ESTIMATION OF THE RATE OF CONVERGENCE TO THE NORMAL LAW OF TRUNCATED LINEAR COMBINATIONS OF ORDER STATISTICS

N. V. Gribkova

1. In [1] one has considered the class of statistics of the following form:

$$T = \tau(X_1, \ldots, X_N), \tag{1}$$

where $\tau: \mathbf{R}^N \to \mathbf{R}$ is a symmetric function of N variables, $X_1,\ldots,X_N$ are independent, identically distributed random variables. In [1] one has investigated the rate of convergence to the normal law of random variables of the form (1), when the number of arguments of the function τ tends to infinity, and one has proved the validity of the following result.

THEOREM A. Assume that $\mathsf{E}T = 0$, $\mathsf{E}T^2 = 1$ and there exist positive constants A and B such that

$$\mathsf{E}\,|\,\mathsf{E}(T \mid X_1)\,|^3 \leqslant AN^{-3/2}, \tag{2}$$

$$1 + \mathsf{E}\{\mathsf{E}(T \mid X_1, \ldots, X_{N-2})\}^2 - 2\mathsf{E}\{\mathsf{E}(T \mid X_1, \ldots, X_{N-1})\}^2 \leqslant BN^{-3}. \tag{3}$$

Then

$$\sup_x |\,\mathsf{P}(T \leqslant x) - \Phi(x)\,| \leqslant C(A + B)N^{-1/2}, \tag{4}$$

where C is an absolute constant, while $\Phi(x)$ is the standard normal distribution function.

One of the consequences of this theorem, obtained in [1], refers to linear combinations of order statistics. We denote $X_{(1)} \leqslant \ldots \leqslant X_{(N)}$ the order statistics of the sample $X_1,\ldots,X_N$, we set $F(x) = \mathsf{P}\{X_1 \leqslant x\}$ and

$$L = N^{-1/2}\sum_{i=1}^{N} c_i X_{(i)}, \tag{5}$$

where c_i are real numbers. Preserving the notations of [1], we set

$$\max_{1\leqslant i\leqslant N} |c_i| = a, \quad N \max_{2\leqslant i\leqslant N} |c_i - c_{i-1}| = b. \tag{6}$$

COROLLARY. If

$$0 < \sigma^2(L) < \infty,$$

then

$$\sup_x \left| \mathsf{P}\left\{\frac{L - \mathsf{E}L}{\sigma(L)} \leqslant x\right\} - \Phi(x)\right| \leqslant C\left[\frac{a^3\mathsf{E}\,|X_1|^3}{\sigma^3(L)} + \frac{b^2\{\mathsf{E}\,|X_1|\}^2}{\sigma^2(L)}\right]N^{-1/2}, \tag{7}$$

where C is an absolute constant.

If $\mathsf{E}\,|X_1|^3 < \infty$, the constants a and b are bounded from above uniformly with respect to N, while the variance $\sigma^2(L)$ stays away from zero, then (7) means an estimate of the rate of convergence of order $N^{-1/2}$.

Such an estimate has been obtained for the first time in [2] for the case when the linear combination is truncated according to the quantiles of the samples, i.e., $c_i = 0$ if $i \leqslant \alpha N$ or $i > \beta N$ $(0 < \alpha < \beta < 1)$. The result of [2] is not generalized by the above given corollary since, in general, a truncated linear combination does not satisfy the smoothness condition of the weights (6) and, in addition, in [2] one does not require the assumptions on the finiteness of the moments of the initial distribution; however, in [2] one imposes rigid smoothness conditions on the distribution F(x), which is not done in [1]. In this paper, using Van Zwet's technique, we obtain a result for linear combinations, truncated as in [2],

M. A. Bonch-Bruevich Leningrad Electrical Engineering Institute for Communications.
Translated from Matematicheskie Zametki, Vol. 42, No. 5, pp. 739-746, November, 1987. Original article submitted January 30, 1985.

 0001-4346/87/4256-0906$12.50

without requiring the moment assumptions and the smoothness of $F(x)$, but imposing smoothness conditions on the coefficients.

