Statistics & Probability Letters 9 (1990) 207–213
North-Holland

March 1990

A LOCAL LIMIT THEOREM FOR SUMS OF DEPENDENT RANDOM VARIABLES

Mei WANG and Michael WOODROOFE

Department of Statistics, The University of Michigan, Ann Arbor, MI 48109-1027, USA

Received July 1988
Revised February 1989

Abstract: A local version of the central limit theorem is established for normalized sums of dependent random variables when a global theorem is known and conditional distributions are sufficiently smooth. The proof uses ideas from Statistics, by representing the density as the integral of a score function for a translation family of distributions.

AMS 1980 Subject Classification: 60F99.

Keywords: Central limit theorem, almost differentiability, score function, martingales, stationary sequences, Markov chains.

1. Introduction

There is a substantial literature on the Central Limit Theorem for sums of dependent random variables, especially martingales and stationary sequences; several sets of sufficient conditions are known for the convergence in distribution of normed sums. See, for example, Hall and Heyde (1980, Chapters 3 and 5), the review paper by Peligrad (1986) and the references given there. There is much less work on local versions of the Central Limit Theorem for dependent sequences, however. Lalley (1986) notes the paucity of such work and proves a local theorem for Gibb's states.

The purpose of this note is to show that techniques developed by Klassen (1984), Boos (1985) and Sweeting (1986) may be used to establish local versions of the Central Limit Theorem for sums of dependent random variables, when a global theorem is known and conditional distributions are sufficiently smooth. The approach is suggested by the work of Jeganathan (1987).

2. Preliminaries

If $u \in \mathbb{R}^n$ and f is a Borel measurable function from $\mathbb{R}^n$ into $\mathbb{R}$, then f is said to be almost differentiable in the direction u iff there is a measurable function Df_u from $\mathbb{R}^n$ into $\mathbb{R}$ for which

$$f(x \pm cu) - f(x) = \pm \int_0^c Df_u(x \pm tu)\, \mathrm{d}t \tag{1}$$

for a.e. $x \in \mathbb{R}^n$ (Lebesgue) for each $c > 0$. Here (1) includes the condition that the integral exists as a Lebesgue integral. It follows directly from Lemma 2, below, that the function Df_u is essentially unique.

If f is continuously differentiable, then (1) holds for all x and u with $Df_u(x) = u \cdot \nabla f(x)$, where ∇f denotes the gradient of f and $\cdot$ denotes the dot product in $\mathbb{R}^n$; but (1) does not require continuous differentiability. Almost differentiability is used by Stein (1981).

Research supported by the National Science Foundation, under DMS8413452.

The following properties of almost differentiability are needed. The first of these is just the formula for integration by parts. The next two follow from routine applications of Fubini's theorem. The last then follows easily, since a derivative vanishes at a minimum.

Lemma 1. *If f and g are almost differentiable in the direction $u \in \mathbb{R}^n$, then so is fg, and $D(fg)_u = fDg_u + gDF_u$.* □

Lemma 2. *If f is almost differentiable in the direction u, and if f and Df_u are integrable, then*

$$\lim_{t \to 0} \int_{\mathbb{R}^n} \left| \frac{f(x+tu) - f(x)}{t} - Df_u(x) \right| \mathrm{d}x = 0. \qquad \square$$

Lemma 3. *Suppose that f is almost differentiable in the direction u and that f and Df_u are integrable. Let*

$$\bar{f}(x_1, \dots, x_{n-1}) = \int_{-\infty}^{\infty} f(x_1, \dots, x_{n-1}, x_n) \, \mathrm{d}x_n$$

for a.e. $(x_1, \dots, x_{n-1}) \in \mathbb{R}^{n-1}$. Then $\bar{f}$ is almost differentiable in the direction $v = (u_1, \dots, u_{n-1})$, and

$$D\bar{f}_v(x_1, \dots, x_{n-1}) = \int_{-\infty}^{\infty} Df_u(x_1, \dots, x_{n-1}, x_n) \, \mathrm{d}x_n \quad a.e. \qquad \square$$

Lemma 4. *If $f \geqslant 0$ and f is almost differentiable in the direction $u \in \mathbb{R}^n$, then $\{x \in \mathbb{R}^n\colon f(x) = 0$ and $Df_u(x) \neq 0\}$ is a Lebesgue null set.* □

