September 1964

A CONCENTRATION INEQUALITY FOR THE SUM OF INDEPENDENT SYMMETRICALLY DISTRIBUTED RANDOM VARIABLES.

G. Kallianpur

Technical Report No. 50

University of Minnesota*
Minneapolis, Minnesota

*Research supported by U. S. Army Research Office - Durham - under Grant DA-ARO-D-31-124-G562.

Bounds on the Lévy concentration function [1] have been obtained by Kolmogorov ([2], [3]) and applied by him in [2]. A strengthened version of Kolmogorov's result has recently been given by B. A. Rogozin [4] who bases his proof on an inequality stated below. It is of interest in the study of central limit theorems (see [1]) to obtain bounds on the concentration function of the sum of independent symmetrically distributed random variables. We shall show how Rogozin's inequality can also be used to derive such a bound.

The concentration function of a random variable X is defined as

(1) $$Q_X(a) = \max_{-\infty<x<\infty} \text{Prob} \left\{ x \leq X \leq x+a \right\}$$

Rogozin's inequality is the following:

Let X_k $(k=1,\dots,n)$ be independent random variables such that

$P(X_k = x_k) = P(X_k = -x_k) = \frac{1}{2}$ where $x_k \geq l_k > 0$.

If $S = \sum_{k=1}^{n} X_k$ and $L \geq \max_k l_k$, then

(2) $$Q_S(L) \leq C.L \left(\sum_k l_k^2\right)^{-\frac{1}{2}}$$ when C is a positive absolute constant.

From the above inequality it is easy to deduce that (2) holds for independent symmetric random variables X_k whose probability mass is concentrated entirely outside the interval $(-l_k, l_k)$.

<u>Lemma</u>. Let X_k $(k=1,\dots,n)$ be independent random variables, symmetrically distributed about zero and such that $|X_k| \geq l_k$ with probability one. Then for $L \geq \max_k l_k$, $Q_S(L)$ has the same bound as in (2).

To prove the lemma, observe that if we define $T = \sum_{k=1}^{n} \epsilon_k Z_k$ where the random variables $\{\epsilon_k\}$, $\{Z_k\}$ are mutually independent such that Z_k is distributed as $|X_k|$ and $\epsilon_k = +1$ or -1 with equal probability, then $\mathcal{L}(T)$, the probability law or distribution of T equals $\mathcal{L}(S)$. Let z_k $(k=1,\ldots,n)$ be fixed numbers with $z_k \geq l_k > 0$ and let $T' = \sum_{k=1}^{n} \epsilon_k z_k$.

We may then apply Rogozin's result to T' and obtain

$$Q_{T'}(L) \leq C\, L \left(\sum_{k=1}^{n} l_k^2\right)^{-\frac{1}{2}} .$$

An elementary argument now shows that $Q_T(L)$ has also the same bound. The lemma follows since $Q_T(L) = Q_S(L)$. We are now in a position to prove our main result.

<u>Theorem</u>. Let X_k $(k=1,..,n)$ be independent random variables symmetrically distributed about zero. If l_k are arbitrary positive numbers and $L \geq \max_k l_k$, then.

(3) $$Q_S(L) \leq AL\, \delta^{-\frac{1}{2}},$$ where A is a positive absolute constant and

$$\delta = \sum_{k=1}^{n} l_k^2\, P(|X_k| \geq l_k).$$

The following corollaries are of interest in connection with the approximation by infinitely divisible distributions to the law of S. Upon setting $l_k = 1$ in (3) we immediately obtain

Corollary 1. $$Q_S(L) \leq A \frac{L}{l} \left\{\sum_{k=1}^{n} P(|X_k| \geq l)\right\}^{-\frac{1}{2}} \quad (L \geq l > 0).$$

Corollary 2. If the X_k's are no longer assumed to be symmetric we have the following inequality:

(4) $$Q_S(L) \le C\left(\frac{L}{l}\right)^{\frac{1}{2}}\left[\sum_{k=1}^{n} P(|X_k - m_k| \ge 2l)\right]^{-\frac{1}{4}}, \quad (L \ge l > 0)$$

where C is a positive constant and m_k is the median of X_k.

