A MULTIDIMENSIONAL LOCAL LIMIT THEOREM FOR LATTICE DISTRIBUTIONS

by
Olaf Krafft

University of North Carolina
and
Technische Hochschule Karlsruhe
Institute of Statistics Mimeo Series No. 530

This research was supported by the
Air Force Office of Scientific
Research Contract No. AF-AFOSR-760-65

Department of Statistics
University of North Carolina
Chapel Hill, N. C.

In this paper a local limit theorem for the extreme tail of a convolution of lattice distributed s-dimensional random vectors shall be proved. It is applied to multinomial distributions. The theorem is a generalization of a result obtained by Blackwell and Hodges [1] for the one-dimensional case and parallels a result obtained by Borovkov and Rogozin [2] for the case of absolutely continuously distributed s-dimensional random vectors.

Let $x' = (\xi_1 \xi_2, \ldots, \xi_s)$ be a random vector in the s-dimensional Eucledian space R_s, a' be a fixed vector in R_s and H be a square matrix of order s with $h := |\det H| > 0$.[1] A vector $k' = (\kappa_1, \kappa_2, \ldots, \kappa_s) \in R_s$ is called an integer vector if all its components κ_j, $j = 1, 2, \ldots, s$ are integers, positive or negative or zero. The set of all integer vectors k is denoted by K. A set

$$L_{a,H} := \{ b \ : \ b = a + Hk, \ k \in K \}$$

is called a lattice with respect to a and H. L' is a sublattice of L if $b \in L'$ implies $b \in L$, and L" is a proper sublattice of L if L" is a sublattice of L and there is at least one $b \in L$ which is not in L". A random vector x is said to be lattice-distributed (L-distributed) if there is a lattice L such that x is confined to L with probability one. In general this lattice L is not uniquely determined. But this can be rectified by the introduction of a concept analogous to that of the maximum span in the one-dimensional case, see Ranga Rao [6]:

For any L-distributed random vector x we say that its lattice L is maximal if

(i) $P(x \in L) = 1$,

(ii) there is no proper sublattice L" of L such that $P(x \in L") = 1$.

1) The notation "$\alpha := \beta$" is used when α is defined as equal to β.

It is shown by Heckendorff [3] that each non-degenerate lattice-distributed random vector has a maximal lattice.

In order to approximate the probabilities for the extreme tail of a convolution of lattice-distributed random vectors we will make the commonly used shift from the tails to the center To this end let G denote the domain of convergence of the moment generating function M(z) of x,

$$M(z) := \sum_{\tilde{x}} \exp(z' \tilde{x}) \, P(x = \tilde{x}).$$

By means of M(z) one can define a class of distributions P_z "conjugate" to the distribution P of x, see e.g. Khinchine [5],

$$P_z(x = \tilde{x}) := M^{-1}(z) \exp(z' \tilde{x}) \, P(x = \tilde{x})$$

which are defined for all $z \in G$. Let F be an arbitrary non-void compact subset of the interior G° of G in R_s and

$$\Phi := \{ b : b = \operatorname{grad} \ln M(z), \; z \in F \}.$$

For $b \in \Phi$ we consider random vectors $y = y(b)$ which take values $\tilde{y} = \tilde{x} - b$ with probabilities

$$(1) \qquad P_{z(b)}(y = \tilde{y}) = M^{-1}(z(b)) \exp(z'(b)\tilde{x}) \, P(x = \tilde{x})$$

where $z(b) \in F$ is that value of z for which $b = \operatorname{grad}_{z(b)} \ln M(z)$. It may be noted that z(b) is that value of z for which the function

$$m(z,b) := M(z) \exp(-z'b)$$

attains its minimum. As long as the distribution P is not concentrated on a hyperplane in R_s the vector z(b) is uniquely determined by b and we can define

$$m(b) := M(z(b)) \exp(-z'(b)b),$$

see Hoeffding [4], Lemma 2.

In the following we will consider a sequence $x_1, x_2, \dots$ of independent random vectors with the same distribution function. $\tilde{x}_{(n)}$ may be defined as $\tilde{x}_{(n)} := \sum_{j=1}^{n} \tilde{x}_j$ and only such values b will be of interest for which there is an $\tilde{x}_{(n)}$ such that $b_n := n^{-1} \tilde{x}_{(n)} = b$.

