Approximating Martingales in Continuous and Discrete Time Markov Processes

Rohan Shiloh Shah

May 6, 2005

Contents

1 Introduction

Markov Chains are often so complex that an exact solution for the steady-state probabilities (or other 'features' of the Markov Chain) are not computable. It is therefore necessary to use variance reducing approximations. There are many techniques currently in use; importance sampling, smoothing, and "control variates and regression" [1]. All these methods make use of simulations of the complex process or simulations of a simpler, related process that is used as an external control in simulating the complex process. The problem with this approach is that both the complex and simple process need to be simulated *simultaneously*. The approach we consider differs in that it constructs a Martingale from the simpler related process which is then used as an internal control in the simulation of the more complicated process. The construction of the martingale is entirely dependent on the 'feature' we choose to measure.[1] The benefit of such an approach is that the simple and complex markov processes do *not* need to be simulated simultaneously.

1.1 *Approximating Martingale Process Method*

We construct a system of linear equations that give us the exact solution for some feature of a simpler Markov Chain that approximates our complex Markov Chain. A martingale is then constructed from this 'exact-approximate' solution. This martingale will then be used to define an unbiased, variance reducing estimate (using linear combinations of more standard estimates) of a measure of the feature we choose. This new estimate can then be used to construct another newer estimate, again using an appropriate martingale and in this way we can get unlimited variance reduction. I will illustrate this mechanism using several examples.

2 Two State Continuous-Time Markov Chain

Let us define a continuous-time markov chain $\{Y_t : t \geq 0\}$ on the state space $S = \{y_0, y_1\}$, where the time spent in state y_i is exponentially distributed with mean $\frac{1}{\lambda_i}$. By the Markov property, the time spent in each state is memoryless - the remaining time spent in any state, before a transition to

[1] More specifically, the construction of the martingale is dependent on a system of linear equations which depend on the 'feature' we choose to measure.

another state occurs, is dependent only on the state and not on the amount of time already spent in that state. The Markov process is determined in this case by a generator or rate matrix of the form

$$p_{i,j} = \lim_{\delta t \to 0} \frac{P(Y_{t+\delta t} = j | Y_t = i)}{\delta t} \tag{1}$$

where each i, j entry is the probability per time unit that a transition between states i and j result. The rate matrix for the continuous-time Markov chain Y_t is then

$$A = \begin{bmatrix} -\lambda_0 & \lambda_0 \\ \lambda_1 & -\lambda_1 \end{bmatrix} \tag{2}$$

Now define a reward function f such that $f(y_0) = 1$ and $f(y_1) = 0$. We can approximate $\alpha(t)$, the fraction of time spent in state y_0 in the interval $[0, t]$ by

$$\alpha(t) \approxeq \frac{1}{t} \int_0^t f(Y_s) ds \longrightarrow \frac{\lambda_1}{\lambda_0 + \lambda_1} \text{ as } t \to \infty \tag{3}$$

Define $\mathcal{F}_t$ to be the σ-algebra generated by $\{Y_s : 0 \leq s \leq t\}$ and construct the history or filtration $\{\mathcal{F}_t : t \geq 0\}$ such that $\mathcal{F}_t \subset \mathcal{F}_s$ whenever $s < t$. Now let $P_{y_0}(y) = P(Y_t = y | Y_0 = y_0)$ and then $P_\mu(y) = \int_S P_{y_0}(Y_t = y) \mu(dy_0)$.

