# Temp

Tainze Wang

November 11, 2021

We want to design an unbiased Monte Carlo estimator for $g\left(\mathbb{E}_{\pi}\left(X\right)\right)$ where $X$ cannot be i.i.d. sampled. First, multiple calling the algorithm from Jacob, O'Leary, and Atchadé [2] gives us independent samples of an estimator $\hat{H}$ such that $\mathbb{E}(\hat{H}) = \mathbb{E}_{\pi}(X)$. Then, with these samples and method in Blanchet and Glynn [1], we are able to create an unbiased estimator $\hat{Z}$ such that $\mathbb{E}(\hat{Z}) = g\left(\mathbb{E}_{\pi}\left(X\right)\right)$.

## Estimator

First we briefly recall the unbiased MCMC mechanism developed by Jacob, O'Leary, and Atchadé. Suppose one is interested in the quantity $\mathbb{E}_{\pi}(X)$ but is only able to sample $\pi$ asymptotically through some Markov chain. The unbiased MCMC method utilizes a coupled pair of Markov chains $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$. Both start from the same distribution and evolve according to the same Markov transition kernel with stationary distribution $\pi$. We assume the two chains will meet as some random but almost surely finite time $\tau$, and stay together after meeting, i.e., $X_t = Y_{t-1}$ for all $t \geq \tau$. Define

$$\hat{H}(X,Y) = X_0 + \sum_{t=1}^{\tau-1}(X_t - Y_{t-1})$$

Then $\hat{H}(X,Y)$ is an unbiased estimator for the mean of $\pi$ assuming certain regularity conditions, see [2] for details.

Assuming the above algorithm is called independently for $m$ times, and we denote the outputs by $H_1, H_2, \cdots, H_m$. These outputs can be viewed as $i.i.d.$ samples from some distribution $\tilde{\pi}$ which has the same expectation as $\pi$. We may assume $m$ is an even number and define the following notations:

$$\begin{aligned} S_H(m) &= H(1) + ... + H(m) \\ S_H^O(m) &= H(1) + H(3) + ... + H(2m-1) \\ S_H^E(m) &= H(2) + H(4) + ... + H(2m) \end{aligned}$$

Finally, define

$$\triangle_n = g\left(\frac{S_H(2^{n+1})}{2^{n+1}}\right) - \frac{1}{2}\left(g\left(\frac{S_H^O\left(2^n\right)}{2^n}\right) + g\left(\frac{S_H^E\left(2^n\right)}{2^n}\right)\right)$$

Let $K \in \mathbb{N}^+$ be a random variable independent of $H(k)$'s and $p(k) = P(K = k)$. Our estimator is given by:

$$\hat{Z} = \frac{\triangle_N}{p(N)} + g\left(H\left(1\right)\right).$$

We denote by $V_n \subset \mathbb{R}^d$ the range of $(H(1) + \cdots + H(n))/n$ for every $n$, and let $V := \cup_{n=1}^{\infty} V_n$. Then we pose the following conditions on both $g$ and $H$:

- **Assumption 1:** The function $g : O \to \mathbb{R}$ where $O \subset \mathbb{R}^d$ contains $V$. Moreover, the mean vector $\mu = (\mathbb{E}_{\pi_1}(X_1), \cdots, \mathbb{E}_{\pi_d}(X_d)) \in O^\circ$.
- **Assumption 2:** The function $g$ is continuously differentiable in a neighborhood of $\mu$, and $Dg(\cdot)$ is locally Hölder continuous with exponent $\alpha > 0$. In other words, there exists $\varepsilon > 0$, $\alpha > 0$ and $c = c(\epsilon) > 0$ such that the following inequality holds for every $x, y \in (\mu - \epsilon, \mu + \epsilon)$:
$$\|Dg(x) - Dg(y)\|_2 \le c\|x - y\|_2^\alpha.$$
- **Assumption 3:** There exists some $l > 2 + \alpha$ such that $H$ has finite $l$-th moments, i.e., $\mathbb{E}\left(\|H\|^l\right) < \infty$
- **Assumption 4:** There exist constants $s > 1$, $\alpha_s \ge 0$, and $C_s > 0$ such that $2\alpha_s + (s-1)l > 2s$ and $\mathbb{E}(\|\Delta_n\|^{2s}) \le C_s 2^{-\alpha_s n}$ for every $n$.