2. We assume that the coefficients of the linear combinations (5) satisfy the following conditions:

(I) $c_i = 0$ if $1 \leqslant i < k$ or $m < i \leqslant N$, where k, m are integers such that $k/N \to \alpha$, $m/N \to \beta$ when $N \to \infty$ $(0 < \alpha < \beta < 1)$.

(II) $F^{-1}(u)$ is defined and satisfies the Lipschitz condition in Ω_α, Ω_β, fixed neighborhoods of α and β.

We select an arbitrary $\varepsilon > 0$ and we consider the truncated random variables (r.v.)

$$\overline{X}_i = \begin{cases} F^{-1}(\alpha) - \varepsilon, & \text{if } X_i < F^{-1}(\alpha) - \varepsilon, \\ X_i, & \text{if } X_i \in [F^{-1}(\alpha) - \varepsilon, F^{-1}(\beta) + \varepsilon,] \\ F^{-1}(\beta) + \varepsilon, & \text{if } X_i > F^{-1}(\beta) + \varepsilon, \end{cases} \tag{8}$$

where $i = 1,\ldots,N_1$, $\overline{X}_i$ are independent and identically distributed, their distribution function (d.f.) being denoted by $V(x)$. We introduce the order statistics (o.s.) $\overline{X}_{(1)} \leqslant \ldots \leqslant \overline{X}_{(N)}$ of the sample $\overline{X}_1,\ldots,\overline{X}_N$ and let

$$\overline{L} = N^{-1/2} \sum_{i=1}^{N} c_i \overline{X}_{(i)} \tag{9}$$

be their linear combination with the same collection of coefficients as for L. We assume that the variance

(III) $\sigma^2(\overline{L}) > 0$.

We consider the normalized linear combination

$$L^* = (L - \mathsf{E}\overline{L})/\sigma(\overline{L}).$$

We set

$$a = \max_{k \leqslant i \leqslant m} |c_i|, \quad b = N \max_{k \leqslant i \leqslant m-1} |c_{i+1} - c_i|. \tag{10}$$

<u>THEOREM 1.</u> If conditions (I) and (II) hold, a and b are bounded from above uniformly with respect to N, and $\sigma^2(\overline{L}) > \gamma > 0$, then there exists a positive constant C, depending on α, β, F, γ, and ε, but not on N, such that

$$\sup_x |\mathsf{P}(L^* \leqslant x) - \Phi(x)| \leqslant CN^{-1/2} \tag{11}$$

for sufficiently large N.

In fact, we shall prove another, more general theorem. We set $K = \max(|F^{-1}(\alpha) - \varepsilon|, |F^{-1}(\beta) + \varepsilon|)$.

<u>THEOREM 2.</u> Assume that the conditions (I)-(III) are satisfied. Then there exists an absolute positive constant C, and positive constants C_1, C_2, C_3, independent of N, such that for all N we have

$$\leqslant e^{-C_1 N} + C\left[\frac{4a^3K^3}{\sigma^3(\overline{L})} + \frac{\max^2(a,b)K^2}{\sigma^2(\overline{L})}(C_2 + N^2 e^{-C_3 N})\right]N^{-1/2}. \tag{12}$$

<u>Remarks:</u> 1°. The constants C_1, C_3 in Theorem 2 depend only on α, β, F, ε and on the properties of the sequences k and m, while C_2 depends on α, β, F, and ε.

2°. The normalizing constants and the right-hand side of the estimate (12) depend on the arbitrarily selected ε, but, in general, this dependence is complicated and an optimization of the estimate with respect to ε fails.

3°. If to the assumptions of Theorem 1 we adjoin the existence of the moment $\mathsf{E}|X_1|^\delta$ for some $\delta > 0$, then with the aid of the same technique we can easily prove that inequality (11) is valid with a constant C, depending only on α, β, F, γ, and δ, for the normalizing constants we can take $\mathsf{E}L$ and $\sigma(L)$, and in this case the truncation of the initial r.v. is not required.

