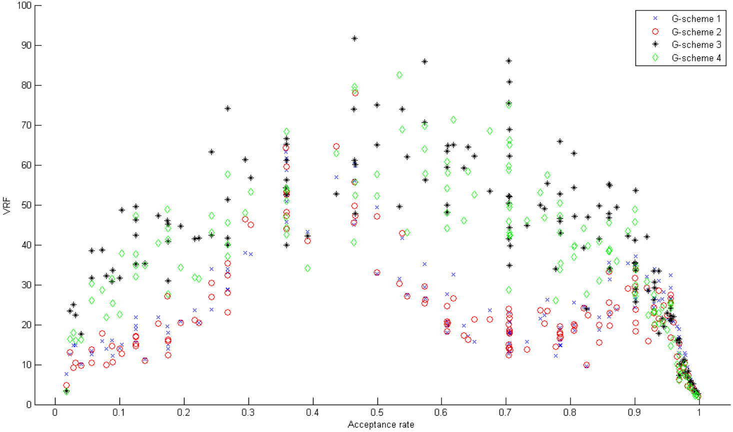

Figure: Plot of VRF by different acceptance rates

Application of control variates to RW-MH algorithms - Univariate Normal

Study of the effect of the proposal distribution

Table: VRF's in Univariate Normal case - T-student proposal (3 df's, variance 160) - Simplest function $G(x) = x$

	Length of Markov chain (number of iterations n)					
n_{PG}	**1,000**	**2,000**	**5,000**	**10,000**	**20,000**	**50,000**
5	6.73	7.16	10.25	8.59	8.64	5.69
10	10.69	8.70	12.45	13.68	10.39	15.76
20	17.68	28.25	21.26	20.21	19.98	27.64
50	22.64	36.28	37.14	44.02	48.17	52.30
100	28.88	48.08	30.06	42.89	54.28	53.40
200	51.35	50.93	52.94	54.31	84.32	46.03
500	55.65	63.28	60.25	86.49	68.40	57.05

Application of control variates to RW-MH algorithms - Univariate Normal

Study of the effect of the proposal distribution

Table: VRF's in Univariate Normal case - Uniform proposal $(-5.5 \cdot \sigma_t, +5.5 \cdot \sigma_t)$ - Simplest function $G(x) = x$

	Length of Markov chain (number of iterations n)					
n_{PG}	**1,000**	**2,000**	**5,000**	**10,000**	**20,000**	**50,000**
5	5.52	5.81	8.21	5.38	5.90	7.25
10	7.94	8.77	11.11	10.71	8.59	8.72
20	11.08	18.53	19.59	15.27	20.23	17.35
50	27.22	26.33	18.14	17.94	17.35	21.84
100	19.74	26.00	19.72	15.23	18.58	21.51
200	18.14	22.50	23.82	26.99	30.12	22.76
500	18.84	18.61	18.36	26.25	25.85	20.14

Application of control variates to RW-MH algorithms - Survival analysis

Source: Albert, J. (2007). Bayesian Computation with R, Springer.

Data:
Lifetime of a number of patients some of which had a heart transplant.

Model assumptions:

- Non-transplant patients: $X_i \sim$ Exponential with mean $1/\eta$
- Transplant patients: $X_i \sim$ Exponential with mean $1/(\tau\eta)$
- Parameter $\eta \sim Gamma(p, \lambda)$, i.e. $f(\eta) = \frac{\lambda^p}{\Gamma(p)}\eta^{p-1}exp(-\lambda\eta)$
- Unknown parameter vector: (τ, λ, p) (all positive)

Notation:

N non-transplant patients:

- n: died
- $N - n$: censored
- x_i survival time

.