We briefly comment on the assumptions above. The first assumption requires the domain of $g$ covers all the possible values of [GW: To be finished].

[GW: Rewrite the statement of the theorem]

### Theorem 1:

Under **assumptions 1** to **4**, if N [GW: $N$] is geometrically distributed with success parameter $p \in \left(\frac{1}{2}, 1 - \frac{1}{2^{(1+\alpha)}}\right)$, then $\mathbb{E}(\hat{Z}) = g\left(\mathbb{E}\left(X\right)\right)$, $Var(\hat{Z}) < \infty$ [GW: Var]

***Proof:***

We first show the unbiasedness of estimator $\hat{Z}$.

$$\begin{aligned}
\mathbb{E}(\triangle_n) &= \mathbb{E}\left(g\left(\frac{S_H(2^{n+1})}{2^{n+1}}\right) - \frac{1}{2}\left(g\left(\frac{S_H^O(2^n)}{2^n}\right) + g\left(\frac{S_H^E(2^n)}{2^n}\right)\right)\right) \\
&= \mathbb{E}\left(g\left(\frac{S_H(2^{n+1})}{2^{n+1}}\right)\right) - \mathbb{E}\left(g\left(\frac{S_H(2^n)}{2^n}\right)\right)
\end{aligned}$$

Law of large numbers gives us

$$\lim_{n\to\infty} \mathbb{E}\left(g\left(\frac{S_H(n)}{n}\right)\right) = g\left(\mathbb{E}(H)\right)$$

Therefore,

$$\sum_{n=0}^{\infty} \mathbb{E}(\triangle_n) = g\left(\mathbb{E}(H)\right) - \mathbb{E}\left(g\left(H\right)\right)$$

Then,

$$\begin{aligned}
\mathbb{E}(\hat{Z}) = \mathbb{E}\left(\mathbb{E}\left(\hat{Z}|N\right)\right) &= \mathbb{E}\left(\mathbb{E}\left(\frac{\triangle_N}{p(N)}|N\right)\right) + \mathbb{E}\left(g\left(H\right)\right) \\
&= \mathbb{E}\left(\sum_{n=0}^{\infty} \frac{\triangle_n}{p(n)} \cdot p(n)\right) + \mathbb{E}\left(g\left(H\right)\right) \\
&= \sum_{n=0}^{\infty} \mathbb{E}(\triangle_n) + \mathbb{E}\left(g\left(H\right)\right) \\
&= g\left(\mathbb{E}(H)\right) = g\left(\mathbb{E}_\pi\left(X\right)\right)
\end{aligned}$$

where the interchangeability between expectation and infinite sum is given by $\mathbb{E}\left(\frac{\triangle_N}{p(N)}\right)^2 < \infty$, which will be proved below.
Now we show $\mathbb{E}(\triangle_n^2) = O(2^{-(1+\gamma)n})$ for some $\gamma > 0$. Let $\delta > 0$ and in a neighborhood of size $\delta$ around $\mu$ **assumption 2** is satisfied.

$$\begin{aligned}
|\triangle_n| &= |\triangle_n| I\left(\max\left(\left\|\frac{S_H^O(2^n)}{2^n} - \mu\right\|_2, \left\|\frac{S_H^E(2^n)}{2^n} - \mu\right\|_2\right) > \frac{\delta}{2}\right) \\
&\quad + |\triangle_n| I\left(\left\|\frac{S_H^O(2^n)}{2^n} - \mu\right\|_2 \leq \frac{\delta}{2}, \left\|\frac{S_H^E(2^n)}{2^n} - \mu\right\|_2 \leq \frac{\delta}{2}\right) \\
&\leq |\triangle_n| I\left(\left\|\frac{S_H^O(2^n)}{2^n} - \mu\right\|_2 > \frac{\delta}{2}\right) + |\triangle_n| I\left(\left\|\frac{S_H^E(2^n)}{2^n} - \mu\right\|_2 > \frac{\delta}{2}\right) \\
&\quad + |\triangle_n| I\left(\left\|\frac{S_H^O(2^n)}{2^n} - \mu\right\|_2 \leq \frac{\delta}{2}, \left\|\frac{S_H^E(2^n)}{2^n} - \mu\right\|_2 \leq \frac{\delta}{2}\right)
\end{aligned}$$

