Expressions for the Markov Chain CLT Variance

by

Jeffrey S. Rosenthal*

(April 2007)

1. Introduction.

This short paper considers certain issues surrounding the variance in the central limit theorem (CLT) for Markov chains. This subject is particularly important when using Markov chain Monte Carlo (MCMC) algorithms, see e.g. [11], [18], [7], [4], [5], [8], and [10].

Let $\{X_n\}$ be a stationary, reversible Markov chain on a state space $\mathcal{X}$, and let $h : \mathcal{X} \to \mathbf{R}$ be a mean 0 measurable function. Often, $n^{-1/2} \sum_{i=1}^{n} h(X_i)$ will converge weakly to Normal$(0, \sigma^2)$ for some $\sigma^2 < \infty$. The asymptotic variance σ^2 is very important in applications, and various alternate expressions for it are available in terms of limits, autocovariances, and spectral theory.

This paper considers three such expressions, denoted A, B, and C, which are known to "usually" equal σ^2. These expressions arise in different applications in different ways. For example, it is proved by Kipnis and Varadhan [11] that if $C < \infty$, then a $\sqrt{n}$-CLT exists for h, with $\sigma^2 = C$. In a different direction, it is proved by Roberts [13] that Metropolis algorithms satisfying a certain condition must have $A = \infty$. Such disparate results indicate the importance of sorting out the relationships between A, B, C, and σ^2.

2. Notation.

Let $\{X_n\}$ be a stationary, time homogeneous Markov chain on the measurable space $(\mathcal{X}, \mathcal{F})$, with transition kernel P, reversible with respect to the probability measure $\pi(\cdot)$, so $\mathbf{P}[X_n \in A] = \pi(A)$ for all $n \in \mathbf{N}$ and $A \in \mathcal{F}$. Let $P^n(x, A) = \mathbf{P}[X_n \in A \,|\, X_0 = x]$ be the n-step transitions. Recall that P is *ergodic* if $\lim_{n\to\infty} \sup_{A\in\mathcal{F}} |P^n(x, A) - \pi(A)| = 0$ for π-a.e. $x \in \mathcal{X}$. This follows (cf. [18], [16], [15]) if P is ϕ-irreducible and aperiodic.

Write $\pi(g) = \int_{\mathcal{X}} g(x)\, \pi(dx)$, and let $L^2(\pi) = \{f : \mathcal{X} \to \mathbf{R} \ s.t.\ \pi(f^2) < \infty\}$. Write $\langle f, g\rangle = \int_{\mathcal{X}} f(x)\, g(x)\, \pi(dx)$ for $f, g \in L^2(\pi)$; by reversibility, $\langle f, Pg\rangle = \langle Pf, g\rangle$.

*Department of Statistics, University of Toronto, Toronto, Ontario, Canada M5S 3G3. Email: jeff@math.toronto.edu. Web: probability.ca/jeff. Supported in part by NSERC of Canada.

Let $h : \mathcal{X} \to \mathbf{R}$ be a fixed, measurable function, with $\pi(h) = 0$ and $\pi(h^2) < \infty$. Let $\gamma_k = \mathbf{E}[h(X_0)\, h(X_k)] = \langle h,\, P^k h\rangle$ be the corresponding lag-k autocovariance. Say that a $\sqrt{n}$-CLT exists for h if $n^{-1/2} \sum_{i=1}^n h(X_i)$ converges weakly to $\text{Normal}(0, \sigma^2)$ for some $\sigma^2 < \infty$. (We allow for the degenerate case $\sigma^2 = 0$.)

We shall also require spectral measures. Let $\mathcal{E}$ be the spectral decomposition measure (e.g. [17], Theorem 12.23) associated with P, so that

$$f(P) \;=\; \int_{-1}^{1} f(\lambda)\; \mathcal{E}(d\lambda)$$

for all bounded analytic functions $f : [-1, 1] \to \mathbf{R}$, and $\mathcal{E}(\mathbf{R}) = \mathcal{E}([-1, 1]) = I$ is the identity operator. Let $\mathcal{E}_h$ be the induced spectral measure for h (cf. [7], p. 1753), viz.

$$\mathcal{E}_h(S) \;=\; \langle h,\; \mathcal{E}(S)h\rangle\,, \qquad S \subseteq [-1, 1] \;\; \text{Borel}$$

with $\mathcal{E}_h(\mathbf{R}) = \langle h,\, \mathcal{E}(\mathbf{R})h\rangle = \langle h,\, h\rangle = \pi(h^2) < \infty$.