4°. The continuity and the strict monotonicity of the d.f. $F(x)$ is assumed only in the neighborhoods of the quantiles $F^{-1}(\alpha)$ and these assumptions cannot be dropped. For example, if the conditions c_i are determined with the aid of the function $h(u)$, $u \in (0, 1)$, $c_i = h(i/(N+1))$, then in our case h(u) is discontinuous at the points α and β, but in [3] one gives an example which shows that the asymptotic normality of L need not exist if the discontinuity points of $F^{-1}(u)$ and of the weight function h(u) coincide.

In order to prove Theorem 2 we need an auxiliary statement.

LEMMA. Assume that $k/N \to \alpha$ when $N \to \infty$, k is an integer; if $F(\Delta_1) < \alpha < (F\Delta_2)$, then there exists a positive constant θ, depending on α, F, Δ_1, Δ_2 and on the properties of the sequence k, but not on N, such that for all N we have $\mathsf{P}\{X_{(k)} \notin [\Delta_1, \Delta_2]\} \leqslant e^{-\theta N}$.

This lemma is proved in [2] (see Lemma 1.1 [2, p. 357]. In the formulation of this lemma one has the condition $k = \alpha N + O(1)$ as $N \to \infty$, but its proof in [2] is based only on the convergence of k/N to α).

3. Proof of Theorem 2. We consider L and $\bar{L}$, defined in (5) and (9), and we note that by the condition (1) we have the imbeddings of the events

$$\{L \leqslant x\} \subset \{\bar{L} \leqslant x\} \cup \{X_{(k)} < F^{-1}(\alpha) - \varepsilon\} \cup \{X_{(m)} > F^{-1}(\beta) + \varepsilon\},$$

$$\{\bar{L} \leqslant x\} \subset \{L \leqslant x\} \cup \{X_{(k)} < F^{-1}(\alpha) - \varepsilon\} \cup \{X_{(m)} > F^{-1}(\beta) + \varepsilon\},$$

and, therefore, we have, uniformly with resepct to $x \in \mathbf{R}$,

$$|\mathsf{P}\{L \leqslant x\} - \mathsf{P}\{\bar{L} \leqslant x\}| \leqslant \mathsf{P}\{X_{(k)} < F^{-1}(\alpha) - \varepsilon\} + \mathsf{P}\{X_{(m)} > F^{-1}(\beta) + \varepsilon\}. \tag{13}$$

From (13), by the lemma and by condition (I) there follows that there exists a constant $C_1 > 0$, depending on α, β, F, ε and on the properties of the sequences k and m, but not on N, such that for all N we have

$$\sup_x |\mathsf{P}\{L^* \leqslant x\} - \Phi(x)| \leqslant e^{-c_1 N} + \sup_x |\mathsf{P}\{\bar{L}^* \leqslant x\} - \Phi(x)|, \tag{14}$$

where $\bar{L}^* = (\bar{L} - \mathsf{E}\bar{L})/\sigma(\bar{L})$. $\bar{L}^*$ is a normalized symmetric function of independent, identically distributed random variables $\bar{X}_1, \ldots, \bar{X}_N$, and the subsequent proof reduces to the verification of the conditions (2), (3) of Theorem A (see the introduction) for it. In order to establish the validity of (2), it is sufficient to refer to the analogous verification in [1, p. 437]. In [1] it is proved that for any linear combination L of the o.s. $X_{(1)}, \ldots, X_{(N)}$ of the sample $X_1, \ldots, X_N$, under the condition $0 < \sigma^2(L) < \infty$ we have the inequality

$$\mathsf{E}\left|\mathsf{E}\left(\frac{L - \mathsf{E}L}{\sigma(L)} \middle| X_1\right)\right|^3 \leqslant \frac{4a^3 \mathsf{E}|X_1|^3}{\sigma^3(L)} N^{-3/2}, \tag{15}$$

where $a = \max_{1 \leqslant i \leqslant N} |c_i|$; the condition of the smoothness of the coefficients (6) has not been used for the proof of (15); therefore, for a linear combination $\bar{L}$ of the o.s.s of the sample $\bar{X}_1, \ldots, \bar{X}_N$, $|\bar{X}_i| \leqslant K$ (i = 1,...,N), we have

$$\mathsf{E}|\mathsf{E}(\bar{L}^* | \bar{X}_1)|^3 \leqslant \frac{4a^3}{\sigma^3(\bar{L})} K^3 N^{-3/2}, \tag{16}$$

i.e. (2) holds with the constant $A = 4a^3K^3/\sigma^3(\bar{L})$.

