Remarks on Nonstandard CLTs for Markov chains

Corresponding author:
Bernardo B. de Andrade
Departamento de Economia,
Universidade de Brasília,
Campus Asa Norte,
Brasília, DF 70910-900, Brazil.
bernardo-andrade@ouhsc.edu

Charles J. Geyer
School of Statistics,
University of Minnesota,
224 Church Street,
Minneapolis, MN 55455, USA.
charlie@stat.umn.edu

February, 2010

Abstract

This article summarizes recently proved nonstandard central limit theorems for Markov chains under the "radically elementary" probability theory (REPT) developed by Nelson [22], suggesting a wide front for future research.

Keywords: Markov chains, nonstandard analysis, radically elementary probability.

1 Introduction

Markov chain Monte Carlo (MCMC) is the primary method of simulating samples from distributions not amenable to direct sampling. It is a necessary *tour de force* in all but the simplest Bayesian applications [10, 26] and also a useful technique in complicated maximum likelihood problems [13]. MCMC theory relies on the heavily measure-theoretic and functional-analytic apparatuses used in the limit theory of Markov chains on general state spaces [21]. The nonstandard tools we use in this paper allow for a much simpler formulation of the Markov chain limit theory that supports MCMC methodology.

Nelson introduced a formalism for probability theory which he called "radically elementary". In this theory all sample spaces are finite and all probability models assign positive probability to every point in the sample space, so expectations always exist, being given by finite sums, and conditional expectations always exist and are given by ratios of finite sums. These restrictions entail no loss of generality when used in conjunction with nonstandard analysis, which is the rigorous use of infinitesimals. Nelson goes from axioms for nonstandard analysis to the law of large numbers, the martingale convergence theorem, the central limit theorem, and Brownian motion in only eighty pages with complete proofs and no lack of rigor.

To show how Nelson's theory works, we give a brief discussion of the De Moivre–Laplace central limit theorem in the style of Nelson. Suppose X is a binomial random variable with sample size n and success probability p. Suppose n is unlimited (which means $1/n$ is infinitesimal) and $p(1-p)$ is non-infinitesimal. Then the random variable $Z = (X - np)/\sqrt{np(1-p)}$ is nearly standard normal in the sense that its distribution function differs from that of a conventional standard normal random variable by only an infinitesimal amount at any point [11, Theorem 8.2, Corollaries 8.3 and 10.6]. We see two differences from conventional theory. First, n does not go to infinity (an infinite sequence of random variables requires an infinite sample space, which Nelson's theory does not allow). Second, although Nelson's theory does not allow the continuous random variables of conventional theory, such as standard normal, it does have random variables, such as the Z described above, that serve the same purposes.

Later we will present a Markov chain CLT in the style of Nelson. Let X_1, X_2, $\ldots$ X_n, be a stationary Markov chain and

$$\bar{f} = \frac{1}{n} \sum_{i=1}^{n} f(X_i).$$

Then under certain conditions given below $[\bar{f} - \mathbb{E}(\bar{f})]/\operatorname{Var}(\bar{f})$ is nearly standard normal in the sense that when n is unlimited its distribution function differs from that of a conventional standard normal random variable by only an infinitesimal amount at any point.

So why redo this known mathematics in "radically elementary" theory? One reason is to do some interesting new mathematics, but another important reason is pedagogical. Our experience is that many users of Markov chains, especially statisticians using MCMC, would benefit from a non–measure-theoretic treatment of Markov chains. Such treatments exist, but they cover only discrete state space chains. Since most applications of MCMC are general state space Markov chains — the state at time t is a continuous random vector — the appropriate theory can only be found in books such as [23] and [21], which are heavily measure-theoretic. Similarly, many MCMC papers, such as [29] have more measure theory than many researchers in statistics, engineering, physics and computer science are interested in. So there is a pedagogical point to a Markov chain central limit theorem that does not involve measure theory and defines Markov chains in terms of transition probability matrices. It is also true that the actual Markov chains involved in MCMC do have finite state space, although in theoretical discussion we pretend that the computer's "real" numbers are the analyst's real numbers so the state space is a continuum. As says [22, p. 13]

> The conventional approach involves an idealization, because one cannot actually complete an infinite number of observations. The second approach also involves an idealization, because one cannot actually complete a non-standard number of observations. In fact, it is the nature of mathematics to deal with idealizations. The choice of a formalism must be based on aesthetic considerations, such as directness of expression, simplicity, and power.