Theorem: Let $x_1, x_2, \dots$ be a sequence of independent $L_{a,H}$-distributed random vectors with the same distribution function and let the following conditions be satisfied:

(A) The lattice $L_{a,H}$ is maximal,

(B) the determinant of the convariance matrix of x_1 is not equal to zero,

(C) $b_n \in \Phi$ for all n,

(D) the series $M(z + z(b_n)) := \sum_{\tilde{x}} \exp\{(z' + z'(b_n)) \tilde{x}\} P(x_1 = \tilde{x})$ converges for all n and all z from a sphere with positive radius and 0 as origin.

Then the following asymptotic expansion is valid uniformly for all $b_n \in \Phi$:

$$(2) \quad P\left(\sum_{j=1}^{n} x_j = \tilde{x}_{(n)}\right) = \frac{m^n(b_n)\, h}{(2\pi n)^{s/2} \Delta^{1/2}(b_n)} \left\{ 1 + \sum_{j=1}^{\rho-1} c_j(b_n) n^{-j} + O(n^{-\rho}) \right\}.$$

Here $\Delta(bn)$ denotes the determinant of the covariance matrix of $y = y(b_n)$, the coefficients $c_j(b_n)$ are dependent only on the moments of y up to the order $(2j + 2)$ and ρ is any positive integer.

Proof. $$P\left(\sum_{j=1}^{n} x_j = \tilde{x}_{(n)}\right) = \sum_{\tilde{x}_j : \Sigma \tilde{x}_j = \tilde{x}_{(n)}} \prod_{j=1}^{n} P(x_j = \tilde{x}_j) =$$

$$= m^n(b_n) \sum_{\tilde{x}_j : \Sigma \tilde{x}_j = \tilde{x}_{(n)}} \exp\left\{-z'(b_n) \sum_{j=1}^{n} (\tilde{x}_j - b_n)\right\} \prod_{j=1}^{n} P(y_j = \tilde{x}_j - b_n) =$$

$$= m^n(b_n) \sum_{\tilde{x}_j : \Sigma(\tilde{x}_j - b_n) = 0} \prod_{j=1}^{n} P(y_j = \tilde{x}_j - b_n) = m^n(b_n)\, P\left(\sum_{j=1}^{n} y_j = 0\right).$$

For $P(\sum_{j=1}^{n} y_j = 0)$ we may apply the saddlepoint method used by Richer [8]. To that end we first note that if x_1 is $L_{a,H}$-distributed then $\tilde{x}_{(n)}$ will be of the form na + Hk. Hence $y(b_n)$ will be $L_{o,H}$-distributed. Condition (A) of theorem 2 in [8] can be replaced by our condition (A) as it is pointed out by Heckendorff [3]. For the expectation Ey we get

$$Ey = \sum_{\tilde{y}(b_n)} \tilde{y}(b_n)\, P(y = \tilde{y}(b_n))$$

$$= M^{-1}(z(b_n)) \{ \sum_{\tilde{x}_1} \tilde{x}_1 \exp(z'(b_n)\tilde{x}_1)\, P(x_1 = \tilde{x}_1) - b_n \sum_{\tilde{x}_1} \exp(z'(b_n)\tilde{x}_1)\, P(x_1 = \tilde{x}_1) \}$$

$$= M^{-1}(z(b_n)) \{ \text{grad}_{z(b_n)} M(z) - (\text{grad}_{z(b_n)} \ln M(z)) M(z(b_n)) \}$$

$$= M^{-1}(z(b_n)) \{ \text{grad}_{z(b_n)} M(z) - \text{grad}_{z(b_n)} M(z) \} = 0.$$

Since by (1) it is seen that $P_{z(b_n)}(y = \tilde{y}) = 0$ holds if and only if $P(x_1 = \tilde{x}_1) = 0$ condition (B) implies that $\Delta(b_n) \neq 0$ for all b_n.