Definition 1. *$M = \{M_t : t \geq 0\}$ is a martingale under probability measure P_μ with respect to filtration $\{\mathcal{F}_t : t \geq 0\}$ iff (1) M_t is $\mathcal{F}_t$-measurable for all $t > 0$ (2) M_t is integrable for all $t > 0$ (i.e. $E\{M_t | Y_0 = y_0\} < \infty$) and (3) if $s \leq t$ then $E(M_t | \mathcal{F}_s) = M_s$ a.s. where the expectation is taken with respect to probability measure P_μ*

We now construct a P_μ martingale with mean zero as follows (see Theorem 1)

$$M(t) = u(Y_t) - u(Y_0) - \int_0^t Au(Y_s) ds \tag{4}$$

where u is any function $u : S \to \mathbb{R}$. If we now simply construct a new unbiased estimator of the fraction of time spent in state y_0 as follows $\hat{\alpha}_t = \alpha_t - \frac{M(t)}{t}$ then we have reduced the variance since $var\{\hat{\alpha}_t\} = var\{\alpha_t - \frac{M(t)}{t}\} \leq var\{\alpha_t\}$, without adding any bias so the estimators have the same expected value. This holds for any choice of function u.

There is however an optimal choice for u that will change the martingale and estimator such that unlimited variance reduction will be possible. Let us take u as follows; $u(0) = 0$ and $u(1) = -(\lambda_0 + \lambda_1)^{-1}$. Then

$$Au(x) = -f(x) + \frac{\lambda_1}{\lambda_0 + \lambda_1} \tag{5}$$

and so we have

$$\begin{aligned}
\hat{\alpha}_t &= \alpha_t - \frac{M(t)}{t} \\
&= \frac{1}{t}\int_0^t f(Y_s)ds - \frac{u(Y_t) - u(Y_0)}{t} + \frac{1}{t}\int_0^t Au(Y_s)ds \\
&= \frac{\lambda_1}{\lambda_0 + \lambda_1} + \frac{u(Y_0) - u(Y_t)}{t} \qquad (6)
\end{aligned}$$

where $var\{\alpha(t)\} \propto \frac{1}{t}$ and $var\{\hat{\alpha(t)}\} \propto \frac{1}{t^2}$ so we can continue the above process by setting $\alpha(t) = \hat{\alpha}(t)$ and using the martingale M to derive a new estimator and in this way unlimited variance reduction is possible. We now apply the above ideas to the approximation of several other relevant 'features'.

3 Expected total reward prior to absorption

3.1 *In a Discrete-Time Markov Chain*

Let us define a discrete-time markov chain $\{X_t : t \geq 0\}$ on the state space $\Sigma = \{x_0, x_1, \cdots\}$ which is countably infinite. Also define a reward function f on Σ. If C is a subset of Σ such that each $x_i \in C$ is an absorbing state then we define $T = \inf\{n \geq 1 : X_n \in C\}$ to be the time to absorption. We can also define the expected total cumulative reward before absorption given that the state starts in $x \in C^c$ by

$$u^*(x) = E\{\sum_{i=0}^{T-1} f(X_i)|X_0 = x\} \qquad (7)$$

Let us make two assumptions that will be useful at a later stage; first assume that $E\{T|X_0 = x\}$ is finite $[A1]$ and that the reward function f is also bounded for all $x \in C^c$ $[A2]$. Combining these assumptions we have that u^* is also bounded since

$$\begin{aligned}
u^*(x) &= E\{\sum_{i=0}^{T-1} f(X_i)|X_0 = x\} \\
&\leq \|f\|E\{T|X_0 = x\} \\
&= \sup_{x \in C^c} |f(x)|E\{T|X_0 = x\} < \infty \qquad (8)
\end{aligned}$$

Now given a total of N sample trajectories $\{X_i : 1 \leq i \leq T | X_0 = x\}_{k=1}^N$, let us define U_k to be the observed cumulative reward for the kth trajectory $\{\sum_{i=0}^{T-1} f(X_i) | X_0 = x\}_k$. Then $\alpha = u^*(x)$ can be estimated by $\alpha_N = \frac{1}{N} \sum_{k=1}^N U_k$. By the markov property we can show that u^* satisfies the linear system

$$u(x_t) = f(x_t) + Bu(x_{t+1}) \tag{9}$$

where B is the restriction of P to C^c (i.e. $P(x, y) = B(x, y) \forall x, y \in C^c$) and $Bu(x_{t+1}) = \int_{C^c} u(y) B(x_{t+1}, dy)$ for $x_{t+1} \in C^c$. For $x \in C$ we will assume that $u^*(x) = f(x) = 0$ and hence