3. Equivariant random variables

Now let $X_1, \dots, X_n$ denote jointly distributed random variables, defined on some probability space $(\Omega, \mathscr{A}, P)$; write $X = (X_1, \dots, X_n)$ for the random vector; and suppose that X has a density f (with respect to n-dimensional Lebesgue measure). Let $\psi : \mathbb{R}^n \to \mathbb{R}$ be a Borel measurable, translation equivariant function; that is,

$$\psi(x_1 + b, \dots, x_n + b) = \psi(x_1, \dots, x_n) + cb$$

for all $x = (x_1, \dots, x_n)$ and $-\infty < b < \infty$ for some $0 < c < \infty$, called the multiplier. Finally, let Y denote a random variable of the form

$$Y = \psi(X) = \psi(X_1, \dots, X_n)$$

and let

$$H(y) = P\{Y \leqslant y\}, \quad -\infty < y < \infty,$$

denote its distribution function.

Proposition 1. *If f is almost differentiable in the direction $1 := (1, \dots, 1)$ and if Df_1 is integrable, then H has density*

$$h(y) = \frac{1}{c} \int_{\{Y \leqslant y\}} l(X) \, \mathrm{d}P, \quad -\infty < y < \infty,$$

where

$$l(x) = \frac{Df_1(x)}{f(x)} I\{f(x) > 0\}, \quad x \in \mathbb{R}^n. \tag{2}$$

Proof. That $E|l(X)| < \infty$ follows directly from the assumed integrability of Df_1. For any y, $-\infty < y < \infty$, and $t > 0$,

$$H(y+t) = \int_{\{x:\psi(x)\leqslant y+t\}} f(x_1,\ldots,x_n)\,\mathrm{d}x_1\cdots\mathrm{d}x_n$$
$$= \int_{\{x:\psi(x)\leqslant y\}} f\left(x_1+\frac{t}{c},\ldots,x_n+\frac{t}{c}\right)\mathrm{d}x_1\cdots\mathrm{d}x_n,$$

where the second equality follows from the change of variables $x_i = x_i' + t/c$ for $i = 1,\ldots,n$. Thus, in more compact notation,

$$\frac{H(y+t)-H(y)}{t} = \int_{\{x:\psi(x)\leqslant y\}} \frac{1}{t}\left[f\left(x+\frac{t}{c}\right)-f(x)\right]\mathrm{d}x$$
$$\to \int_{\{x:\psi(x)\leqslant y\}} \frac{1}{c} Df_1(x)\,\mathrm{d}x$$
$$= \int_{\{x:\psi(x)\leqslant y\}} \frac{1}{c} l(x) f(x)\,\mathrm{d}x = \int_{\{Y\leqslant y\}} \frac{1}{c} l(X)\,\mathrm{d}P, \tag{3}$$

as $t \to 0$ by Lemmas 2 and 4. Thus, H has a derivative at every y. It follows that H is continuous and then that the right side of (3) is continuous. That the derivative of H is its density follows easily. □

4. Normalized sums

Let X_k, $k = 1, 2,\ldots$, denote a sequence of jointly distributed random variables, defined on a probability space $(\Omega, \mathscr{B}, P)$, for which $E(X_k) = 0$ and $0 < E(X_k^2) < \infty$ for all $k = 1, 2,\ldots$. Let

$$S_n^* = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} X_k$$

and

$$H_n(s) = P\{S_n^* \leqslant s\}, \quad s \in \mathbb{R},\ n \geqslant 1.$$

Next, let G_k be a regular conditional distribution function for X_k, given $X_1,\ldots,X_{k-1}$, that is

$$G_k(x_1,\ldots,x_{k-1};\ x_k) = P\{X_k \leqslant x_k \mid X_1 = x_1,\ldots,X_{k-1} = x_{k-1}\}$$

for $-\infty < x_1,\ldots,x_k < \infty$ and $k = 1, 2,\ldots$; and suppose that G_k has a continuous density g_k,

$$g_k(x_1,\ldots,x_{k-1};\ x) = \frac{\partial G_k}{\partial x}(x_1,\ldots,x_{k-1};\ x) \quad \forall x,\ k.$$

Then

$$f_n(x_1,\ldots,x_n) = \prod_{k=1}^{n} g_k(x_1,\ldots,x_{k-1};\ x_k)$$

defines a joint density for $X_1,\ldots,X_n$. Thus, if f_n is almost differentiable in the direction 1 and $Df_{n,1} := D(f_n)_1$ is integrable, then S_n^* has density

$$h_n(s) = \frac{1}{\sqrt{n}} \int_{\{S_n^* \leqslant s\}} L_n\,\mathrm{d}P, \quad s \in \mathbb{R},\ n \geqslant 1, \tag{4}$$

where

$$L_n = l_n(X_1, \ldots, X_n), \quad n \geqslant 1,$$

and l_n is as in (2) with f replaced by f_n.