The proof for that

$$P(\sum_{j=1}^{n} y_j = 0) = \frac{h}{(2\pi n)^{s/2}\, \Delta^{1/2}(b_n)} \{ 1 + \sum_{j=1}^{\rho-1} c_j(b_n)\, n^{-j} + O(n^{-\rho}) \}$$

holds is then with slight modifications the same as it is given by Richter [8]. The saddlepoint will be at $z_o = 0$, the expansions of the logarithm of the moment generating function of $y(b_n)$ in Taylor series will be uniform for $b_n \in \Phi$. It should be mentioned here that the coefficients of the expansions in Richter's papers [7] and [8] are partially incorrect. So in formula (11) in [7] the factor $\rho_3^{\,2}(3!)^{-2}$ should be multiplied by 1/2 - the same correction should be made in [8], p. 102, line 8 f.b. - and the formula for I_1 in the second line from the top, p. 214 in [7], becomes

$$I_1 = \frac{i\sqrt{2\pi}\, e^{n(K(z_o) - z_o \tau\sigma}}{\sqrt{n\, K''(z_o)}} \left[1 + \{ \frac{\rho_4(z_o)}{8} - \frac{5}{24} \rho_3^2(z_o) \} \frac{1}{n} + o(\frac{1}{n}) \right].$$

Whereas in applications of this theorem to special cases the evaluation of $z(b_n)$ and of $\Delta(b_n)$ might be rather cumbersome it can easily be done for multinomial distributions:

Let x be an s-dimensional random vector and e_j, $j = 1,2,\ldots, s$, be the unit vectors in R_s. The distribution of x may be given by

$$P\ (x = e_j) = p_j,\ j = 1, 2,\ldots,s,\quad P\ (x = 0) = p_{s+1},\ \sum_{j=1}^{s+1} p_j = 1,\ p_j > 0,\ j=1,2,\ldots,s+1.$$

x is then lattice-distributed with maximal lattice $L_{o,I}$ where I is the s-dimensional unit matrix, $h = 1$. The sum of n such vectors x defines an $(s + 1)$-dimensional multinomial distribution, i. e., if there are $(s + 1)$ pairwise disjoint events B_j as results of an experiment with $P\ (B_j) = p_j$, $j = 1, 2,\ldots, s + 1$, and $p\ (n_1, n_2, \ldots, n_{s+1})$ is the probability that in n independent repetitions of the experiment the event B_1 occurs n_1 times, B_2 occurs n_2 times,..., B_{s+1} occurs n_{s+1} times, $\sum_{j=1}^{s+1} n_j = n$, $n_j \geq 0$, $j = 1, a,\ldots,s + 1$, then

$$p(n_1, n_2,\ldots,n_{s+1}) = P\ (\sum_{j=1}^{n} x_j = \sum_{j=1}^{s} n_j e_j) = \frac{n\,!}{n_1!\, n_2!\, \ldots n_{s+1}!}\, p_1^{n_1} p_2^{n_2} \ldots p_{s+1}^{n_{s+1}}.$$

For the moment generating function $M(z)$ of x we get with $z' = (\zeta_1, \zeta_2, \ldots \zeta_s)$

$$M(z) = p_{s+1} + \sum_{j=1}^{s} e^{\zeta_j} p_j .$$

Let $n^{-1} \sum_{j=1}^{n} \tilde{x}'_j = b'_n = (\beta_1, \beta_2, \ldots \beta_s)$ and consider for fixed $\epsilon > 0$ values b'_n with $\beta_j > \epsilon$, $j = 1, 2,\ldots,s$, $\beta_{s+1} := 1 - \sum_{j=1}^{s} \beta_j < 1 - \epsilon$. The conditions of the theorem are then satisfied. The vector $z(b_n)$ can be calculated differentiating with respect to z the function

$$M(z)\ \exp\ (z' b_n) = p_{s+1} \exp\ (- \sum_{j=1}^{s} \zeta_j \beta_j) + \sum_{j=1}^{s} p_j \exp(\zeta_j (1-\beta_j) - \sum_{k \neq j} \zeta_k \beta_k).$$