In order to prove (3), we introduce, as in [1], the r.v.

$$Z = \bar{L}_1 - \mathsf{E}(\bar{L}_1 | \bar{X}_1, \ldots, \bar{X}_{N-1}) - \mathsf{E}(\bar{L}_1 | \bar{X}_1, \ldots, \bar{X}_{N-2}, \bar{X}_N) + \mathsf{E}(\bar{L}_1 | \bar{X}_1, \ldots, \bar{X}_{N-2}),$$

where $\bar{L}_1 = \bar{L} - \mathsf{E}\bar{L}$, for which

$$\mathsf{E}Z^2 = \mathsf{E}\bar{L}_1^2 + \mathsf{E}\{\mathsf{E}(\bar{L}_1 | \bar{X}_1, \ldots, \bar{X}_{N-2})\}^2 - 2\mathsf{E}\{\mathsf{E}(\bar{L}_1 | \bar{X}_1, \ldots, \bar{X}_{N-1})\}^2. \tag{17}$$

Comparing (3) and (17), we see that for the verification of (3) we have to estimate $\mathsf{E}Z^2$. We need one more formula [1]; we introduce additional notations. Assume that for $n \leqslant N$, $\bar{X}_{1,n} \leqslant \ldots \leqslant \bar{X}_{n,n}$ denote, as in [1], the o.s. of the r.v. $\bar{X}_1, \ldots, \bar{X}_n$, and we set $\bar{X}_{0,n} = F^{-1}(\alpha) - \varepsilon$, $\bar{X}_{n+1,n} = F^{-1}(\beta) + \varepsilon$. Let R_{N-1}, R_N be the ranks of $\bar{X}_{N-1}$, $\bar{X}_N$ in the sample $\bar{X}_1, \ldots, \bar{X}_N$. We introduce $k_1 = \min(R_{N-1}, R_N)$, $k_2 = \max(R_{N-1}, R_N)$ and the functions

$$G(x) = \int_{-\infty}^{x} V(y)\,dy, \quad H(x) = \int_{x}^{\infty} (1 - V(y))\,dy,$$
$$M(x) = \int_{-\infty}^{x} V(y)(1 - V(y))\,dy, \tag{18}$$

where $V(y)$ is the d.f. of the r.v. $\bar{X}_1$. The functions G and M do not decrease with respect to x, H does not increase. We have the formula

$$N^{1/2}Z = \sum_{i=1}^{N-1} (c_{i+1} - c_i)(M(\bar{X}_{i,N-2}) - M(\bar{X}_{i-1,N-2})) -$$
$$- \sum_{i=1}^{k_1} (c_{i+1} - c_i)(G(\bar{X}_{i,N}) - G(\bar{X}_{i-1,N})) + \sum_{i=k_2}^{N} (c_i - c_{i-1})(H(\bar{X}_{i+1,N}) - H(\bar{X}_{i,N})) \qquad (19)$$

(see [1, p. 437]). If we denote the sums in the right-hand side of (19) in succession by S_1, S_2, S_3, then $N^{1/2}Z = S_1 - S_2 + S_3$ and

$$N^{1/2} |Z| \leqslant |S_1| + |S_2| + |S_3|. \qquad (20)$$

By the definition (10) of a and b and by condition (I), we have

$$|S_1| \leqslant \frac{b}{N}(M(\bar{X}_{m-1,N-2}) - M(\bar{X}_{k-1,N-2})) +$$
$$+ a(M(\bar{X}_{m,N-2}) - M(\bar{X}_{m-1,N-2})) + a(M(\bar{X}_{k-1,N-2}) - M(\bar{X}_{k-2,N-2})). \qquad (21)$$