Proposition 2. *If g_k is almost differentiable in the direction $1 = (1, \ldots, 1)$ (k times) for all $k \geqslant 1$, then f_n is almost differentiable in the direction $1 = (1, \ldots, 1)$ (n times) and*

$$\left(\frac{Df_{n,1}}{f_n}\right)(x_1, \ldots, x_n) = \sum_{k=1}^{n} \left(\frac{Dg_{k,1}}{g_k}\right)(x_1, \ldots, x_k)$$

a.e. on $\{x\colon f_n(x_1, \ldots, x_n) > 0\} \subseteq \mathbb{R}^n$ for all $n = 1, 2, \ldots$. If, in addition,

$$\tau_k^2 := E\left\{\left[\left(\frac{Dg_{k,1}}{g_k}\right)(X_1, \ldots, X_{k-1};\ X_k)\right]^2\right\} < \infty \tag{5}$$

for all $k \geqslant 1$, then L_n, $n \geqslant 1$, defines a martingale for which

$$E(L_n^2) = \sum_{k=1}^{n} \tau_k^2 \quad \forall n \geqslant 1.$$

Proof. The first assertion follows directly from Lemma 1, induction, and division. That

$$L_n = \sum_{k=1}^{n} \left(\frac{Dg_{k,1}}{g_k}\right)(X_1, \ldots, X_{k-1};\ X_k)$$

is integrable for all n follows directly from (5). For the martingale property,

$$\begin{aligned} E\{L_k - L_{k-1} \mid X_1 = x_1, \ldots, X_{k-1} = x_{k-1}\} &= \int_{-\infty}^{\infty} Dg_{k,1}(x_1, \ldots, x_{k-1};\ x_k)\, \mathrm{d}x_k \\ &= D\bar{g}_{k,1}(x_1, \ldots, x_{k-1}) = 0 \end{aligned}$$

for a.e. $x_1, \ldots, x_{k-1}$, by Lemma 3, since $\bar{g}_k(x_1, \ldots, x_{k-1}) = 1$ for all $x_1, \ldots, x_{k-1}$. The final assertion follows, since martingale differences are orthogonal. □

Theorem 1. *With the notation of the previous paragraph, suppose that*

$$H_n \xrightarrow{\mathrm{d}} \text{Normal}(0, \alpha^2), \quad \exists \alpha > 0, \tag{6}$$

and

$$\sup_n E\left\{\frac{1}{n} L_n^2\right\} < \infty. \tag{7}$$

Then

$$\lim_n h_n(s) = \frac{1}{\alpha} \phi\left(\frac{s}{\alpha}\right) \tag{8}$$

uniformly in $-\infty < s < \infty$, where ϕ denotes the standard normal density.

Proof. First observe that for all $s \in \mathbb{R}$, $t > 0$ and $n \geqslant 1$,

$$h_n(s+t) - h_n(s) = \int_{\{s < S_n^* \leqslant s+t\}} \frac{1}{\sqrt{n}} L_n\, \mathrm{d}P,$$

by Proposition 1. Since $\sup_n(1/n)E(L_n^2) < \infty$, $L_n/\sqrt{n}$, $n \geqslant 1$, are uniformly integrable. Moreover, since H_n is continuous for all n and converges to a normal distribution as $n \to \infty$, it is easily seen that $\lim_{t \to 0+} \sup_{n \geqslant 1} H_n(s+t) - H_n(s) = 0$ for all s. So, h_n, $n \geqslant 1$, are equicontinuous. That $h_n(s)$ converges to $\alpha^{-1}\phi(\alpha^{-1}s)$ uniformly on compacts in $-\infty < s < \infty$ now follows from the main result of Boos (1985) and Sweeting (1986). So, it remains to show that $\lim_{|s| \to \infty} \sup_{n \geqslant 1} h_n(s) = 0$. That $\lim_{s \to \infty} \sup_{n \geqslant 1} h_n(-s) = 0$ follows directly from (4), since $L_n/\sqrt{n}$, $n \geqslant 1$, are uniformly integrable and $\lim_{s \to -\infty} \sup_{n \geqslant 1} P\{S_n^* \leqslant -s\} = 0$; and $\lim_{s \to \infty} \sup_n h_n(s) = 0$, by a similar argument. □

