An Elementary Proof of the
Local Central Limit Theorem

by

Burgess Davis*
Purdue University

Technical Report #93-41

Department of Statistics
Purdue University

August 1993

*Supported by a NSF grant.

△

AN ELEMENTARY PROOF OF THE
LOCAL CENTRAL LIMIT THEOREM

by

Burgess Davis
Department of Statistics
Purdue University
West Lafayette, IN 47907-1399

ABSTRACT

We give an elementary proof of the local central limit theorem for i.i.d. integer valued and vector valued random variables, and an extension of this theorem to non-identically distributed variables.

1. Integer valued variables. In this paper we give elementary, characteristic function free, proofs of some results of Gamkrelidze [1], results which include the local central limit theorem for i.i.d. variables. For more general theorems than those proved here, see [4]. Let ϕ and Φ be the standard normal density and distribution functions, and let $a \wedge b$ stand for the minimum of a and b. To keep the notation simple, we first state and prove a local limit theorem for sequences of random variables. At the end of this section we state a more general result for triangular arrays, which follows from exactly the same argument, and derive the local central limit theorem for i.i.d. variables from it. Vector valued variables are considered in Section 2.

Theorem 1: *Let $X_i, i \geq 1$, be independent integer valued random variables, and put $S_n = X_1+X_2+\ldots+X_n$. Suppose there exists a positive number α such that $\max_k P(X_j = k) \wedge P(X_j = k+1) > \alpha$ for at least αn of the integers $j = 1, 2, \ldots, n$, for all $n \geq n_0$ for some integer n_0. Suppose also that there are numbers $a_n, b_n, n \geq 1$, such that $c\sqrt{n} < b_n < C\sqrt{n}$, for positive constants c and C, and such that $\lim_{n\to\infty} P\left(\frac{S_n - a_n}{b_n} < t\right) = \Phi(t), -\infty < t < \infty$. Then $\sup_k |b_n P(S_n = k) - \phi(\frac{k-a_n}{b_n})| \to 0$ as $n \to \infty$.*

The condition involving α was invented by McDonald. See [2] and [3]. Throughout, c and C will stand for positive absolute constants, not necessarily the same at each occurrence. We just mention that the hypothesis $b_n > c\sqrt{n}$ in Theorem 1 is implied by the other hypotheses.

Proof of Theorem 1: Let $e_{k,n} = (k - a_n)/b_n$. Let $h_{k,n} = \Phi(e_{k+1,n}) - \Phi(e_{k,n})$, and note

$$\max_k |b_n h_{k,n} - \phi(e_{k,n})| \to 0 \text{ as } n \to \infty. \tag{1}$$

If $W_n, n \geq 1$, are integer valued random variables, then $(W_n - a_n)/b_n$ converges in distribution to the standard normal if and only if

$$\sup_{j<k} \sum_{i=j}^{k} P(W_n = i) - h_{i,n} \to 0 \text{ as } n \to \infty. \tag{2}$$

We say $(W_n - a_n)/b_n$ has standard normal local limits if $\sup_k |b_n P(W_n = k) - \phi(e_{k,n})| \to 0$ as $n \to \infty$, or, equivalently, if

$$\sup_k b_n |P(W_n = k) - h_{k,n}| \to 0 \; as \; n \to \infty. \tag{3}$$

If X is an integer valued random variable, we put $s(X) = sup_j |P(X = j+1) - P(X = j)|$, and note that if Y is integer valued and independent of X then

$$s(X+Y) \le s(X), \tag{4}$$

and also that if $A_1, A_2, \ldots, A_m$ is a partition of the sample space,

$$s(X) \le \sum_{k=1}^{m} s(X|A_k) P(A_n), \tag{5}$$

where the conditional s is computed from the conditional probabilities. The following lemma is well known, and essentially used in [1]–[4]. Since the proof is easy, we provide it.

