Notes About Markov Chain CLTs

[Rough notes by Jeffrey S. Rosenthal, February 2007, based on very helpful conversations with J.P. Hobert, N. Madras, G.O. Roberts, and T. Salisbury. For discussion and clarification only – not for publication. Comments appreciated.]

1. Introduction.

These notes concern various issues surrounding central limit theorems (CLTs) for Markov chains, important notably for MCMC algorithms. A number of other papers have discussed related matters ([8], [13], [5], [3], [6], [7]), and probably much of the discussion below is already known, but we wanted to write it up for our own clarification.

Let $\pi(\cdot)$ be a probability measure on a measurable space $(\mathcal{X}, \mathcal{F})$. Let P be a Markov chain operator reversible with respect to $\pi(\cdot)$. Write $\langle f, g\rangle = \int_{\mathcal{X}} f(x)\, g(x)\, \pi(dx)$; by reversibility, $\langle f, Pg\rangle = \langle Pf, g\rangle$.

Let $h : \mathcal{X} \to \mathbf{R}$ be measurable, with $\pi(h^2) < \infty$ and (say) $\pi(h) = 0$. Let $\{X_n\}_{n=0}^{\infty}$ follow the transitions P in stationarity, so $\mathcal{L}(X_n) = \pi(\cdot)$ and $\mathbf{P}[X_{n+1} \in A \,|\, X_n] = P(X_n, A)$ for all $A \in \mathcal{F}$, for $n = 0, 1, 2, \ldots$. Let $\gamma_k = \mathbf{E}[h(X_0)\, h(X_k)] = \langle h,\, P^k h\rangle$. Let $r(x) = \mathbf{P}[X_1 = x \,|\, X_0 = x]$ for $x \in \mathcal{X}$. Let $\mathcal{E}$ be the spectral measure (e.g. [12]) associated with P, so that

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

for "all" analytic functions $f : \mathbf{R} \to \mathbf{R}$, and also $\mathcal{E}(\mathbf{R}) = I$. Let $\mathcal{E}_h$ be the induced measure for h, viz.

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

a positive Borel measure (cf. [5], p. 1753), which is finite if $\pi(h^2) < \infty$ since then $\mathcal{E}_h(\mathbf{R}) = \langle h,\, \mathcal{E}(\mathbf{R})h\rangle = \langle h,\, h\rangle = \pi(h^2) < \infty$.

We are interested in the question of whether/when a root-n CLT exists for h, meaning that $n^{-1/2} \sum_{i=1}^{n} h(X_i)$ converges weakly to $\text{Normal}(0, \sigma^2)$ for some $\sigma^2 < \infty$.

2. Representations of the Variance.

There are a number of possible formulae for σ^2 in the literature (e.g. [8], [5], [3]), including:

$$A \;=\; \lim_{n\to\infty} n^{-1} \mathbf{Var}\left(\sum_{i=1}^{n} h(X_i)\right)\;;$$

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

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

It is proved in [8] that if $C<\infty$, then a CLT exists for h (with $\sigma^2=C$). And, it is proved in [9] that if $\lim_{n\to\infty} n\,\mathbf{E}[h^2(X_0)\,r(X_0)^n]=\infty$, then $A=\infty$. So, it seems important to sort out the relationship between A, B, and C. It is various implied (e.g. [5]) that A, B, and C are usually all equivalent, and here we consider conditions which make that true.

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

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

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

Obviously, $A'=A$ and $B'=B$ provided the limits in A and B exist. But it may be possible that A' and/or B' are well-defined even if A and/or B are not.

We begin with a lemma (somewhat similar to Theorem 3.1 of [5]).

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

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

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

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

To continue, recall that P is *ergodic* if $\lim_{n\to\infty}\sup_{A\in\mathcal{F}}|P^k(x,A)-\pi(A)|=0$ for π-a.e. $x\in\mathcal{X}$. This follows (cf. [13], [11], [10]) if P is ϕ-irreducible and aperiodic.

Lemma 2. *If P is reversible and ergodic, then $\lim_{k\to\infty}\gamma_k=0$.*

Proof. Since P is ergodic, 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. [5], Lemma 5). Hence, by dominated convergence (since $|\lambda^k|\le 1$, and $\int 1\,\mathcal{E}_h(d\lambda)=\pi(h^2)<\infty$), we have:

$$\lim_{k\to\infty}\gamma_k \;=\; \lim_{k\to\infty}\langle h,\,P^kh\rangle \;=\; \lim_{k\to\infty}\int_{-1}^{1}\lambda^k\,\mathcal{E}_h(d\lambda)$$

$$=\; \int_{-1}^{1}\left(\lim_{k\to\infty}\lambda^k\right)\mathcal{E}_h(d\lambda) \;=\; \int_{-1}^{1}0\,\mathcal{E}_h(d\lambda) \;=\; 0\,. \quad ∎$$

Proposition 3. *If P is reversible and ergodic, then $A' = B'$. (We allow for the possibility that $A' = B' = \infty$.)*

Proof. We compute directly (by expanding the square) that

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

Hence,

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

$$=\; \gamma_0 + 2\gamma_1 + 2\sum_{i=1}^{j}\frac{\gamma_{2i}}{2j+1} + 2\sum_{i=1}^{j}\frac{2j+1-2i-1}{2j+1}\,(\gamma_{2i}+\gamma_{2i+1})\,.$$