$$u^* = f + Pu^* \text{ for } x \in \Sigma \tag{10}$$

Theorem 1. *As in the previous section, define $\mathcal{F}_t$ to be the σ-algebra generated by $\{X_s : 0 \leq s \leq t\}$ and construct the history or filtration $\{\mathcal{F}_t : t \geq 0\}$ such that $\mathcal{F}_t \subset \mathcal{F}_s$ whenever $s < t$. Now let $P_{x_0}(x) = P(X_t = x | X_0 = x_0)$ and then let the two assumptions [A1] and [A2] hold so that we have for a function $u : \Sigma \rightarrow \mathbb{R}$ such that $u(x) = 0$ for $x \in C$ and $0 \leq u(x) \leq \infty$ for $x \in C^c$ the following*

$$M_n = u(X_n) - u(X_0) - \sum_{k=0}^{n-1} (P - I) u(X_k) \tag{11}$$

where for a fixed and bounded constant b we have $|(P - I)u(x)| \leq b$ for $x \in C^c$. It follows that $M = (M_n : n \geq 0)$ is a P_x martingale for all $x \in \Sigma$.

If we have the additional condition $E\{\sum_{i=0}^{T-1} u(X_k) | X_0 = x\} < \infty$, then

$E(M_T|X_0 = x) = 0$ for all $x \in C^c$. To see this observe that

$$\begin{aligned}
E_x(\sum_{k=1}^{T} |M_{k+1} - M_k|\,|\mathcal{F}_k) &= E_x(\sum_{k=1}^{T} |u(X_{k+1}) - u(X_0) - \sum_{i=0}^{k}(P-I)u(X_i) \\
&- u(X_k) + u(X_0) + \sum_{i=0}^{k-1}(P-I)u(X_i)||\mathcal{F}_k) \\
&= E_x(\sum_{k=1}^{T} |u(X_{k+1}) - Pu(X_k)||X_k) \\
&\leq \sum_{k=1}^{T} E_x(u(X_{k+1})|X_k) + \sum_{k=1}^{T} Pu(X_k) \\
&= \sum_{k=1}^{T} 2Pu(X_k) \\
&\leq \sum_{k=1}^{T} 2(u(X_k) + b) < \infty \qquad (12)
\end{aligned}$$

Now applying the dominated convergence theorem we have that $E\{M_T|X_0 = x\} = 0$ since the following hold

$$\begin{aligned}
&E\{M_{\min(T,m)}|X_0 = x\} = 0 \text{ for all } m \\
&\lim_{m\to\infty} M_{\min(T,m)} = M_T a.s. \\
&\sup_m \left| M_{\min(T,m)}| \sum_{k=1}^{T} |M_{k+1} - M_k| \right| \qquad (13)
\end{aligned}$$

We now define the martingale estimator as

$$\hat{\alpha}_n = \frac{1}{n}\sum_{k=1}^{n}(U_k - M_T(k)) = \alpha_n - \frac{1}{n}\sum_{k=1}^{n} M_T(k) \qquad (14)$$

where $M_T(k)$ is M_T (as defined in Theorem 1) for the kth trajectory $\{X_i : 0 \leq i \leq T\}_k$. If we take u in 11 to be equal to u^* (from Equation 7) then under the probability law P_x defined in section 2 we have

$$M_T = -u^*(x) + U \qquad (15)$$

which implies that $var\{\hat{\alpha}_n\} = \frac{1}{n^2} var\{\sum_{k=1}^{n} U_k - M_T(k)\} = \frac{1}{n^2} var\{\sum_{k=1}^{n} u^*(x)\} =$

$var\{u^*(x)\} = 0$. This follows from the following

$$\begin{aligned} E_x M_T &= E_x\{u^*(X_T) - u^*(X_0) - \sum_{k=0}^{T-1}(P_x - I)u^*(X_k)\} \\ &= E_x\{-u^*(X_0) + \sum_{k=0}^{T-1} f\} \\ &= -u^*(X_0 = x) + E_x\{\sum_{k=0}^{T-1} f\} \\ &= -u^*(X_0 = x) + u^*(X_0 = x) = 0 \end{aligned} \tag{16}$$

where Equation 10 gives us that $(P_x - I)u^* = -f$ and $u^*(X_T) = 0$ since $X_T \in C$ and since $X_0 = x$ under P_x.