We add that the conventional approach involves other idealizations: continuous random variables do not actually exist because measurements can be made to only a finite number of decimal places. In the "radically elementary" approach we do not have continuous random variables, but do allow random variables whose possible values have only infinitesimal separation, and this too is an idealization. We would also add pedagogical and historical considerations to the aesthetic ones Nelson mentions.

Finally we should warn against one tempting misconception. The point of most work on nonstandard analysis is to prove theorems of conventional mathematics using nonstandard methods. The statements of the theorems are entirely conventional nowhere mentioning infinitesimals or other nonstandard concepts. Only in the proofs are infinitesimals used. This includes most work on nonstandard probability theory such as [18]. Nelson [22] and our work are different in that theorem statements involve nonstandard concepts and the objects described by the theorems are nonstandard, e.g., a binomial distribution with nonstandard sample size. Thus analogies between theory in the style of Nelson and conventional probability theory are not exact. The two theories are different in mathematical content, although much the same purposes are served. For one thing, countable additivity is vacuous when sample spaces are finite, so Nelson is in some ways more analogous to finitely additive probability theory ([9, Sections III.1–III.3], [8] and [2]) than to countably additive theory. It also bears some resemblance to constructive mathematics [4] in that theorems of conventional mathematics that are not constructively valid, such as the martingale convergence theorem and the Birkhoff ergodic theorem, are also not valid in "radically elementary" probability theory without additional assumptions.

We use the following notation. When x and y are real numbers, $x \simeq y$ means $x - y$ is infinitesimal, $x \ll y$ means $x \le y$ and not $x \simeq y$, and $x \lesssim y$ means $x \le y$ or $x \simeq y$. When x is a real number, $x \simeq \infty$ means x is positive and unlimited (i. e., $1/x$ is infinitesimal) and $x \ll \infty$ means not $x \simeq \infty$. When $y \neq 0$, we use $x \sim y$ to mean $x/y \simeq 1$. For the most part we use only the most rudimentary concepts of nonstandard analysis, such as the sum of a limited number of infinitesimals is infinitesimal. Readers wanting complete background should consult [22, Chapters 4–6] and [11, Part I].

The only counterintuitive principle of nonstandard analysis that we use is overspill. Set formation, $\{\, x \in A : B(x) \,\}$, where A is a set and $B(x)$ is a predicate, is allowed only when the predicate is from conventional mathematics, i.e., does not involve the notions infinitesimal, unlimited, and other concepts derived from them, in which case it is called *internal*. Sets $\{\, x \in A : B(x) \,\}$, where B is internal are called *internal sets* in nonstandard analysis, and in [22] Nelson allows only such sets as mathematical objects. It is provable ([22, Chapter 5]; [11, Section 3.4]) that the set of all infinitesimals does not exist and similarly for the set of unlimited numbers. This has the consequence that there is no least upper bound for the infinitesimals (which would exist by order completeness of the real numbers if there existed a set containing all the infinitesimals and nothing else). This has the further consequence that if $B(x)$ is an internal property that holds for all infinitesimal x, then it must also hold for some non-infinitesimal x (called "overspill"), or if $B(x)$ is an internal property that holds for all unlimited x, then it must also hold for some limited x.

The analog of convergence in distribution in "radically elementary" probability theory is near equivalence of random variables, which has already be alluded to above. Random variables X and Y are *nearly equivalent* [22, Chapter 17] if for every limited nearly continuous function $\mathbb{E}h(X) \simeq \mathbb{E}h(Y)$ where limited means every value is limited and nearly continuous means $x \simeq y$ implies $h(x) \simeq h(y)$. A random variable X is limited almost surely if for every $\varepsilon \gg 0$ there is an $M_\varepsilon \ll \infty$ such that $\Pr(|X| \leq M_\varepsilon) \geq 1 - \varepsilon$. Random variables that are limited almost surely are nearly equivalent if and only if the Lévy distance between their distributions is infinitesimal [11, Chapter 10]. Also, random variables that are *limited almost surely* are nearly equivalent if and only if their characteristic functions are nearly equal for all limited values of the argument [11, Chapter 11]. We finalize by noting that Nelson's *definition of* L_p is quite different from the conventional theory. A random variable X is L_1 if