When $\left\|\frac{S_H^O(2^n)}{2^n} - \mu\right\|_2 \leq \frac{\delta}{2}$ and $\left\|\frac{S_H^E(2^n)}{2^n} - \mu\right\|_2 \leq \frac{\delta}{2}$, we have $\left\|\frac{S_H(2^{n+1})}{2^{n+1}} - \mu\right\|_2 \leq \delta$.

$$\begin{aligned}
\triangle_n &= \frac{1}{2}\left(g\left(\frac{S_H(2^{n+1})}{2^{n+1}}\right) - g\left(\frac{S_H^O(2^n)}{2^n}\right)\right) + \frac{1}{2}\left(g\left(\frac{S_H(2^{n+1})}{2^{n+1}}\right) - g\left(\frac{S_H^E(2^n)}{2^n}\right)\right) \\
&= \frac{1}{2} Dg(\xi_n^O)\left(\frac{S_H(2^{n+1})}{2^{n+1}} - \frac{S_H^O(2^n)}{2^n}\right) + \frac{1}{2} Dg(\xi_n^E)\left(\frac{S_H(2^{n+1})}{2^{n+1}} - \frac{S_H^E(2^n)}{2^n}\right) \\
&= \frac{1}{4}\left(Dg(\xi_n^O) - Dg(\xi_n^E)\right) \frac{S_H^O(2^n) - S_H^E(2^n)}{2^n}
\end{aligned}$$

where $\xi_n^O$ between $\frac{S_H(2^{n+1})}{2^{n+1}}$ and $\frac{S_H^O(2^n)}{2^n}$, $\xi_n^E$ between $\frac{S_H(2^{n+1})}{2^{n+1}}$ and $\frac{S_H^E(2^n)}{2^n}$. Thus

$$\|\triangle_n\|_2 \leq c\left\|\xi_n^O - \xi_n^E\right\|_2^\alpha \cdot \left\|\frac{S_H^O(2^n) - S_H^E(2^n)}{2^n}\right\|_2 \leq c\left\|\frac{S_H^O(2^n) - S_H^E(2^n)}{2^n}\right\|_2^{1+\alpha}$$

Then,

$$\begin{aligned}
&\mathbb{E}\left(|\triangle_n|^2 I\left(\left\|\frac{S_H^O(2^n)}{2^n} - \mu\right\|_2 \leq \frac{\delta}{2}, \left\|\frac{S_H^E(2^n)}{2^n} - \mu\right\|_2 \leq \frac{\delta}{2}\right)\right) \\
\leq& c\mathbb{E}\left(\left\|\frac{S_H^O(2^n) - S_H^E(2^n)}{2^n}\right\|_2^{2(1+\alpha)}\right) = O(2^{-(1+\alpha)n})
\end{aligned}$$

To analyze $\mathbb{E}\left(|\triangle_n|^2 I\left(\left\|\frac{S_H^O(2^n)}{2^n} - \mu\right\|_2 > \frac{\delta}{2}\right)\right)$,

$$\begin{aligned}
\mathbb{E}\left[|\Delta_n|^2 I\left(\left\|\frac{S_H^O(2^n)}{2^n} - \mu\right\|_2 > \frac{\delta}{2}\right)\right] &\leq (\mathbb{E}[|\Delta_n|^{2s}])^{1/s} \mathbb{P}\left(\left\|\frac{S_H^O(2^n)}{2^n} - \mu\right\|_2 > \frac{\delta}{2}\right)^{(s-1)/s} \\
&\leq C_s^{1/s} 2^{-\alpha_s n/s}\left(\frac{\delta}{2}\right)^{-l(s-1)/s} \mathbb{E}\left[\|\frac{S_H^O(2^n)}{2^n} - \mu\|_2^l\right] \\
&\leq C(s,l,d) 2^{-\alpha_s n/s - ln/2} = C(s,l,d) 2^{-n(\alpha_s + 0.5(s-1)l)/s}
\end{aligned}$$

where the last inequality follows from the Marcinkiewicz–Zygmund inequality.