3.2 *In a Continous-Time Markov Chain*

Use the definitions in Section 2 for a CTMC (and let the state space S be countably infinite) and redefine the following from Section 3

$$T = \inf\{n \geq 1 : Y_n \in C\} \tag{17}$$

$$u^*(y) = E\{\int_0^T f(Y_i)|Y_0 = y\} \tag{18}$$

$$\alpha_N = \frac{1}{N}\sum_{k=1}^{N} U_k \tag{19}$$

$$U_k = \sum_{i=0}^{T-1} f(Y_i(k)) \tag{20}$$

Note that (18) implies $u^*(x) = 0 \forall x \in C$.

Theorem 2. *Again, define $\mathcal{F}_t$ to be the σ-algebra generated by $\{X_s : 0 \leq s \leq t\}$ and construct the history or filtration $\{\mathcal{F}_t : t \geq 0\}$ such that $\mathcal{F}_t \subset \mathcal{F}_s$ whenever $s < t$. Now let $P_{y_0}(y) = P(Y_t = y|Y_0 = y_0)$ and $P_\mu(y) = \int_S P_{y_0}(Y_t = y)\mu(dy_0)$ so that we have for a function $u : S \to \mathbb{R}$ the following*

$$M_n = u(Y_n) - u(Y_0) - \int_0^t Au(Y_n)ds \tag{21}$$

It follows that $M = (M_n : n \geq 0)$ is a P_μ martingale for any choice of μ. Also setting $u = u^\star$ makes $\hat{\alpha}$ have zero variance.

4 Infinite Horizon Expected Discounted Rewards

Let us define a discrete-time markov chain $\{X_t : t \geq 0\}$ on the state space $\Sigma = \{x_0, x_1, \cdots\}$ which is countably infinite. Also define a reward function f on Σ and *further assume that f is bounded.* We now define a discounting function h such that $0 \leq h(x) \leq \delta < 1$ for all $x \in \Sigma$. We can now discount rewards in the infinite horizon as follows:

$$U = \sum_{i=0}^{\infty} f(X_i) \times \left[\prod_{k=0}^{i-1} h(X_k)\right] \tag{22}$$

since $\left[\prod_{k=0}^{i-1} h(X_k)\right] < 1$. The 'measure' we are interested in is the following

$$u^*(x) = E(U|X_0 = x) \leq \frac{\|f\|}{1-\delta} \tag{23}$$

and is bounded. We cannot directly simulate the process to estimate U or u^* because of the infinite horizon. We could use an unbiased estimate that runs the process for a random number of time steps and then uses a finite sum in 22 to estimate U. We know that u^* satisfies the following linear system of equations

$$u(x) = f(x) + h(x)Pu(x) \ \forall x \in \Sigma \tag{24}$$

Theorem 3. *For a function h such that $0 \leq h(x) \leq \delta < 1$ for all $x \in \Sigma$ and a bounded function $u : \Sigma \to \mathbb{R}$ ($|u(x)| \leq b$), we define*

$$M_n = \sum_{k=1}^{n} [u(X_k) - Pu(X_{k-1})] \prod_{j=0}^{k-1} h(X_j) \tag{25}$$

then $M = \{M_t : t \geq 0\}$ is a P_x martingale for all $x \in \Sigma$ so that $M_n \to M_\infty$ a.s. such that $E_x M_\infty$.