3. Expressions for the Variance.

We are interested in the question of whether/when

$$n^{-1/2} \sum_{i=1}^{n} h(X_i) \;\Rightarrow\; \text{Normal}(0, \sigma^2)\,, \qquad \sigma^2 < \infty\,, \tag{1}$$

and the corresponding variance σ^2. (In fact, the convergence in (1) does not require stationarity; see e.g. Proposition 29 of [15].) There are a number of possible formulae for σ^2 in the literature (e.g. [11], [7], [4]), including:

$$A \;=\; \lim_{n\to\infty} n^{-1} \mathbf{Var}\left(\sum_{i=1}^{n} h(X_i)\right) \;=\; \gamma_0 + 2 \lim_{n\to\infty} \sum_{k=1}^{n-1} \Big(\frac{n-k}{n}\Big)\, \gamma_k\,;$$

$$B \;=\; \gamma_0 + 2 \sum_{k=1}^{\infty} \gamma_k \;=\; \gamma_0 + 2 \lim_{n\to\infty} \sum_{k=1}^{n} \gamma_k\,;$$

$$C \;=\; \int_{-1}^{1} \frac{1+\lambda}{1-\lambda}\; \mathcal{E}_h(d\lambda)\,.$$

We shall also have occasion to consider versions of A and B where the limit is taken over odd integers only:

$$A' \;=\; \lim_{j\to\infty} (2j+1)^{-1} \mathbf{Var}\left(\sum_{i=1}^{2j+1} h(X_i)\right)\,;$$

$$B' \;=\; \gamma_0 + 2 \lim_{j\to\infty} \sum_{k=1}^{2j+1} \gamma_k\,.$$

Obviously, $A' = A$ and $B' = B$ provided the limits in A and B exist. But it may be possible that, say A' is well-defined even though A is not.

The following result is implicit in some earlier works (e.g. [11], [7], [4]), but does not appear to have previously been written down precisely.

Theorem 1. *If P is reversible and ergodic, then $A = B = C$ (though they may all be infinite).*

Theorem 1 is proved in Section 5. We first note that the assumption of ergodicity cannot be omitted:

Example 2. Let $\mathcal{X} = \{-1, 1\}$, with $\pi\{-1\} = \pi\{1\} = 1/2$, and $P(1, \{-1\}) = P(-1, \{1\}) = 1$, so P is reversible with respect to $\pi(\cdot)$. Let h be the identity function. Then $\left|\sum_{i=1}^{n} h(X_i)\right| \leq 1$, so $A = 0$. On the other hand, $\gamma_k = (-1)^k$, so $B' = 0$ but B is an oscillating sum and thus undefined. So $A \neq B$, but Theorem 1 is not violated since the chain is periodic and hence not ergodic. And, a (degenerate) $\sqrt{n}$-CLT does hold, with $\sigma^2 = A = B' = 0$.

Now, Kipnis and Varadhan [11] proved for reversible chains that if $C < \infty$, then a CLT exists for h, with $\sigma^2 = C$. Combining this with Theorem 1, we have:

Corollary 3. *If P is reversible and ergodic, and any one of A, B, and C is finite, then a CLT exists for h, with $\sigma^2 = A = B = C < \infty$.*

Also, Roberts [13], considered the quantity $r(x) = \mathbf{P}[X_1 = x \mid X_0 = x]$, the probability of remaining at x, which is usually positive for Metropolis-Hastings algorithms. He proved that if $\lim_{n\to\infty} n\, \mathbf{E}[h^2(X_0)\, r(X_0)^n] = \infty$, then $A = \infty$ (and used this to prove that $A = \infty$ for some specific independence sampler examples). Combining his result with Theorem 1, we have:

Corollary 4. *If P is reversible and ergodic, and if $\lim_{n\to\infty} n\, \mathbf{E}[h^2(X_0)\, r(X_0)^n] = \infty$, then $A = B = C = \infty$.*