Lemma 1. *If* $\sup_n ns(W_n) < \infty$, *and* $(W_n - a_n)/b_n$ *converges in distribution to the standard normal, then* $(W_n - a_n)/b_n$ *has standard normal local limits.*

Proof: We have

$$\sup_n n|h_{k+1,n} - h_{k,n}| < \infty. \tag{6}$$

Let $d_{k,n} = P(W_n = k) - h_{k,n}$. In view of (6), and the first hypotheses on $W_n, n \ge 1$, there is a positive constant c such that $|d_{k,n} - d_{k+1,n}| \le c/n$ for all n and k, which implies

$$|d_{k,n} - d_{j,n}| < c|j-k|/n \text{ for all } n, k, \text{ and } j. \tag{7}$$

Let $\varepsilon > 0$. We show that $|d_{k,n}| > \varepsilon/\sqrt{n}$ can happen for only a finite number of n. Suppose, first, that $d_{k,n} > \varepsilon/\sqrt{n}$. Then by (7), $d_{k,n} > \varepsilon/2\sqrt{n}$ for all j such that $c|j-k|/n \le \varepsilon/2\sqrt{n}$. The number of such j exceeds $\varepsilon\sqrt{n}/2c$. Summing $d_{j,n}$ over these j yields a number exceeding $(\varepsilon\sqrt{n}/2c)(\varepsilon/2\sqrt{n})$, which by (2) can happen for only a finite number of n. The case $d_{k,n} < -\varepsilon/\sqrt{n}$ is handled similarly. □

We will now show that Lemma 1 can be applied to S_n. It is here that this paper differs from its predecessors.

Lemma 2. Let $b(k,n) = \binom{n}{k} 2^{-n}$, $1 \le n, 0 \le k \le n, b(k,n) = 0$ otherwise. Then

$$|b(k+1,n) - b(k,n)| \le 32/n, n \ge 1, 0 \le k < n. \tag{8}$$

Proof: Since $b(k,n), 0 \le k \le n$ is symmetric around $k = n/2$ and monotone on $0 \le k \le n/2$, we have $b(k,n) \le 1/(|n-2k|+1)$, since there are $|n-2k|+1$ integers no further from $n/2$ than k is. Thus, for either $k \le n/4$ or $k \ge 3n/4$, (8) holds. Otherwise, both $k+1 \ge n/4$ and $n-k \ge n/4$, so

$$|b(k+1,n) - b(k,n)| = nb(k,n-1)\frac{|n-2k-1|}{(k+1)(n-k)}$$
$$\le \frac{n}{|(n-1)-2k|+1} \cdot \frac{|n-2k-1|}{(k+1)(n-k)} \le \frac{16}{n} \cdot \frac{|n-2k-1|}{|(n-1)-2k|+1} \le \frac{32}{n}. \qquad \square$$

Let Γ be those positive integers n for which there is an integer $j(n)$ such that $P(X_n = j(n)+1) > \alpha$ and $P(X_n = j(n)) > \alpha$. Let $Z_n, n \ge 1$, be random variables taking on no values exept 0 and 1, such that $Z_n = 0$ if $n \notin \Gamma$, while if $n \in \Gamma, Z_n = 1$ only if $X_n = j(n)$ or $X_n = j(n)+1$. We also require that the random vectors $(X_n, Z_n), 1 \le n < \infty$, are independent, and that $P(Z_n = 1, X_n = j(n)) = P(Z_n = 1, X_n = j(n)+1) = \alpha$. It is easy to construct such Z_n by randomization: If you observe $X_n = j(n)$, flip a coin with probability of heads equal to $\alpha/P(X_n = j(n))$. If you observe $X_n = j(n)+1$, flip a coin with probability of heads equal to $\alpha/P(X_n = j(n)+1)$. Put $Z_n = 1$ if a head is tossed, and if a tail is tossed or if no coin is flipped, put $Z_n = 0$.

Now $X_1, X_2, \ldots, X_n$ are conditionally independent given $Z_1, Z_2, \ldots, Z_n$. Let $A = \{n_1, n_2, \ldots, n_k\} \subset \{1,2,\ldots,n\} \cap \Gamma$. Conditioned on the event Θ that $Z_i = 1$ for exactly those i in A, $X_{n_1} - j(n_1), X_{n_2} - j(n_2), \ldots, X(n_k) - j(n_k)$ are i.i.d., each taking on the values 0 and 1 with probability one half. Write

$$S_n = \sum_{i=1}^{k}(X_{n_i} - j(n_i)) + \sum_{i=1}^{k} j(n_i) + \sum_{\{1,2,\ldots,n\}\setminus A} X_i \tag{9}$$

Then conditioned on Θ, the three sums in (9) are independent, and the first sum has a binomial $(k, 1/2)$ distribution. Thus $s(S_n|\Theta) \le 32/k$, by Lemma 2, and (4), or, writing this a different way, if we let $\lambda(n)$ be the number of integers k in $\{1,2,\ldots,n\}$ such that $Z_k = 1$, we have $s(S_n|\Theta) \le 32/\lambda(n)$.