M transplant patients:

- m: died
- $M - m$: censored
- y_i time to transplant
- z_i survival time

Application of control variates to RW-MH algorithms - Survival analysis

Likelihood:

$$L(\tau,\lambda,p) = \prod_{i=1}^{n} \frac{p\lambda^p}{(\lambda+x_i)^{p+1}} \prod_{i=n+1}^{N} \left(\frac{\lambda}{\lambda+x_i}\right)^p \prod_{j=1}^{m} \frac{\tau p\lambda^p}{(\lambda+y_j+\tau z_j)^{p+1}} \prod_{j=m+1}^{M} \left(\frac{\lambda}{\lambda+y_j+\tau z_j}\right)^p$$

Prior distribution of parameters: $g(\tau,\lambda,p) \propto 1$

Posterior distribution of parameters: $g(\tau,\lambda,p|\text{data}) \propto L(\tau,\lambda,p)$

Transformation

$$\phi = (\phi_1 := log\tau, \phi_2 := log\lambda, \phi_3 := logp,)$$

$$g(\phi|\text{data}) \propto L\left(e^{\phi_1}, e^{\phi_2}, e^{\phi_3}\right) \cdot e^{\sum_{i=1}^{3}\phi_i}$$

Application of control variates to RW-MH algorithms - Survival analysis

Table: Function $F(\tau, \lambda, p) = log(\tau) = \phi_1$, $n = 10,000$, $n_{PG} = 50$

Form of G	A. $\begin{bmatrix} G_1 = \phi_1 \\ G_2 = \phi_2 \\ G_3 = \phi_3 \end{bmatrix}$	B. $\begin{bmatrix} G_1 = \phi_1^2 \\ G_2 = \phi_2^2 \\ G_3 = \phi_3^2 \end{bmatrix}$	C. $\begin{bmatrix} G_1 = \phi_1 \\ G_2 = \phi_1^2 \end{bmatrix}$	D. $\begin{bmatrix} G_1 = \phi_1 \\ G_2 = e^{\phi_1} \end{bmatrix}$
VRF	29.9	1.4	41.6	30.7
Form of G	E. $\begin{bmatrix} G_1 = \phi_1 \\ G_2 = \phi_1^2 \\ G_3 = \phi_2 \\ G_4 = \phi_2^2 \\ G_5 = \phi_3 \\ G_6 = \phi_3^2 \end{bmatrix}$	F. $\begin{bmatrix} G_1 = \phi_1 \\ G_2 = \phi_1^2 \\ G_3 = \phi_2 \\ G_4 = \phi_2^2 \end{bmatrix}$	G. $\begin{bmatrix} G_1 = \phi_1 \\ G_2 = \phi_1^2 \\ G_3 = \phi_3 \\ G_4 = \phi_3^2 \end{bmatrix}$	
VRF	42.3	37.4	35.4	

Application of control variates to RW-MH algorithms - Survival analysis

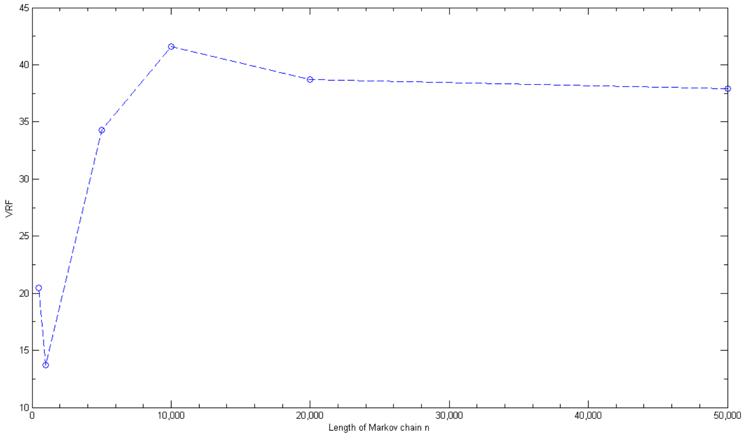

Figure: Plot of VRF by n, function $F(\tau, \lambda, p) = log(\tau) = \phi_1$,

Application of control variates to RW-MH algorithms - Poisson generation

Setting:

- Target distribution π: Poisson(λ),
- Proposal: discrete bell-shaped:

$$P_{\Delta}^{M}(\Delta = \delta) = \frac{M + 1 - |\delta|}{M(M+1)}, \quad {\scriptstyle \delta \in \{-M, -M+1, \ldots, -1, +1, \ldots, M-1, M\}}$$