If the Markov chain is not reversible, then the spectral measure required to define C becomes much more complicated, and we do not pursue that here. However, it is still possible to compare A and B. It follows immediately from the definitions and the dominated convergence theorem (cf. [2], p. 172; [4]) that if $\sum_k |\gamma_k| < \infty$, then $A = B < \infty$ (though this might not imply a $\sqrt{n}$-CLT for h). The condition $\sum_k |\gamma_k| < \infty$ is known to hold for uniformly ergodic chains (see [2]), and for reversible geometric chains (since that implies [14] that $|\gamma_k| \leq \rho^k \, \pi(h^2)$ for some $\rho < 1$), but it does not hold in general. This leads to the following question:

Open Problem #1. *If the Markov chain is ergodic, but not necessarily reversible or geometrically ergodic or uniformly ergodic, does it necessarily follow that $A = B$ (allowing that they may both be infinite)?*

4. Converse: CLT Necessity.

The result from [11] raises the question of the *converse*. Suppose $n^{-1} \sum_{i=1}^{n} h(X_i)$ converges weakly to Normal$(0, \sigma^2)$ for some $\sigma^2 < \infty$. Does it necessarily follow that any of A, B, and C are finite? In particular, an affirmative answer to this question would allow a strengthening of Corollary 4 to conclude that no $\sqrt{n}$-CLT holds for such h, and in particular a $\sqrt{n}$-CLT does not hold for the independence sampler examples considered by Roberts [13].

Even in the i.i.d. case (where $P(x, A) = \pi(A)$ for all $x \in \mathcal{X}$ and $A \in \mathcal{F}$), this question is non-trivial. However, classical results (cf. Sections IX.8 and XVII.5 of Feller [6]; for related results see e.g. [3], [1]) provide an affirmative answer in this case:

Theorem 5. *The converse to the result in [11] holds in the i.i.d. case. That is, if $\{X_i\}$ are i.i.d., and $n^{-1/2} \sum_{i=1}^{n} h(X_i)$ converges weakly to Normal$(0, \sigma^2)$ for some $\sigma^2 < \infty$, then A, B, and C are all finite, and $\sigma^2 = A = B = C$.*

Proof. Let $Y_i = h(X_i)$, and let $U(z) = \mathbf{E}[Y_1^2 \, \mathbf{1}_{|Y_1| \leq z}]$. Then since the $\{Y_i\}$ are i.i.d. with mean 0, Theorem 1a on p. 313 of [6] says that there are positive sequences $\{a_n\}$ with $a_n^{-1}(Y_1 + \ldots + Y_n) \Rightarrow$ Normal$(0, 1)$ if and only if $\lim_{z \to \infty} [U(sz)/U(z)] = 1$ for all $s > 0$. Furthermore, equation (8.12) on p. 314 of [6] (see also equation (5.23) on p. 579 of [6]) says that in this case,

$$\lim_{n \to \infty} n \, a_n^{-2} \, U(a_n) \;=\; 1 \,. \tag{2}$$

Now, the hypotheses imply that $a_n^{-1}(Y_1 + \ldots + Y_n) \Rightarrow$ Normal$(0, 1)$ where $a_n = c \, n^{1/2}$ with $c = \sigma^{-1} > 0$. Thus, from (2), we have $\lim_{n \to \infty} c \, U(cn^{1/2}) = 1$. It follows that $\lim_{z \to \infty} U(z) = c^{-1} < \infty$, i.e. $\mathbf{E}(Y_1^2) < \infty$. We then compute that $\gamma_k = 0$ for $k \geq 1$, so $B = \gamma_0 = \mathbf{E}(Y_1^2) = \sigma^2 < \infty$. Hence, by Corollary 3, $\sigma^2 = A = B = C = \mathbf{E}(Y_1^2) < \infty$. ∎

Remark 6. In the above proof, if $\mathbf{E}(Y_1^2) = \sigma^2 < \infty$, then of course $U(z) \to \sigma^2$, so $U(sz)/U(z) \to \sigma^2/\sigma^2 = 1$, and the (classical) CLT applies. On the other hand, there are many distributions for the $\{Y_i\}$ which have infinite variance, but for which the corresponding U is still slowly varying in this sense. Examples include the density function $y^{-3}\mathbf{1}_{|y|\geq 1}$, and the cumulative distribution function $1-(1+y)^{-2}$ for $y \geq 0$. The results from [6] say that we cannot have $a_n = c\, n^{1/2}$ in such cases. (In the $y^{-3}\mathbf{1}_{|y|\geq 1}$ example, we instead have $a_n = c\,(n \log n)^{-1/2}$.)