In order to estimate $|S_2|$, we note that $\bar{X}_{i,N} = \bar{X}_{i,N-2}$ for $i \leqslant k_1 - 1$ and $\bar{X}_{k_1,N} \leqslant \bar{X}_{k_1,N-2}$; therefore,

$$G(\bar{X}_{k_1,N}) - G(\bar{X}_{k_1-1,N}) \leqslant G(\bar{X}_{k_1,N-2}) - G(\bar{X}_{k_1-1,N-2})$$

and

$$|S_2| \leqslant \sum_{i=1}^{k_1} |c_{i+1} - c_i| (G(\bar{X}_{i,N-2}) - G(\bar{X}_{i-1,N-2})) \leqslant \sum_{i=1}^{N-1} |c_{i+1} - c_i| (G(\bar{X}_{i,N-2}) - G(\bar{X}_{i-1,N-2})).$$

Consequently, for $|S_2|$ we have the estimate (21), but with the replacement of M by G in the right-hand side. In a similar manner we obtain that for $|S_3|$ we have (21) if instead of M we set $-H$. We note that for the derivative we have the inequalities $0 \leqslant (M(x) + G(x) - H(x))' \leqslant 5/4$ and, therefore, from (20) and from the fact that $|\bar{X}_i| \leqslant K$ for i = 1,...,N, there follow the inequalities

$$N^{1/2}|Z| \leqslant \frac{5}{4}\Big(\frac{b}{N}(\bar{X}_{m-1,N-2} - \bar{X}_{k-1,N-2}) + a(\bar{X}_{m,N-2} -$$
$$- \bar{X}_{m-1,N-2}) + a(\bar{X}_{k-1,N-2} - \bar{X}_{k-2,N-2})\Big) \leqslant \frac{5}{2}\frac{b}{N}K +$$
$$+ \frac{5}{4}a((\bar{X}_{m,N-2} - \bar{X}_{m-1,N-2}) - (\bar{X}_{k-1,N-2} - \bar{X}_{k-2,N-2})). \qquad (22)$$

We estimate the second moments of the differences in the right-hand side of (22). We consider the first one, $\bar{X}_{m,N-2} - \bar{X}_{m-1,N-2}$, the estimates for the second one are similar. If we denote by $V_{m,m-1}(x,y)$ the joint d.f. of the o.s. $\bar{X}_{m,N-2}$ and $\bar{X}_{m-1,N-2}$, then

$$\mathsf{E}(\bar{X}_{m,N-2} - \bar{X}_{m-1,N-2})^2 = \iint_{\mathbf{R}^2} (x-y)^2 \, dV_{m,m-1}(x,y) = I.$$

We note that $V^{-1}(u)$ satisfies the Lipschitz condition in the neighborhood $\Omega_\beta' = \Omega_\beta \cap (F(F^{-1}(\alpha) - \varepsilon), F(F^{-1}(\beta) + \varepsilon))$ of the point β with the same constant as $F^{-1}(u)$; we denote it by c_β, we denote by Ω the preimage of this neighborhood, $\Omega = V^{-1}(\Omega_\beta')$, and we divide I into two intervals $I = I_1 + I_2$, where

$$I_1 = \iint_{\Omega\times\Omega} (x-y)^2 \, dV_{m,m-1}(x,y),$$
$$I_2 = \iint_{\mathbf{R}^2\setminus\Omega\times\Omega} (x-y)^2 \, dV_{m,m-1}(x,y).$$

We estimate I_1. We denote by $U_{(m,m-1)}(u, v)$ the joint d.f. of the o.s. $U_{m,N-2}$ and $U_{m-1,N-2}$ of the uniform distributions on (0, 1), and after a change of variables we obtain

$$I_1 = \iint_{\Omega_\beta'\times\Omega_\beta'} (V^{-1}(u) - V^{-1}(v))^2 \, dU_{m,m-1}(u,v),$$

and, by the Lipschitz condition,

$$I_1 \leqslant c_\beta^2 \iint_{\Omega_\beta'\times\Omega_\beta'} (u-v)^2 \, dU_{m,m-1}(u,v) \leqslant c_\beta \mathsf{E}(U_{m,N-2} - U_{m-1,N-2})^2.$$