If we generate n different finite trajectories under P_x then we can estimate $\alpha = u^*(x)$ by the average $\alpha_n = \frac{1}{n}\sum_{i=1}^{n} U_i$. We now define the martingale estimator

$$\hat{\alpha}_n = \frac{1}{n}\sum_{i=1}^{n} U_i - M_\infty(i) \tag{26}$$

which is an unbiased estimator. If we take u in 25 to be equal to u^* then $M_\infty = U - u^*(X_0)$, so that under P_x, $\hat{\alpha}_n$ is an estimator for $u^*(x)$ such that

$var(\hat{\alpha}_n) = 0$. To see this observe that

$$
\begin{aligned}
M_\infty &= \sum_{k=1}^{\infty} \left[\prod_{j=0}^{k-1} h(X_j) \right] u(X_k) - Pu(X_{k-1}) \\
&= -h(X_0)Pu(X_0) + \sum_{k=1}^{\infty} \left[\prod_{j=0}^{k-1} h(X_j) \right] u(X_k) - h(X_k)Pu(X_{k-1}) \\
&= f(X_0) - u(X_0) + \sum_{k=1}^{\infty} f(X_k) \left[\prod_{j=0}^{k-1} h(X_j) \right] \\
&= U - u^*(X_0) \qquad (27)
\end{aligned}
$$

since $u(X_0) = f(X_0) + h(X_0)Pu(X_0)$ and $u(X_n) = f(X_n) + h(X_n)Pu(X_{n-1})$.

5 Finite Horizon Cumulative Rewards

As in the previous section, we define a discrete-time markov chain $\{X_t : t \geq 0\}$ on the state space $\Sigma = \{x_0, x_1, \cdots\}$ which is countably infinite. Also assume that the chain terminates with probability one. Also define a reward function f on Σ and *further assume that f is bounded $\|f\| < \infty$.* We now define

$$u_n^*(x) = E\left(\sum_{k=0}^{n} f(X_k) | X_0 = x \right) \qquad (28)$$

Next we define α_m and the martingale estimator as follows

$$
\begin{aligned}
\alpha_m &= \sum_{i=0}^{m} \left(\{ \sum_{k=0}^{n} f(X_k(i)) \} \right) \\
\hat{\alpha}_m &= \sum_{i=0}^{m} \left(\{ \sum_{k=0}^{n} f(X_k(i)) \} - M_n(i) \right) \\
M_n(i) &= \sum_{k=1}^{n} [u(X_k(i)) - Pu(X_{k-1}(i))] \prod_{j=0}^{k-1} h(X_j(i)) \qquad (29)
\end{aligned}
$$

where $\{X_k(i) : k \geq 0\}$ is the ith (of a total of m) trajectory simulated. We also know that $u_n^*(x)$ satisfies the linear system

$$
\begin{aligned}
u_0 &= f \\
u_j &= Pu_{j-1} + f \;\; j = 1, \cdots, n \qquad (30)
\end{aligned}
$$

If we now take $u_j = u_j^*$ for $j = 1, \cdots, n$ then

$$
\begin{aligned}
M_n(i) &= \sum_{k=1}^{n} u_{n-k}(X_k(i)) - P u_{n-k}(X_{k-1}(i)) \\
&= \sum_{k=0}^{n} f(X_k(i)) - u_n(X_0(i)) \qquad (31)
\end{aligned}
$$

so that under P_x we have that

$$
\begin{aligned}
var\{\hat{\alpha}_n\} &= var\left\{\sum_{i=0}^{m}\left(\{\sum_{k=0}^{n} f(X_k(i))\} - M_n(i)\right)\right\} \\
&= var\left\{\sum_{i=0}^{m} u_n(X_0(i))\right\} \\
&= 0 \qquad (32) \\
& \qquad (33)
\end{aligned}
$$

6 Average Steady-State Reward

References

[1] Heymann, Ott and Andradottir - "Variance Reduction through Smoothing and Control Variates for Markov Chain Simulations"

[2] Remi Munos, Emmanuel Gobet - "Sensitivity analysis using Ito-Malliavin calculus and martingales"

[3] Henderson and Glynn - "Approximating Martingales for Variance Reduction in Markov Process Simulation"