Theorem 5 is specific to the i.i.d. case, leading to the following question:

Open Problem #2. *Does Theorem 5 hold in the non-i.i.d. case, i.e. when $\{X_n\}$ is assumed only to be a reversible stationary ergodic Markov chain? (And, to what extent are reversibility and ergodicity necessary?)*

There are various results in the stationary process literature (e.g. [9], [12]) that are somewhat related to those from [6] used in the proof of Theorem 5, but their applicability to Open Problem #2 is unclear. Also, if $\{n^{-1}\sum_{i=1}^{n} h(X_i)^2\}$ is *uniformly integrable*, then whenever a $\sqrt{n}$-CLT exists we must have $A = \sigma^2$, which implies by Theorem 1 (assuming reversibility) that $\sigma^2 = A = B = C < \infty$, but it is not clear when this uniform integrability condition will be satisfied. Finally, we note that Open Problem #2 is specific to the $n^{-1/2}$ normalisation and the Normal limiting distribution; other normalisations and limiting distributions may sometimes hold, but we do not consider them here.

5. Proof of Theorem 1.

Theorem 1 follows from Corollary 10 and Proposition 11 below. We begin with a lemma (somewhat similar to Theorem 3.1 of [7]).

Lemma 7. *If P is reversible, then $\gamma_{2i} \geq 0$, and $|\gamma_{2i+1}| \leq \gamma_{2i}$, and $|\gamma_{2i+2}| \leq \gamma_{2i}$.*

Proof. By reversibility, $\gamma_{2i} = \langle f, P^{2i} f\rangle = \langle P^i f, P^i f\rangle = \|P^i f\|^2 \geq 0$.

Also, $|\gamma_{2i+1}| = \langle f, P^{2i+1} f\rangle = |\langle P^i f, P(P^i f)\rangle| \leq \|P^i f\|^2 \|P\| \leq \|P^i f\|^2 = \gamma_{2i}$.

Similarly, $|\gamma_{2i+2}| = \langle f, P^{2i+2} f\rangle = |\langle P^i f, P^2(P^i f)\rangle| \leq \|P^i f\|^2 \|P^2\| \leq \|P^i f\|^2 = \gamma_{2i}$. ∎

Lemma 8. *If P is reversible and ergodic, then $\lim_{k\to\infty} \gamma_k = 0$.*

Proof. Since P is ergodic, it does not have an eigenvalue 1 or -1. Hence (cf. [17], Theorem 12.29(b)) its spectral measure $\mathcal{E}$ does not have an atom at 1 or -1, i.e. $\mathcal{E}(\{-1,1\}) = 0$, so also $\mathcal{E}_h(\{-1,1\}) = 0$ (cf. [7], Lemma 5). Hence, by the dominated convergence theorem (since $|\lambda^k| \leq 1$ for $-1 \leq \lambda \leq 1$, and $\int_{-1}^{1} 1\ \mathcal{E}_h(d\lambda) = \pi(h^2) < \infty$), we have:

$$\lim_{k\to\infty} \gamma_k \ = \ \lim_{k\to\infty} \langle h,\, P^k h\rangle \ = \ \lim_{k\to\infty} \int_{-1}^{1} \lambda^k\ \mathcal{E}_h(d\lambda)$$
$$= \ \int_{-1}^{1} \left(\lim_{k\to\infty} \lambda^k\right) \mathcal{E}_h(d\lambda) \ = \ \int_{-1}^{1} 0\ \mathcal{E}_h(d\lambda) \ = \ 0\,. \qquad \blacksquare$$

Proposition 9. *If P is reversible and ergodic, then $A' = B'$. (We allow for the possibility that $A' = B' = \infty$.)*