Since always $s(X) \le 1$, we get $s(S_n|\Theta) \le (32/\lambda(n)) \wedge 1$. Together with (5) this gives

$$s(S_n) \le E(32/\lambda(n)) \wedge 1. \tag{10}$$

To complete the proof of Theorem 1, in view of Lemma 1 it suffices to show the expectation on the right of (10) to be smaller that c/n for an absolute constant c and large enough n. But $\lambda(n)$ is itself binomial, with success probability exactly 2α and number of trials at most n and at least αn, if $n \ge n_0$. Thus Var $\lambda(n) \le n2\alpha(1-2\alpha) \le n/4$, and Chebyshev's inequality gives

$$P(\lambda(n) < \alpha^2 n) \le P(|\lambda(n) - E\lambda(n)| \ge \alpha^2 n) \le \text{ Var } \lambda(n)/(\alpha^2 n)^2 \le 1/4(\alpha^2 n)^2,$$

so $E(32/\lambda(n)) \wedge 1 \leq 32/\alpha^2 n + P(\lambda(n) < \alpha^2 n) < 32/\alpha^2 n + 1/4(\alpha^2 n)^2$, if $n \geq n_0$. □

We note that it is easy to show that if W_n is integer valued and a_n, b_n are constants such that $c\sqrt{n} < b_n < C\sqrt{n}, n \geq 1$, then $(W_n - a_n)/b_n$ converges in distribution to the standard normal if $(W_n - a_n)/b_n$ has standard normal local limits. The following theorem can be proved in the same way Theorem 1 was proved. In fact, it follows from Theorem 1, and the observation that its proof shows that the rate of convergence there depends only on α, c, C, and the rate of convergence of $\sup_t |P((S_n - a_n)/b_n < t) - \Phi(t)|$ to zero.

Theorem 2: *Let $X_{k,n} 1 \leq k \leq n, 1 \leq n < \infty$, be independent integer valued random variables. Suppose there exists a positive number α such that $\max_j P(X_{k,n} = j) \wedge P(X_{k,n} = j+1) > \alpha$ for at least αn integers $1 \leq k \leq n$, if $n \geq n_0$, for some integer n_0. Put $T_n = \sum_{k=1}^{n} X_{k,n}$. Suppose there are numbers $a_n, b_n, n \geq 1$, such that $c\sqrt{n} < b_n < C\sqrt{n}, n \geq 1$, for positive constants c and C, and such that $\frac{T_n - a_n}{b_n}$ converges in distribution to the standard normal. Then* $\sup_k |b_n P(T_n = k) - \phi\left(\frac{k-a_n}{b_n}\right)| \to 0$ as $n \to \infty$.

If $Y_1, Y_2, \ldots$ are i.i.d. random variables with mean μ and variance σ^2 such that $gcd\{j\colon P(Y_i = j) > 0\} = 1$, and $V_n = Y_1 + \ldots + Y_n$, then there is an m such that $P(V_m = k) > 0$ and $P(V_m = k+1) > 0$ for some k. Writing $V_n = V_m + (V_{2m} - V_m) + \ldots + (V_{jm} - V_{(j-1)m}) + (V_n - V_{jm})$, and noting that the first j random variables in this sum have the distribution of V_m, we can apply Theorem 2 and the central limit theorem for i.i.d. random variables (applied to $Y_n, n \geq 1$) to get the local central limit theorem for i.i.d. integer valued variables stated in Spitzer [5], p. 79, namely

$$\sup_k |P(V_n = k)\sigma\sqrt{n} - \phi(\tfrac{k-n\mu}{\sigma\sqrt{n}})| \to 0 \text{ as } n \to \infty. \tag{11}$$

We remark that the analog of Theorem 1, where the condition involving α is replaced by the condition that there is a positive constant γ such that $\sum_k P(X_j = k) \wedge P(X_j = k+1) > \gamma$ for at least γn of the integers $j = 1, 2, \ldots, n$, remains true. See [1]. It is not difficult to adapt the argument just given to prove this analog.