Corollary 1. *If* $\sup_k \tau_k^2 < \infty$ *and* (6) *holds, then* (8) *holds.* □

Remark 1. The quantity L_n is the score function (at $\theta = 0$) for the statistical problem in which one observe $Y_k = X_k + \theta$ for $k = 1, \ldots, n$ for some unknown $\theta \in \mathbb{R}$; and $E(L_n^2)$ is the Fisher information for this model. Thus, (7) requires that the Fisher information grow no faster than n.

5. Stationary Markov processes

In this section, $X_1, X_2, \ldots,$ denotes a strictly stationary sequence with conditional densities g_k, $k \geqslant 1$, as in the previous section. Moreover, X_n, $n \geqslant 1$, is assumed to be a Markov chain of order $m-1$, where $2 \leqslant m < \infty$. That is, there are versions of G_k, $k \geqslant 1$, for which

$$G_k(x_1, \ldots, x_{k-1};\ x) = G_m(x_{k-m+1}, \ldots, x_{k-1};\ x)$$

for all $(x_1, x_2, \ldots) \in \mathbb{R}^\infty$, $x \in \mathbb{R}$ and $k \geqslant m$.

Corollary 2. *If* g_k *are almost differentiable in the direction* 1 *for all* $k \leqslant m$, $\tau_k^2 < \infty$ *for all* $k \leqslant m$, *and* (6) *holds, then so does* (8).

Proof. In this case $\tau_k^2 = \tau_m^2$ for all $k > m$, so that $\sup_k \tau_k^2$ is finite. □

In the remainder of this section, it is assumed that the densities g_k, $k \leqslant m$, are almost differentiable in all directions, and that $Dg_{k,u}$ is of the form $Dg_{k,u} = u \cdot \nabla g_k$ for all u for some essentially unique function ∇g_k for all $k \leqslant m$. Let

$$\iota_k = E\left\{\left\| \frac{\nabla g_k}{g_k}(X_1, \ldots, X_{k-1};\ X_k)\right\|^2\right\}, \quad 1 \leqslant k \leqslant m,$$

where $\|\cdot\|$ denotes the Euclidean norm, and observe that $\tau_k^2 \leqslant k\iota_k$ for all $k = 1, \ldots, m$.

Theorem 2. *If* $\iota_m < \infty$ *and* (6) *holds, then* (8) *holds.*

Proof. It suffices to show that $E(L_{m-1}^2) < \infty$, since then

$$E(L_n^2) = E(L_{m-1}^2) + (n-m+1)\tau_m^2 \leqslant E(L_{m-1}^2) + (n-m+1)m\iota_m = O(n)$$

as $n \to \infty$. To see this, first observe that

$$f_{m-1}(x_2, \ldots, x_m) = \int_{-\infty}^{\infty} g_m(x_1, \ldots;\ x_m) f_{m-1}(x_1, \ldots, x_{m-1})\, dx_1$$

for a.e. $x_2, \ldots, x_m$. So, letting $\partial_i = e_i \cdot \nabla$, where $e_i = (0, \ldots, 1, \ldots, 0)$ denotes the ith unit vector,

$$\partial_i f_{m-1}(x_2, \ldots, x_m) = \int_{-\infty}^{\infty} \partial_i g_m(x_1, \ldots; x_m) f_{m-1}(x_1, \ldots, x_{m-1})\, \mathrm{d}x_1 + \int_{-\infty}^{\infty} g_m(x_1, \ldots; x_m) \partial_i f_{m-1}(x_1, \ldots, x_{m-1})\, \mathrm{d}x_1$$

for all $i = 2, \ldots, m$ by Lemmas 1 and 3. Here the second term is to be interpreted as zero when $i = m$. By Schwarz' inequality,