Proof. We have that

$$(2j+1)^{-1}\,\mathbf{Var}\left(\sum_{i=1}^{2j+1} h(X_i)\right) \ = \ \gamma_0 + 2\gamma_1 + 2\sum_{i=1}^{j}\left(\frac{2j+1-2i}{2j+1}\,\gamma_{2i} + \frac{2j+1-2i-1}{2j+1}\,\gamma_{2i+1}\right)$$
$$= \ \gamma_0 + 2\gamma_1 + 2\sum_{i=1}^{j}\frac{\gamma_{2i}}{2j+1} + 2\sum_{i=1}^{j}\frac{2j+1-2i-1}{2j+1}\,(\gamma_{2i}+\gamma_{2i+1})\,. \qquad (3)$$

By Lemma 7, $\gamma_{2i} + \gamma_{2i+1} \geq 0$, so as $j \to \infty$, for fixed i,

$$\frac{2j+1-2i-1}{2j+1}\,(\gamma_{2i}+\gamma_{2i+1}) \ \nearrow \ \gamma_{2i}+\gamma_{2i+1}\,,$$

i.e. the convergence is *monotonic*. Hence, by the monotone convergence theorem,

$$\lim_{j\to\infty} 2\sum_{i=1}^{j}\frac{2j+1-2i-1}{2j+1}\,(\gamma_{2i}+\gamma_{2i+1}) \ = \ \lim_{j\to\infty} 2\sum_{i=1}^{j}(\gamma_{2i}+\gamma_{2i+1})\,.$$

By Lemma 8, $\gamma_{2i} \to 0$ as $i \to \infty$, so $\sum_{i=1}^{j}\frac{\gamma_{2i}}{2j+1} \to 0$ as $j \to \infty$. Putting this all together, we conclude from (3) that

$$\lim_{j\to\infty}\ (2j+1)^{-1}\,\mathbf{Var}\left(\sum_{i=1}^{2j+1} h(X_i)\right) \ = \ \gamma_0 \ + \ 2\gamma_1 \ + \ 2\ \lim_{j\to\infty}\ \sum_{i=1}^{j}(\gamma_{2i}+\gamma_{2i+1})\,,$$

i.e. $A' = B'$. $\blacksquare$

Corollary 10. *If P is reversible and ergodic, then $A = B$. (We allow for the possibility that $A = B = \infty$.)*

Proof. If P is ergodic, then by Lemma 8, $\gamma_k \to 0$, so $B = B'$. Also,

$$(n+1)^{-1}\mathbf{Var}\left(\sum_{i=1}^{n+1} h(X_i)\right) \; - \; n^{-1}\mathbf{Var}\left(\sum_{i=1}^{n} h(X_i)\right) \tag{4}$$

$$= \; n^{-1}\left[\mathbf{Var}\left(\sum_{i=1}^{n+1} h(X_i)\right) \; - \; \mathbf{Var}\left(\sum_{i=1}^{n} h(X_i)\right)\right] \; + \; [n(n+1)]^{-1}\mathbf{Var}\left(\sum_{i=1}^{n+1} h(X_i)\right)$$

Now, the first term above is equal to $n^{-1}\sum_{i=1}^{n}\gamma_i$ (which goes to 0 since $\gamma_k \to 0$), plus $n^{-1}\mathbf{E}[h^2(X_{i+1})]$ (which goes to 0 since $\pi(h^2) < \infty$). The second term is equal to

$$\frac{\gamma_0}{n(n+1)} \; + \; 2\sum_{k=1}^{n-1} \frac{n-k}{n^2(n+1)}\,\gamma_k$$

which is $O(1/n)$ and hence also goes to 0. We conclude that the difference in (4) goes to 0 as $n \to \infty$, so that $A = A'$. Hence, by Proposition 9, $A = A' = B' = B$. ∎

Proposition 11. *If P is reversible and ergodic, then $B = C$. (We allow for the possibility that $B = C = \infty$.)*