For all m = 2,...,N − 2 we have $\mathsf{E}(U_{m,N-2} - U_{m-1,N-2})^2 = 2/(N(N-1))$ (see [4, Chap. 5]); therefore

$$I_1 \leqslant 2\left(1+\frac{1}{N-1}\right)c_\beta^2 N^{-2}. \qquad (23)$$

We consider I_2. In view of the fact that $|\bar{X}_i| \leqslant K$ ($i = 1,\ldots,N$), we have

$$I_2 \leqslant 4K^2 \int_{R\setminus\Omega} dV_m(x) = 4K^2 \mathbf{P}\{\bar{X}_{m,N-2} \notin \Omega\},$$

where $V_m(x)$ is the d.f. of $\bar{X}_{n,N-2}$, and, by the lemma, one can find a positive constant θ_1, depending on β, F, ε, and on the properties of the sequence m, but not on N, such that

$$I_2 \leqslant 4K^2 e^{-\theta_1 N}. \qquad (24)$$

From (23), (24) and from the similar estimates for $\mathbf{E}(\bar{X}_{k-1,N-2} - \bar{X}_{k-2,N-2})^2$, by (22) and by the Schwarz inequality there follows that for all N we have

$$\mathbf{E}Z^2 \leqslant \max{}^2(a, b)\, K^2 (C_2 + N^2 e^{-C_3 N})\, N^{-3}, \qquad (25)$$

C_2 depends only on α, β, F and ε, C_3 only on α, β, F, ε and on the properties of the sequences m and k. From (17) and (25) we obtain that (3) is satisfied for $\bar{L}^*$ with a constant $B = \max^2(a, b)\, K^2 (C_2 + N^2 e^{-C_3 N})\, \sigma^{-2}(\bar{L})$. From here, from (14) and (16), by Theorem A we obtain (12). The theorem is proved.

LITERATURE CITED

1. W. R. van Zwet, "A Berry – Esseen bound for symmetric statistics," Z. Wahrscheinlichkeitstheorie Verw. Gebiete, 66, No. 3, 425-440 (1984).
2. S. Bjerve, "Error bounds for linear combinations of order statistics," Ann. Statist., 5, No. 2, 357-369 (1977).
3. S. M. Stigler, "Linear functions of order statistics with smooth weight functions," Ann. Statist., 2, No. 4, 676-693 (1974).
4. H. A. David, Order Statistics, Wiley, New York (1970).

SOME REMARKS ON SUMS OF DEPENDENT COMPONENTS

L. V. Kir'yanova and V. I. Rotar'

This note contains some very easily proven assertions which, in the opinion of the authors provide some useful illustrations to the question of possible limit distributions for sums of, generally speaking, dependent components. Some simplified variants of these assertions were discussed previously in [1].

Let n = 1, 2,...

$$\{X_{1n}, X_{2n}, \ldots\} \qquad (1)$$

be series of sequences of random variables (r.v.) taking values ±1.

Considering the sum $S_n = X_{1n} + \ldots + X_{nn}$, we can asume without any loss of generality that the r.v. in (1) are symmetrically dependent (s.d.) – otherwise we can turn to the series of sequences obtained from (1) by such a random permutation of X_{jn} that all versions are equally distributed. In the sequel we assume the r.v. in (1) to be s.d. and consequently equally distributed.

Let F_n be the distribution of the r.v. $\zeta_n = S_n/\sqrt{n}$.

Proposition 1. For any distribution F there exists a series of sequences of symmetrically dependent random variables (1) such that

$$\mathbf{M}X_{1n} = 0 \text{ for all } n = 1, 2, \ldots, \qquad (2)$$

and

Moscow Institute of Electrical Engineering. Translated from Matematicheskie Zametki, Vol. 42, No. 5, pp. 747-750, November, 1987. Original article submitted January 21, 1987.

0001-4346/87/4256-0910$12.50