2. Vector valued variables. In this section we state Theorem 3, the multivariate analog of Theorem 1, and indicate how the proof of Theorem 1 can be altered to prove Theorem 3. The multivariate analog of Theorem 2, and the derivation of the multidimensional analog of (11) exactly parallel their counterparts in Section 1 and are omitted. We fix the dimension $d > 1$, let L denote the d dimensional integer lattice, denote vectors in L with arrows, $\vec{x} = (x_1, x_2, \ldots, x_d)$, and let $\vec{i}_j, 1 \leq j \leq d$, be the unit vectors in $\mathbf{R}^d$, so that $\vec{i}_j$ is the vector of 0s except for 1 in the jth position. The Euclidean distance between $\vec{x}$ and $\vec{y}$ is denoted $|\vec{x} - \vec{y}|$. We let ν be a fixed non-degenerate d-dimensional multivariate normal distribution, and let η be its density.

Theorem 3. *Let $\overrightarrow{X}_1, \overrightarrow{X}_2, \ldots$ be a sequence of independent L-valued random vectors. Suppose there is a positive number α such that for each $j, 1 \le j \le d, \max_{\vec{x}\in L} P(\overrightarrow{X}_k = \vec{x}) \wedge P(\overrightarrow{X}_k = \vec{x} + \vec{i}_j) > \alpha$ for at least αn of the integers $k, 1 \le k \le n$. Suppose also that there is a sequence $\vec{a}_n, n \ge 1$, positive constants c, C and a sequence of real numbers $b_n, n \ge 1$, such that $c\sqrt{n} < b_n < C\sqrt{n}$ and such that $(\vec{S}_n - \vec{a}_n)/b_n \to \nu$ in distribution, where $\vec{S}_n = \sum_{i=1}^{n} \overrightarrow{X}_i$. Then*

$$\sup_{\vec{k}\in L} |b_n^{d/2} P(S_n = \vec{k}) - \eta\left(\frac{\vec{k} - \vec{a}_n}{b_n}\right)| \to 0 \text{ as } n \to \infty.$$

The proof of Theorem 3 is an easy modification of the proof of Theorem 1, which will now be described. We let $\vec{e}_{\vec{k},n} = (\vec{k} - \vec{a}_n)/b_n, \vec{k} = (k_1, \ldots, k_d)$, and replace $h_{k,n}$ with

$$h_{\vec{k},n} = \nu\{\vec{t} \in \mathbf{R}^d\colon\ (k_i - a_i)/b_n \le t_i \le (k_i + 1 - a_i)/b_n, 1 \le i \le d\}.$$

The analog of (1) is: $\max_{\vec{k}} |b_n^{d/2} h_{\vec{k},n} - \eta(\vec{e}_{\vec{k},n})| \to 0$ as $n \to \infty$. If $\overrightarrow{X}$ is an L-valued random vector we put

$$s(\overrightarrow{X}) = \sup\{|P(\overrightarrow{X} = \vec{x}) - P(\overrightarrow{X} = \vec{y})|\colon\ |\vec{x} - \vec{y}| = 1, \vec{x}, \vec{y} \in L\}.$$

Exact analogs of (4) and (5) hold.

Lemma 3: *If $sup_n b_n^{(d+1)/2} s(\overrightarrow{W}_n) < \infty$, and $(\overrightarrow{W}_n - \vec{a}_n)/b_n$ converges in distribution to ν, then*

$$\sup_{\vec{x}\in L} |b_n^{d/2} P(\overrightarrow{W}_n = \vec{x}) - \eta(e_{\vec{x},n})| \to 0 \text{ as } n \to \infty.$$

The proof parallels the proof of Lemma 1: If $b_n^{d/2} P(\overrightarrow{W}_n = \vec{x}) - \eta(e_{\vec{x},n})$ exceeds ε, then using the hypothesis on $s(\overrightarrow{W}_n)$, it exceeds $\varepsilon/2$ for all $\vec{y}$ a distance of $c\sqrt{n}$ or less from $\vec{x}$, which contradicts the fact that $(\overrightarrow{W}_n - \vec{a}_n)/b_n$ converges in distribution to ν, since we can find a "square" centered at $\vec{x}$, of these $\vec{y}$, such that $\overrightarrow{W}_n/n$ has a large probability of belonging to this square, compared to the ν-probability of the square.