- Inference is focused on:

$$F(\lambda) = \sqrt{\lambda}$$

- To enhance estimation of $\pi(F)$:

$$F^{G,\theta}(\lambda) = F(\lambda) - \theta \cdot U(\lambda)$$

where $U = G - PG$

Application of control variates to RW-MH algorithms - Poisson generation

- The form of function G used here is

$$G(\lambda) = \lambda$$

- In the present setting:

$$\begin{aligned} PG(x) &= E_\pi\left[G(\lambda_{t+1})|\lambda_t = x\right] \\ &= \dots \\ &= x + \sum_{j=-k,\neq 0}^{k}\left[P_\Delta^k(j)\cdot j\cdot \rho(x, x+j)\right] \end{aligned}$$

- The terms of $PG(x)$ are assessed using two approaches:
  - (i) Using the exact formula
  - (ii) Using Monte Carlo estimates from P_Δ^k distribution

Application of control variates to RW-MH algorithms - Poisson generation

Table: VRF's for $F(x) = \sqrt{x}$, $G(x) = x$ for $n = 5,000$ (analytic formula for $PG(x)$)

	Proposal distribution P^M_Δ						
Target $P(\lambda)$	1	10	15	20	30	40	70
$\lambda = 5$	34.7	42.2	43.3	27.1	-	-	-
$\lambda = 10$	50.8	-	89.6	83.1	64.9	-	-
$\lambda = 100$	10.2	23.9	-	-	-	40.0	174.1

Application of control variates to RW-MH algorithms - Heavy tailed distributions

Source: Jarner, S.F. and Roberts, G.O. (2007). Convergence of Heavy-tailed Monte Carlo Markov Chain Algorithms. *Scandinavian Journal of Statistics*. **34**, 781-815.

Main result for RW-MH:
Heavy tailed proposals lead to higher rates of convergence

Focus:
Polynomially ergodic Markov chains

For polynomial target distributions:
they derive polynomial rates of convergence

Application of control variates to RW-MH algorithms - Heavy tailed distributions

Table: Existence of central limit theorems for RW-MH algorithm (from Jarner and Roberts, 2007)

	Target distribution					
Proposal distr.	$t(2.5)$	$t(3)$	$t(4)$	$t(5)$	$t(6)$	$t(7)$
Uniform			L	C	C	C
Normal			L	C	C	C
Cauchy		L	C	C	C	C
$t(0.5)$	L	L	C	C	C	C

C: CLT holds for $|x|$, L: CLT holds for $|x|^s$, $s < 1$

Application of control variates to RW-MH algorithms - Heavy tailed distributions

Table: VRF's for $F(x) = |x|$, with $G_i = x^i$, $i = 1, 2, 3$ ($n = 200,000$, $n_{PG} = 50$)

	Target distribution					
Proposal distr.	$t(2.5)$	$t(3)$	$t(4)$	$t(5)$	$t(6)$	$t(7)$
Uniform	8.95	5.79	3.78	3.17	3.24	2.28
Normal	0.09	4.54	4.08	2.73	2.72	2.80
Cauchy	3.54	3.95	3.46	3.75	3.83	4.93
$t(0.5)$	3.47	3.27	2.89	3.61	6.06	9.28

Summary and Further Research

Summary

- A solid methodological framework has been provided for the development and use of control variates in MCMC
- For given G function, consistent estimates for optimal coefficients θ are defined

Main reference:

Dellaportas, P. and Kontoyiannis, I. (2008). Control Variates for Reversible MCMC samplers. Submitted to JRSS(Series B)

Summary and Further Research

Summary

- A solid methodological framework has been provided for the development and use of control variates in MCMC
- For given G function, consistent estimates for optimal coefficients θ are defined

Further research

- Approaches for elicitation of G functions and decision on k
- Techniques for more efficient derivation of $PG(x)$
- Extension to Reversible-Jump MCMC algorithms

Main reference:

Dellaportas, P. and Kontoyiannis, I. (2008). Control Variates for Reversible MCMC samplers. Submitted to JRSS(Series B)

Summary and Further Research

Thank you for listening :)