$$\mathbb{E}\left\{|X|\mathbb{I}_{\{|X|>a\}}\right\} \simeq 0,$$

for all unlimited a. Theorem 8.1 in Nelson [22] provides equivalent conditions for a random variable to be L_1 and is often used as the definition. It also shows that the L_1 property is stronger than limited absolute expectation. X is L_p, $1 < p < \infty$, if $|X|^p$ is L_1.

2 A CLT under α-Mixing

Consider a sequence of random variables $X_1, \ldots, X_\nu$ defined on a finite sample space such that $\nu \simeq \infty$. Let $\sigma(X_i, \ldots X_j)$ denote the algebra generated by $X_i, \ldots, X_j$. Define the α-mixing coefficients (see [5])

$$\alpha_n = \max_{1 \leq k \leq \nu - n} \max_{\substack{A \in \sigma(X_1, \ldots, X_k) \\ B \in \sigma(X_{k+n}, \ldots X_\nu)}} |\Pr(A \cap B) - \Pr(A)\Pr(B)| \tag{2.1}$$

We will be interested in the case where $X_1, \ldots, X_\nu$ are *α-mixing*, also called *strongly mixing*, which means α_n nearly converges to zero. This concept has two characterizations, which are equivalent by overspill: one is $\alpha_n \simeq 0$ whenever $n \simeq \infty$, and the other is for every $\varepsilon \gg 0$ there exists $N \ll \infty$ such that $\alpha_n \leq \varepsilon$ whenever $n \geq N$.

We make the following assumptions throughout this section:

(A.0) $X_1, \ldots, X_\nu$ is *stationary*, meaning for each $i \geq 0$ the joint distribution of $(X_k, \ldots, X_{k+i})$ is the same for all k such that $1 \leq k \leq k + i \leq \nu$.

Moreover, $\mathbb{E}X_k = 0$. Hence $\mathrm{cov}(X_k, X_{k+i}) = \mathbb{E}(X_1 X_{1+i})$ for $1 \le k \le k+i \le \nu$.

(A.1) There exists $M \ll \infty$ such that $n^5 \alpha_n \le M$ for $n \simeq \infty$;

(A.2) $\mathbb{E}X_1^{12} \ll \infty$.

We note that (A.1) implies by overspill that there exists $K \ll \infty$ such that $n^5 \alpha_n \le K$ for $1 \le k \le \nu$. These assumptions parallel those in [3, Theorem 27.4]. The conditions above may be stronger than necessary for a CLT – in conventional theory a more general result for strongly mixing sequences is [15, Theorem 18.5.2]. We conjecture that a nonstandard counterpart to this result would follow similar steps. Jones [16] surveys a number of conventional Markov chain CLTs.

Define $S_n = \sum_{i=1}^n X_i$, $\gamma_i = \mathbb{E}(X_1 X_{1+i})$ and

$$\sigma_n^2 = \frac{1}{n} \mathrm{Var}(S_n) = \mathrm{Var}(X_1) + 2 \sum_{i=1}^{n-1} \frac{n-i}{n} \gamma_i. \tag{2.2}$$

Theorem 2.1. *Let $X_1, \ldots, X_\nu$, with $\nu \simeq \infty$, be a sequence of random variables. Define α-mixing coefficients by* (2.1)*, and assume (A.0), (A.1), and (A.2). Define σ_n^2 by* (2.2)*, and suppose $\sigma_\nu^2 \gg 0$. Then $(X_1 + \cdots + X_\nu)/(\sigma_\nu \sqrt{\nu})$ is nearly standard normal.*

Sketch of proof. Most of the work is in proving that (see [1])

$$\frac{X_1 + \cdots + X_\nu}{\sigma_\nu \sqrt{\nu}} S_\nu$$

is the sum of a nearly standard normal random variable and a limited almost surely random variable. Then application of a radically elementary Slutsky's theorem [11, Corollary 9.11] implies that the sum is nearly standard normal. □

3 A Radically Elementary Markov Chain Limit Theory

Based on Theorem 2.1 we may prove a nonstandard CLT for stationary polynomially ergodic Markov chains (Theorem 3.1). Rosenthal [28] gives an upper bound on the distance to stationarity of a Markov chain. We have [1] combined it with Theorem 3.1 to obtain a CLT under drift and minorization conditions.