$$\begin{aligned} |\partial_i f_{m-1}(x_2, \ldots, x_m)| \leqslant & \sqrt{\int_{-\infty}^{\infty} \frac{\partial_i g_m(x_1, \ldots,; x_m)^2}{g_m(x_1, \ldots,; x_m)} f_{m-1}(x_1, \ldots, x_{m-1})\, \mathrm{d}x_1} \\ & \times \sqrt{\int_{-\infty}^{\infty} g_m(x_1, \ldots,; x_m) f_{m-1}(x_1, \ldots, x_{m-1})\, \mathrm{d}x_1} \\ & + \sqrt{\int_{-\infty}^{\infty} g_m(x_1, \ldots,; x_m) \frac{\partial_i f_{m-1}(x_1, \ldots, x_{m-1})^2}{f_{m-1}(x_1, \ldots, x_{m-1})}\, \mathrm{d}x_1} \\ & \times \sqrt{\int_{-\infty}^{\infty} g_m(x_1, \ldots,; x_m) f_{m-1}(x_1, \ldots, x_{m-1})\, \mathrm{d}x_1} \end{aligned}$$

for a.e. $x_2, \ldots, x_m$. So, since the second and fourth integrals are just $f_{m-1}(x_2, \ldots, x_m)$ and g_m is a conditional density,

$$\begin{aligned} & \int \cdots \int_{\mathbb{R}^{m-1}} \frac{\partial_i f_{m-1}(x_2, \ldots x_m)^2}{f_{m-1}(x_2, \ldots x_m)}\, \mathrm{d}x_2 \cdots \mathrm{d}x_m \\ & \leqslant 2 \int \cdots \int_{\mathbb{R}^m} \frac{\partial_i g_m(x_1, \ldots; x_m)^2}{g_m(x_1, \ldots; x_m)} f_{m-1}(x_1, \ldots, x_{m-1})\, \mathrm{d}x_1 \cdots \mathrm{d}x_m \\ & \quad + 2 \int \cdots \int_{\mathbb{R}^m} g_m(x_1, \ldots; x_m) \frac{\partial_i f_{m-1}(x_1, \ldots, x_{m-1})^2}{f_{m-1}(x_1, \ldots, x_{m-1})}\, \mathrm{d}x_m \cdots \mathrm{d}x_1 \\ & \leqslant 2\iota_m + 2 \int \cdots \int_{\mathbb{R}^{m-1}} \frac{\partial_i f_{m-1}(x_1, \ldots, x_{m-1})^2}{f_{m-1}(x_1, \ldots, x_{m-1})}\, \mathrm{d}x_1 \cdots \mathrm{d}x_{m-1} \end{aligned}$$

for all $i = 2, \ldots, m$. When $i = m$, the last integral is absent, so that the first is finite. Then, when $i = m - 1$, the last integral is finite, so that the first is finite too. By induction the first integral is finite for all $i = 2, \ldots, m$, thus proving the proposition. □

Remark 2. There are potential relationships between (6) and the condition that ι_m be finite. Using Boos' theorem again, it may be shown that if $\iota_m < \infty$ and $X_1, X_2, \ldots,$ is mixing, then $X_1, X_2, \ldots,$ is strongly (uniformly) mixing, one of a set of conditions which imply asymptotic normality.

References

Boos, D. (1985), A converse to Scheffe's theorem, *Ann. Statist.* **13**, 423–427.

Hall, P. and C. Heyde (1980), *Martingale Limit Theory and its Application* (Academic Press, New York).

Jeganathan, P. (1987), Strong convergence of the distributions of estimators, *Ann. Statist.* **15**, 1699–1708.

Klassen, C. (1984), Location estimators and spread, *Ann. Statist.* **12**, 311–321.

Lalley, S. (1986), Ruelle's Perron Frobenius theorem and the central limit theorem for additive functionals of Gibb's states, in: J. Van Ryzin, ed., *Adaptive Procedures and Related Topics* (Inst. Math. Statist., Hayward, CA) pp. 428–446.

Peligrad, M. (1986), Recent advances in the central limit theorem and its weak invariance principle for mixing sequences of random variables (a survey), in: E. Eberlein and M.S. Taqqu, eds., *Dependence in Probability and Statistics* (Birkhauser, Basel).

Stein, C. (1981), Estimation of the mean of a multivariate normal distribution, *Ann. Statist.* **9**, 1135–1152.

Sweeting, T. (1986), A converse to Scheffe's theorem, *Ann. Statist.* **14**, 1252–1256.