By Lemma 1, $\gamma_{2i}+\gamma_{2i+1} \geq 0$, 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,

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

By Lemma 2, $\gamma_{2i} \to 0$ 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 that

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

i.e. $A' = B'$, Q.E.D. ∎

Corollary 4. *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 2, $\gamma_k \to 0$, so $B = B'$. Also,

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

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

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 also goes to 0. We conclude that the difference in (1) goes to 0 as $n \to \infty$, so that $A = A'$. Hence, by Proposition 3, $A = A' = B' = B$. ∎

Remark 5. If $\gamma_{2i} \not\to 0$, then since $\gamma_{2i+2} \leq \gamma_{2i}$ by Lemma 1, we must have $\sum_{i=1}^{\infty}\gamma_{2i} = \infty$. But is it possible that, say, $\gamma_{2i} = 1/i$ and $\gamma_{2i+1} = -1/i$ for all large i, so that B' is finite, but A' is infinite?

Proposition 6. *If P is reversible and ergodic, then $B = C$. (We allow for the possibility that $B = C = \infty$.)*

Proof. We compute (recalling that $\mathcal{E}_h(\{-1,1\}) = 0$) that:

$$B = \lim_{k\to\infty}\Big(\langle h,h\rangle + 2\langle h,\,Ph\rangle + 2\langle h,\,P^2h\rangle + \ldots + 2\langle h,\,P^kh\rangle\Big)$$

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

where the penultimate equality is justified by the monotone convergence theorem, since

$$\{\frac{1+\lambda-\lambda^{k+1}}{1-\lambda}\} \nearrow \frac{1+\lambda}{1-\lambda}\,, \qquad k \to \infty$$

whenever $-1 < \lambda < 1$. ∎

Remark. The above use of the monotone convergence theorem is somewhat subtle, in that the monotonicity is *not* on the original random variables, only for the λ's with respect to the spectral measure.

Corollary 7. *If P is reversible and ergodic, then $A = B = C$ (though they may all be infinite).*

Using the result from [8], we have:

Corollary 8. *If P is reversible and ergodic, and any one of A, B, and C is finite, then a CLT exists for h (with $\sigma^2 = A = B = C$).*

Using the result from [9], we have:

Corollary 9. *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, and C are all infinite.*

3. Converse: CLT Necessity.

The result from [8] 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?

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 appears to be a non-trivial question. However, Sections IX.8 and XVII.5 of Feller [4] appear to resolve the issue, as we now discuss. (For related comments see e.g. [2], [1].)

Theorem 1a on p. 313 of [4] says that a distribution belongs to the domain of attraction of the normal distribution if and only if its truncated variance is slowly varying. More precisely, letting $U(z) = \mathbf{E}[X_1^2 I_{|X_1|\le z}]$, the theorem says that in the i.i.d. case, there are sequences $\{a_n\}$ and $\{b_n\}$ with $a_n^{-1}(X_1 + \ldots + X_n) \Rightarrow N(0,1)$ if and only if $\lim_{z\to\infty} [U(sz)/U(z)] = 1$ for all $s > 0$.

Now, if $\mathbf{E}(X_1^2) = \sigma^2 < \infty$, 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 other distributions which have infinite variance, but for which U is slowly varying as above. Examples include the density function $x^{-3}\mathbf{1}_{|x|\ge 1}$, and the cumulative distribution function $1 - (1+x)^{-2}$ for $x \ge 0$. The result in [4] says that in such cases we still have $a_n^{-1}(X_1 + \ldots + X_n) \Rightarrow N(0,1)$, but the question is whether we could perhaps still have $a_n = c\, n^{1/2}$ even if the variance is infinite.

It appears the answer is no. Specifically, equation (8.12) on p. 314 of [4] (see also equation (5.23) on p. 579 of [4]) says that in such cases, we can always arrange that

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

If we did have $a_n = c\,n^{1/2}$, then this would imply that $\lim_{n\to\infty} cU(cn^{1/2}) = 1$, i.e. that $\lim_{z\to\infty} U(z) < \infty$, i.e. that the variance is finite. (In examples like $x^{-3}\mathbf{1}_{|x|\geq 1}$ we would have something like $a_n = (n \log n)^{-1/2}$ instead.) So, this appears to prove:

Proposition 10. *The converse to the result in [8] 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 $\sigma^2 = A = B = C$.*

Meanwhile, the non-i.i.d. case appears to still be open.

4. Possible Open Questions.

I would appreciate clarification about any of the following questions. Are they known? trivial? interesting? etc.

How much of the above carries over if P is not ergodic, and $\gamma_k \not\to 0$? (See Remark 5.) Do we still always have $A' = B'$ (even though A' and B' may be undefined)? And, could it be that, say, A is defined even though B is not?

How much of the above carries over if $\pi(h^2) = \infty$? Does the spectral measure $\mathcal{E}_h$ still make sense then? Are A and B both necessarily equal to $+\infty$ in this case?

And, most importantly: does Proposition 10 hold in the non-i.i.d. case, i.e. for general reversible Markov chains?

In a different direction, does any of the above carry over to the case where P is not reversible? (Even to the case where $P = P_1 P_2$ where each P_i is reversible?)

Also, I think most of the results presented in Sections 2 and 3 above are already known in some form. But were they previous written down and proved somewhere? If so, where?

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] 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.

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

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

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

[6] 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.

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

[8] 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.

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

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

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

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

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