Let X_1, X_2, $\dots$ X_n, be a stationary Markov chain and

$$\bar{f} = \frac{1}{n} \sum_{i=1}^{n} f(X_i).$$

Then under certain conditions given below $[\bar{f} - \mathbb{E}(\bar{f})] / \operatorname{Var}(\bar{f})$ is nearly standard normal in the sense that when $n \simeq \infty$ its distribution function differs from that of a conventional standard normal random variable by only an infinitesimal amount at any point.

The conventional measure-theoretic analogs of Theorem 3.1 are well known [7, 14, 17, 19, 20, 21, 23, 30]. These analogs use many different methods. Our method is most analogous to a combination of [7, Theorem 2], which is a corollary of the stationary process CLT [15, Theorem 18.5.2] but using the much simpler stationary process CLT [3, Theorem 27.4].

Let X_1, $\dots$, X_ν, with $\nu \simeq \infty$ be a stationary stochastic process taking values in an arbitrary finite set $\mathcal{S}$, and suppose it is also a Markov chain with transition probabilities given by a matrix P with entries

$$P(x, y) = \Pr(X_{n+1} = y \mid X_n = x), \qquad x, y \in \mathcal{S}, \tag{3.1}$$

and invariant distribution given by a vector π with entries

$$\pi(x) = \Pr(X_n = x), \qquad x \in \mathcal{S}. \tag{3.2}$$

A function $f : \mathcal{S} \to \mathbb{R}$ induces a *functional* of the Markov chain

$$\bar{f} = \frac{1}{\nu} \sum_{i=1}^{\nu} f(X_i). \tag{3.3}$$

Define

$$\mu_f = \mathbb{E}(\bar{f}) \tag{3.4}$$

$$\sigma_f^2 = \nu \operatorname{Var}(\bar{f}) \tag{3.5}$$

We establish conditions under which $\sqrt{\nu}(\bar{f} - \mu_f)/\sigma_f$ is nearly standard normal.

The *total variation distance* between two probability vectors π_1 and π_2 on $\mathcal{S}$ is defined by

$$\|\pi_1 - \pi_2\| = \max_{x \in \mathcal{S}} |\pi_1(x) - \pi_2(x)|$$

Consider the inequality

$$\|P^n(x,\,\cdot\,) - \pi(\,\cdot\,)\| \le M(x) r_n, \tag{3.6}$$

where P^n denotes the n-fold product of the transition probability matrix P, where π is the invariant distribution, where M is a nonnegative function on $\mathcal{S}$, and r_n is a nonnegative decreasing sequence. We say the Markov chain is *geometrically ergodic* if $r_n = r^n$ with $r \ll 1$ and $\mathbb{E}M(X_n) \ll \infty$, and we say it is *polynomially ergodic of order* k if $r_n = n^{-k}$ for some positive integer k and again $\mathbb{E}M(X_n) \ll \infty$.

Following [7, Theorem 2] a strong mixing for the Markov chain can be established. First note that the Markov property implies (see [5])

$$\alpha_n = \max_{\substack{A \subset \mathcal{S} \\ B \subset \mathcal{S}}} |\Pr(X_{n+1} \in A \text{ and } X_1 \in B) - \pi(A)\pi(B)|$$

and

$$\begin{aligned} |\Pr(X_{n+1} \in A \text{ and } X_1 \in B) - \pi(A)\pi(B)| &= \left| \sum_{x \in B} (P^n(x, A) - \pi(A))\pi(x) \right| \\ &\le \mathbb{E}M(X_1) \cdot r_n, \end{aligned} \tag{3.7}$$

where $\pi(A) = \sum_{x \in A} \pi(x)$ and $P^n(x, A) = \sum_{y \in A} P(x, y)$. Thus with either geometric ergodicity, or polynomial ergodicity of order $k \ge 5$, the Markov chain will have mixing coefficients satisfying assumption (A.1) of Theorem 2.1.