By our assumption $2\alpha_s + (s-1)l > 2s$, it is clear that $(\alpha_s + 0.5(s-1)l)/s > 1$, and therefore $\mathbb{E}\left(|\triangle_n|^2 I\left(\left\|\frac{S_H^O(2^n)}{2^n} - \mu\right\|_2 > \frac{\delta}{2}\right)\right) = \mathcal{O}(2^{-(1+\tilde{\alpha})n})$ for some $\tilde{\alpha} > 0$. Combining the two parts, we conclude $\mathbb{E}[\Delta_n^2] = \mathcal{O}(2^{-(1+\gamma)n})$ where $\gamma = \min\{\beta, \alpha_s + \frac{(s-1)l}{2s} - 1\} > 0$.

Similarly, we have $\mathbb{E}\left(|\triangle_n|^2 I\left(\left\|\frac{S_H^E(2^n)}{2^n} - \mu\right\|_2 > \frac{\delta}{2}\right)\right) = \mathcal{O}(2^{-n(1+\gamma)})$. Hence, $\mathbb{E}(\triangle_n^2) = O(2^{-(1+\alpha)n})$ for all $n > 0$ is proved.

[GW: The following is still under construction] If we pick N geometrically distributed with parameter $k \in \left(\frac{1}{2}, 1 - \frac{1}{2^{(1+\alpha)}}\right)$ such that $p(n) = (1-k)^n k$, then

$$\mathbb{E}\left(\frac{\triangle_N}{p(N)}\right)^2 = \mathbb{E}\left(\mathbb{E}\left(\left(\frac{\triangle_N}{p(N)}\right)^2 |N\right)\right) = \mathbb{E}\left(\sum_{n=0}^{\infty} \frac{\triangle_n^2}{p^2(n)} \cdot p(n)\right)$$
$$= O\left(\sum_{n=0}^{\infty}\left(\frac{1}{(1-r)\cdot 2^{1+\alpha}})\right)^n\right) < \infty$$

which proves that $\mathbb{E}(\hat{Z}) = g\left(\mathbb{E}\left(X\right)\right)$, $Var(\hat{Z}) < \infty$

### Complexity and parameter of Geometric Distribution

For the selection of parameter in the geometric distribution, we want to minimize the complexity given below

$$\left(\sum_{n=0}^{\infty} 2^n \cdot p(n)\right) \times \left(\sum_{n=0}^{\infty} \frac{2^{-(1+\alpha)n}}{p(n)}\right) = \left(\sum_{n=0}^{\infty} (2(1-k))^n\right) \times \left(\sum_{n=0}^{\infty}\left(\frac{1}{(1-k)2^{1+\alpha}}\right)^n\right)$$
$$= \frac{1}{1-2(1-r)} \cdot \frac{(1-k)2^{1+\alpha}}{(1-k)2^{1+\alpha}-1}$$

Differentiate with respect to k and we can get the optimal selection where $k^* = 1 - 2^{-(2+\alpha)/2}$

$$p(n) = \left(\frac{1}{2^{(2+\alpha)/2}}\right)^n \left(1 - \frac{1}{2^{(2+\alpha)/2}}\right)$$

## References


[1] Jose H. Blanchet and Peter W. Glynn. Unbiased Monte Carlo for optimization and functions of expectations via multi-level randomization. *2015 Winter Simulation Conference (WSC)*, pages 3656–3667, 2015.

[2] Pierre E Jacob, John O'Leary, and Yves F Atchadé. Unbiased markov chain monte carlo methods with couplings. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(3):543–600, 2020.