Proof. We compute that:

$$\begin{aligned}
B \; &= \; \lim_{k\to\infty}\Big(\langle h, h\rangle + 2\,\langle h,\, Ph\rangle + 2\,\langle h,\, P^2 h\rangle + \ldots + 2\,\langle h,\, P^k h\rangle\Big) \\
&= \; \lim_{k\to\infty}\Big\langle h,\; (I + 2P + 2P^2 + \ldots + 2P^k)h\Big\rangle \\
&= \; \lim_{k\to\infty}\int_{-1}^{1} (1 + 2\lambda + 2\lambda^2 + \ldots + 2\lambda^k)\; \mathcal{E}_h(d\lambda) \\
&= \; \lim_{k\to\infty}\int_{-1}^{1} (2\frac{1-\lambda^{k+1}}{1-\lambda} - 1)\; \mathcal{E}_h(d\lambda) \\
&= \; \lim_{k\to\infty}\int_{-1}^{1} (\frac{1+\lambda-\lambda^{k+1}}{1-\lambda})\; \mathcal{E}_h(d\lambda) \\
&= \; C \; - \; \lim_{k\to\infty}\int_{-1}^{1} \lambda^{k+1}\; \mathcal{E}_h(d\lambda) \\
&= \; C\,,
\end{aligned}$$

where the final equality follows by dominated convergence as in the proof of Lemma 8. ∎

Acknowledgements. I thank N. Madras and T. Salisbury for suggesting the i.i.d. case in Theorem 5; G.L. Jones for many useful comments including mentioning the results in [2] and [12]; and C.J. Geyer, J.P. Hobert, and G.O. Roberts for helpful discussions.

References

[1] S.K. Basu (1984), A local limit theorem for attraction to the standard normal law: The case of infinite variance. Metrika **31**, 245–252. Available at http://www.springerlink.com/content/v616742404102u13/

[2] P. Billingsley (1968), Convergence of Probability Measures. Wiley, New York.

[3] M. Breth, J. S. Maritz, and E. J. Williams (1978), On Distribution-Free Lower Confidence Limits for the Mean of a Nonnegative Random Variable. Biometrika **65**, 529–534.

[4] K.S. Chan and C.J. Geyer (1994), Discussion paper. Ann. Stat. **22**, 1747–1758.

[5] X. Chen (1999). Limit theorems for functionals of ergodic Markov chains with general state space. Mem. Amer. Math. Soc. **139**.

[6] W. Feller (1971), An introduction to Probability Theory and its applications, Vol. *II*, 2nd ed. Wiley & Sons, New York.

[7] C.J. Geyer (1992), Practical Markov chain Monte Carlo. Stat. Sci., Vol. **7**, No. **4**, 473-483.

[8] J.P. Hobert, G.L. Jones, B. Presnell, and J.S. Rosenthal (2002), On the Applicability of Regenerative Simulation in Markov Chain Monte Carlo. *Biometrika* **89**, 731–743.

[9] I.A. Ibragimov and Y.V. Linnik (1971), Independent and Stationary Sequences of Random Variables. Wolters-Noordhoff, Groningen (English translation).

[10] G.L. Jones (2004), On the Markov chain central limit theorem. Prob. Surveys **1**, 299–320.

[11] C. Kipnis and S.R.S. Varadhan (1986), Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. Comm. Math. Phys. **104**, 1-19.

[12] M. Peligrad (1986), Recent advances in the central limit theorem and its weak invariance principle for mixing sequences of random variables (a survey). In *Dependence in Probability and Statistics: A Survey of Recent Results*, E. Eberlein and M.S. Taqqu, eds., Birkhauser, Cambridge, Mass., pp. 193–223.

[13] G.O. Roberts (1999), A note on acceptance rate criteria for CLTs for Metropolis-Hastings algorithms. J. Appl. Prob. **36**, 1210–1217.

[14] G.O. Roberts and J.S. Rosenthal (1997), Geometric ergodicity and hybrid Markov chains. Electronic Comm. Prob. **2**, Paper no. 2, 13–25.

[15] G.O. Roberts and J.S. Rosenthal (2004), General state space Markov chains and MCMC algorithms. Prob. Surveys **1**, 20–71.

[16] J.S. Rosenthal (2001), A review of asymptotic convergence for general state space Markov chains. Far East J. Theor. Stat. **5**, 37–50.

[17] W. Rudin (1991), Functional Analysis, 2nd ed. McGraw-Hill, New York.

[18] L. Tierney (1994), Markov chains for exploring posterior distributions (with discussion). Ann. Stat. **22**, 1701–1762.