Theorem 3.1. *Let $X_1, \ldots, X_\nu$, with $\nu \simeq \infty$, be a stationary Markov chain that is polynomially ergodic of order $k \ge 5$. Let f be such that $\mathbb{E}f(X_n)^{12} \ll \infty$ and $\sigma_f^2 \gg 0$, where we use the notation* (3.3), (3.4), *and* (3.5). *Then $\sqrt{\nu}(\bar{f} - \mu_f)/\sigma_f$ is nearly standard normal.*

Proof. Immediate from Theorem 2.1 and the comments preceding this theorem, since $f(X_1), \ldots, f(X_\nu)$ is a stationary stochastic process for which conditions (A.0), (A.1), (A.2) and $\sigma_f \gg 0$ are satisfied. □

Remarks: Convergence Rates for Markov Chains Rosenthal [28] shows how to obtain an upper bound on the left-hand side of (3.6). This result can be used to prove another Markov chain CLT in which drift and minorization conditions imply that $\sqrt{\nu}(\bar{f} - \mu_f)/\sigma_f$ is nearly standard normal for both stationary and nonstationary chains (see [1] where the result is proved and applied to a first order autoregressive process). The CLT in [1] does not work for any starting distribution, only those satisfying $\mathbb{E}\{V(X_0)\} \ll \infty$ where the function V is given by the drift condition. This need to start the chain at a "good" starting point is widely recognized, but not motivated by the conventional Markov chain CLT. In conventional theory [7] a drift condition is enough by itself to ensure geometric ergodicity of the stationary chain, which implies a CLT, and if the chain is also Harris recurrent, the CLT holds regardless of the initial distribution [21, Proposition 17.1.6]. A minorization is not needed since, in fact, it always holds [21, Theorem 5.2.3] but the proof is non-constructive and does not say how large ε is. It could be infinitesimal, which would not do. Hence we need an explicit minorization condition in REPT.

4 Markov Chain Functional CLT

In this section we discuss a functional central limit theorem (FCLT) for Markov chains; the proof of the theorem can be found in [1].

Let X_1, X_2, ... be a stationary Markov chain. Then, under certain assumptions [14, 17, 19, 21], a conventional FCLT holds, meaning that the stochastic process

$$W_n(t) = \frac{1}{\sqrt{n}\sigma_f} \sum_{i=1}^{\lfloor nt \rfloor} [f(X_i) - \mu_f],$$

converges weakly to Brownian motion with $\mu_f = \mathbb{E}\{f(X_i)\}$ and

$$\sigma_f^2 = \operatorname{Var}\{f(X_i)\} + 2 \sum_{j=1}^{\infty} \operatorname{cov}\{f(X_i), f(X_{i+j})\}.$$

Now let X_1, ..., X_ν be a stationary Markov chain having finite state space $\mathcal{S}$ and transition probability matrix P. Consider $f : \mathcal{S} \to \mathbb{R}$ such that $f(X_n)$ is L_2 for some n, and hence for all n by stationarity. Suppose there exists a function $g : \mathcal{S} \to \mathbb{R}$ such that $g(X_n)$ is L_2 for some and hence all n and is a solution to the Poisson equation $(I - P)g = f$, where I is the identity matrix and $g \in \mathbb{R}^{\mathcal{S}}$. This implies $\mathbb{E}f(X_n) = 0$, and since for any constant c, $g + c$ is a solution whenever g is, we may also choose g so that $\mathbb{E}g(X_n) = 0$. Let $h = Pg$, and define the quantity (independent of n by stationarity)

$$\gamma_f^2 = \mathbb{E}\{g(X_n)^2\} - \mathbb{E}\{h(X_n)^2\}. \tag{4.1}$$

We show below that, under certain conditions, the stochastic process

$$Z(t) = \frac{1}{\sqrt{\nu}\gamma_f} \sum_{n=1}^{\nu t-1} f(X_n) \tag{4.2}$$

having *carrier* $T = \{\, k/\nu : k = 0, \dots, \nu \,\}$, $\nu \simeq \infty$, is a Wiener process.