The partial derivatives will be zero for those values of z for which

$$\beta_j \left(p_{s+1} + \sum_{k=1}^{s} p_k e^{\zeta_k}\right) = p_j e^{\zeta_j}, \quad j = 1, 2, \ldots, s, \text{ i. e., if}$$

$$\zeta_j = \ln \left(\beta_j \, p_{s+1} / p_j \, \beta_{s+1}\right) .$$

This result makes it possible to calculate the functions $M(z + z(b_n))$ and $m(b_n)$ to

$$M(z + z(b_n)) = \frac{p_{s+1}}{\beta_{s+1}} \{\beta_{s+1} + \sum_{j=1}^{s} \beta_j e^{\zeta_j}\} \quad \text{and}$$

$$m(b_n) = \prod_{j=1}^{s+1} (p_j/\beta_j)^{\beta_j} .$$

$y(b_n)$ takes values $e_j - b_n$, $j = 1, 2, \ldots, s$, and $(-b_n)$ with probabilities $P(y = e_j - b_n) = \beta_j$, $j = 1, 2, \ldots, s$, $P(y = -b) = \beta_{s+1}$, from which the covariance matrix $C(b_n)$ is calculated as

$$C(b_n) = (c_{ij}), c_{ij} = \begin{cases} -\beta_i\beta_j, & i \neq j \\ (1-\beta_i)\,\beta_i, & i=j. \end{cases}$$

Adding all remaining rows of $C(b_n)$ to the first one we get for $\Delta(b_n)$

$$\Delta(b_n) = \beta_{s+1} \begin{vmatrix} \beta_1, & \beta_2, & \ldots, & \beta_s \\ -\beta_1\beta_2, & (1-\beta_2)\beta_2, & \ldots, & -\beta_1\beta_s \\ \vdots & \vdots & & \vdots \\ -\beta_1\beta_s, & -\beta_2\beta_s, & \ldots, & (1-\beta_s)\beta_s \end{vmatrix} .$$

If we now add the first row, multiplied with β_j, to the j-th row, $j = 2, 3, \ldots, s$, we get

$$\Delta(b_n) = \prod_{j=1}^{s+1} \beta_j .$$

With the help of (2) we can then estimate the probabilities (3) to

(4) $$p(n_1, n_2, \ldots, n_{s+1}) = \{(2\pi n)^s \prod_{j=1}^{s+1} (n_j/n)\}^{-1/2} \prod_{j=1}^{s+1} (np_j/n_j)^{n_j} \{1+O(n^{-1})\},$$

an expression which is obtained also by Hoeffding [4], applying Sterling's formula to (3). The function m(b) is closely related to the function I (A(s), F) used by Hoeffding to derive more general results on probabilities of large deviations.

Acknowledgement. I am grateful to Professor W. Hoeffding for useful discussions and constructive comments concerning this paper.

References

[1] Blackwell, D. and Hodges, J. L., Jr. (1959). Probability in the extreme tail of a convolution. Ann. Math. Statist. 30 1113-1120.

[2] Borovkov, A. A. and Rogozin, B. A. (1965). On the central limit theorem in the multidimensional case. (Russian) Teor. Verojatnost. i Primenen. 10 61-69. English translation in Theor. Probab. and Appl. 10.

[3] Heckendorff, H. (1965). On a multidimensional local limit theorem of random vectors with integer-valued coordinates. (Ukrainian) Dop. Akad. Nauk Ukr. RSR 1 19-23.

[4] Hoeffding, W. ( ). On probabilities of large deviations. Proc. Fifth Berkeley Symp. on Math. Stat. and Prob. To appear.

[5] Khinchin, A. I.(1949). Statistical Mechanics. Dover Publications, New York.

[6] Ranga Rao, R. (1961). On the central limit theorem in R_k. Bull. Amer. Math. Soc. 67 359-361.

[7] Richter, W. (1957). Local limit theorems for large deviations. (Russian) Teor. Verojatnost. i. Primenen. 2 214-229. English translation in Theor. Probab. and Appl. 2 206-220.

[8] Richter, W. (1958). Multi-dimensional local limit theorems for large deviations. (Russian) Teor. Verojatnost. i Primenen. 3 107-114. English translation in Theor. Probab. and Appl. 3 100-106.