For positive integers $m_i, 1 \le i \le d$, all in $\{1, 2, \ldots, n\}$, and integers $j_i \in \{0, \ldots, m_i\}, 1 \le i \le d$, put $r_{\vec{m}}(\vec{j}) = \prod_{i=1}^{d} b(j_i, m_i) 2^{-m_i}$. Let $m = \min_{1\le i\le d} m_i$.

Lemma 4: $\sup_{\substack{\vec{x},\vec{y}\in L \\ |\vec{x}-\vec{y}|=1}} |r_{\vec{m}}(\vec{x}) - r_{\vec{m}}(\vec{y})| < c/m^{(d+1)/2}$.

Lemma 4 follows from Lemma 2 and the fact that $b(k, n) < cn^{-1/2}$.

Let $\Gamma_j, 1 \le j \le d$, be disjoint sets of positive integers such that the cardinality of $\Gamma_j \cap \{1, 2, \ldots, n\}$ exceeds cn for all large enough n, and such that if $n \in \Gamma_j$ there is $\vec{x}_j(n)$ such that $P(\overrightarrow{X}_n = \vec{x}_j(n)) > \alpha$ and $P(\overrightarrow{X}_n = \vec{x}_j(n) + \vec{i}_j) > \alpha$. Let $\psi_n, n \ge 1$, be random variables taking on only the values $0, 1, \ldots, d$, such that if $1 \le j \le d$ then $\psi_n \ne j$ if $n \notin \Gamma_j$, and if $n \in \Gamma_j, \{\psi_n =$

$j\} \subset \{\overrightarrow{X}_n = \vec{x}_j(n)$ or $\overrightarrow{X}_n = \vec{x}_j(n) + \vec{i}_j\}$ and $P(\psi_n = j, \overrightarrow{X}_n = \vec{x}_j(n)) = P(\psi_n = j, \overrightarrow{X}_n = \vec{x}_j(n) + \vec{i}_j) = \alpha$. We also require $(\overrightarrow{X}_n, \psi_n), n \geq 1$, to be independent. Fix n. For $1 \leq d \leq n$, let $E_j \subset \Gamma_j \cap \{1, 2, \ldots, n\}$, and let $E_i, 1 \leq i \leq d$ be disjoint. Let e_i be the number of elements in E_i. Let Γ be the event that $\psi_k = j$ for exactly those $k, 1 \leq k \leq n$, which are in E_j, $1 \leq j \leq d$. Write

$$\sum_{k=1}^{n} \overrightarrow{X}_k = \sum_{j=1}^{d} \left(\sum_{i \in E_j} \overrightarrow{X}_i - \vec{x}_i(n) \right) + \sum_{j=1}^{d} \sum_{i \in E_j} \vec{x}_i(n) + \sum_{i \in \{1,2,\ldots,n\} \setminus \cup_{i=1}^{d} E_i} \overrightarrow{X}_i.$$

Conditioned on Γ, the first two double sums and the last sum are independent; and if we let $\vec{Y}$ stand for the first double sum, then $P(\overrightarrow{Y} = \vec{s}) = r_{\vec{e}}(\vec{s})$.

The rest of the proof of Theorem 3 proceeds as before, noting that the minimum of the d binomials, representing the number of $k, 1 \leq k \leq n$, such that $\psi_k = j, 1 \leq j \leq d$, is smaller than cn with probability no greater than c/n^2, for small enough c.

References

[1] Gamkrelidze, N.G. (1988). On the application of a smoothness function in proving a local limit theorem. *Theory Probab. Appl.*, **33**, 352–355.

[2] McDonald, David (1979). On local limit theorems for integer valued random variables. *Theory of Probability and Its Applications*, **24**, 613–619.

[3] McDonald, David (1979). A local limit theorem for large deviations of sums of independent nonidentically distributed random variables. *Ann. Prob.*, **7**, 526–531.

[4] Mukhin, A.B. (1991). Local limit theorems for lattice random variable. *Theory Probab. Appl.*, **35**, 698–713.

[5] Spitzer, F. (1964). *Principles of random walk.* Van Nostrand, New York.