The regularity condition involving solutions of the Poisson equation is not verifiable in practical applications. However, it is a "usual" condition in probability theory [14, 19, 21] and can help clarify theoretical issues [6, 12]. The REPT substitute for this conventional theory serves the same purposes.

The first assumption needed for a FCLT parallels the square integrability condition found in the conventional theory. We require an additional assumption, which we also put in the form of a LLN.

Theorem 4.1. *Suppose g is L_2 and $\gamma_f \gg 0$, and suppose the stochastic process defined by*

$$V_n = \frac{1}{n} \sum_{i=2}^{n} \bigl[\mathbb{E}\{g(X_i)^2 \mid X_{i-1}\} - h(X_{i-1})^2\bigr], \qquad n = 1, \dots, \nu \tag{4.3}$$

converges almost surely to γ_f^2. Suppose further that the stochastic process defined by

$$U_n = \frac{1}{n} \sum_{i=1}^{n} g(X_i)^2 \qquad n = 1, \dots, \nu \tag{4.4}$$

converges almost surely to a limited quantity. Then (4.2) *is a Wiener process.*

Sketch of the proof [1]. Recall that $(I - P)g = f$ and note that $h = Pg$ means that $h(X_n) = \mathbb{E}\{g(X_{n+1}) \mid X_n\}$ so

$$Y(t) = \frac{1}{\sqrt{\nu}\gamma_f} \sum_{n=2}^{\nu t} \bigl[g(X_n) - h(X_{n-1})\bigr] \tag{4.5}$$

is a martingale having mean zero for all t, where γ_f^2 is determined by (4.1). Then, with Z given by (4.2) and Y by (4.5), we can use (4.3) to establish that (i) Y is a Wiener process and (ii) $Z(t)$ and $Y(t)$ are nearly equivalent for each fixed t. The second part of the proof uses (4.5) to show that almost surely $Z(t) \simeq Y(t)$ for all t, the exception set not depending on t. This step boils down to proving that for any $\lambda \gg 0$ and $\mu \simeq \infty$,

$$\Pr\left\{\max_{\mu \le n \le \nu} \frac{g(X_n)^2}{n} \ge \lambda\right\} \simeq 0. \tag{4.6}$$

□

Remarks: Open Questions. In conventional probability, LLNs like the assumptions in Theorem 4.1 follow from Birkhoff's ergodic theorem and an assumption of ergodicity. One says a Markov chain is ergodic if every L_1 functional has a LLN. In REPT, however, ergodicity and related concepts have not yet been studied, so we made the particular LLN we needed an assumption. We do not yet have a REPT substitute for Birkhoff's ergodic theorem, but even if we did, it would not do much good since a REPT substitute for ergodicity or related notions like irreducibility and Harris recurrence, have not been established. This whole area provides a wealth of research topics in this new mathematics.

We can say that the first assumption cannot be weakened, because it is used to prove that, almost surely, $\tau(t) \simeq t$ for all t, which is essentially equivalent to the assumption. Perhaps (4.4) can be weakened, since it is used to prove that, almost surely, $g(X_n)^2/\nu \simeq 0$ for all n, which seems to be a much weaker statement than the assumption.

5 Concluding Remarks

Most of the new results are part of a program of redoing in REPT the Markov chain theory needed for applications [1, 11]. They show that conventional probability theory that seem to require measure theory and functional analysis can be put in the REPT framework. The proofs follow the conventional ones but replace real analysis with simple considerations about the arithmetic of infinitesimals. That results from conventional probability theory that are incomprehensible without measure theory can be moved to REPT is an interesting accomplishment. It is very important to note that REPT theorems involve terms that are not part of classical analysis. The idea is not to prove conventional results by means of REPT but rather to prove radically elementary theorems which have the same scientific content as conventional ones. Finally, the nonstandard tools used by REPT should not be confused with the more orthodox version of nonstandard analsyis (see Pitt's [24] review article for a succint explanation of differences).

References

[1] B.B. Andrade, C.J. Geyer, Nonstandard Central Limit Theorems for Markov chains, (in preparation)

[2] K.P.S. Bhaskara Rao, M. Bhaskara Rao, *Theory of Charges: A Study of Finitely Additive Measures*. London: Academic Press, 1983.

[3] P. Billingsley, *Probability and Measure*, 3rd edition, Wiley, 1995.

[4] E. Bishop, *Foundations of Constructive Analysis*. McGraw-Hill, 1967.

[5] R.C. Bradley, Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions, *Probability Surveys*, 2 (2005) 107144.

[6] K.S. Chan, On the central limit theorem for an ergodic Markov chain. *Stochastic Processes and their Applications*, 47 (1993) 113-117.

[7] K.S. Chan, C.J. Geyer, Discussion on Markov chains for exploring posterior distributions by Luke Tierney, *The Annals of Statistics*, 22 (1994) 1747-1758.

[8] L.E. Dubins and Savage, L.J. *How to Gamble If You Must: Inequalities for Stochastic Processes*. New York: McGraw-Hill, 1965.

[9] N. Dunford, J.T. Schwartz, *Linear Operators, Part I, General Theory*, Interscience Publishers, 1964.

[10] J. Geweke Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, **57** (1989), 1317-1339.

[11] C.J. Geyer, Radically Elementary Probability and Statistics, Technical Report No. 657, School of Statistics, Univ. of Minnesota, 2007.

[12] C.J. Geyer, Conditioning in Markov chain Monte Carlo. *Journal of Computational and Graphical Statistics*, 4 (1995) 2031-2050.

[13] C.J. Geyer and Thompson, E.A. Constrained Monte Carlo maximum likelihood for dependent data, (with discussion). *Journal of the Royal Statistical Society, Ser. B*, **54** (1992), 657-699.

[14] M.I. Gordin, B.A. Lifšic, The central limit theorem for stationary Markov processes. Soviet Mathematics - Doklady, 19 (1978) 392-394 (English translation).

[15] I.A. Ibragimov, Yu.V. Linnik, *Independent and Stationary Sequences of Random Variables*, Wolters-Noordhoff, 1971.

[16] G.L. Jones, On the Markov chain central limit theorem. *Probability Surveys*, 1 (2004) 299-320.

[17] C. Kipnis, S.R.S. Varadhan, Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Communications in Mathematical Physics*, 104 (1986) 1-19.

[18] P. Loeb, Conversion from nonstandard to standard measure spaces and applications to probability theory, *Transactions of the American Mathematical Society* 211 (1975) 113-122.

[19] N. Maigret, Théorème de limite centrale fonctionnel pour une chaîne de Markov récurrente au sens de Harris et positive, Annales de l'institut Henri Poincaré, B Probabilités et Statistiques, 14 (1978), 425-440.

[20] M. Maxwell, M. Woodroofe, Central limit theorems for additive functionals of markov chains, *The Annals of Probability*, 28 (2000) 713-724.

[21] S.P. Meyn, R.L. Tweedie, *Markov Chains and Stochastic Stability*, Springer Verlag, 1993.

[22] E. Nelson, *Radically Elementary Probability Theory*, Princeton University Press, 1987.

[23] E. Nummelin, *General Irreducible Markov Chains and Non-Negative Operators*, Cambridge University Press, 1984.

[24] L.R. Pitt, Review: Edward Nelson, Radically elementary probability theory. *Bulletin of the American Mathematical Society*, (N.S.), 20 (1989) 240-243.

[25] D. Revuz, *Markov Chains*, 2nd edition, North-Holland, 1984.

[26] A.M. Robert, *Nonstandard Analysis*, Dover, Translated by the author from the 1985 French original, 2003.

[27] A. Robinson, *Non-standard Analysis*, Princeton University Press, Reprint of the 1974 revised edition, 1996.

[28] J.S. Rosenthal, Minorization Conditions and Convergence Rates for Markov Chain Monte Carlo, *Journal of the American Statistical Association*, 90 (1995) 558-566.

[29] Tierney, L. Markov chains for exploring posterior distributions. *The Annals of Statistics* **22** (1994), 1701–1728.

[30] B. Tóth, Persistent random walks in random environment, *Probability Theory and Related Fields* 71 (1986) 615-625.

[31] S.R.S. Varadhan, *Probability Theory*, Courant Lecture Notes, American Mathematical Society, 2001.