In deducing Corollary 2 from Corollary 1 we make use of the easily verifiable facts that if ξ and η are independent, identically distributed random variables with median m, then

$$P(|\xi-\eta| \ge l) \ge \tfrac{1}{2} P(|\xi - m| \ge l) \quad \text{and} \quad Q_\xi^2(l) \le Q_{\xi-\eta}(2l).$$

In the case of symmetric random variables Kolmogorov's inequality follows from Corollary 1. Since $P(|X_k| \ge l) \ge 1 - Q_{X_k}(2l)$ we obtain at once the inequality

$$Q_S(L) \le A.L.l^{-1}\left[\sum_{k=1}^{n}\left\{1 - Q_{X_k}(2l)\right\}\right]^{-\frac{1}{2}}.$$

In the general case the following inequality similarly follows from (4).

$$Q_S(L) \le A\, L^{\frac{1}{2}}.l^{-\frac{1}{2}}\left[\sum_{k=1}^{n}\left\{1 - Q_{X_k}(2l)\right\}\right]^{-\frac{1}{4}}.$$

We now give the proof of the theorem. Let ξ_k $(k=1,..,n)$ be independent random variables such that $P(\xi_k=1) = 1-P(\xi_k=0) = p_k$ where $p_k = P(|X_k| \ge l_k)$. Also introduce independent random variables U_k and V_k such that the distribution of U_k is equal to the conditional distribution of X_k given $|X_k|<l_k$, and the distribution of V_k is the conditional distribution of X_k given $|X_k| \ge l_k$. If we set

$$Z = \sum_k \left\{\xi_k V_k + (1-\xi_k)\, U_k\right\}$$, then clearly we have $\mathcal{L}(Z) = \mathcal{L}(S)$.

Denote by E the event $\left\{ \xi_{k_i}=1,\ \xi_k=0 \text{ if } k \ne k_i,\ i=1,..,\nu \text{ and } U_k=u_k \text{ for } k=1,..,n\right\}$; setting $a = \sum_{k \ne \text{any } k_i} u_k$ we find that the conditional probability

$$P[\ x \leq Z \leq x+L \mid E\ ] = P[\ x-a \leq \sum_{i=1}^{\nu} V_{k_i} \leq x-a+L\ ].$$

Hence

$$P[\ x \leq Z \leq x+L\ ] = \int P[\ x-a \leq \sum_{i=1}^{\nu} V_{k_i} \leq x-a+L\ ]\ dF$$

where F is the joint distribution of $\{U_k\}$ and $\{\xi_k\}$ and the integration is over all possible values $\{u_k\}$ of the U_k and of the variables $\{\xi_k\}$.

We bound the integrand by $C \cdot L\ [\ \sum_{i=1}^{\nu} 1_{k_i}^2\ ]^{-\frac{1}{2}}$ over the region $\sum 1_k^2\ \xi_k^2 > \frac{\delta}{2}$ (which we may do on account of the lemma proved above) and by 1 over the complementary region $\sum 1_k^2\ \xi_k^2 \leq \delta/2$. We then obtain

$$Q_Z(L) \leq C\ L\,(\tfrac{\delta}{2})^{-\frac{1}{2}} + P[\sum_{k=1}^{n} 1_k^2\ \xi_k^2 \leq \tfrac{\delta}{2}].$$

Since, by definition, $\delta = E(\sum 1_k^2\ \xi_k^2)$ and since $\mathrm{Var}\ (\sum 1_k^2\ \xi_k^2) \leq L^2\ \delta$ we have

$$P[\sum 1_k^2\ \xi_k^2 \leq \tfrac{\delta}{2}] \leq P[\ |\sum 1_k^2\ \xi_k^2 - E(\sum 1_k^2\ \xi_k^2)| \geq \tfrac{\delta}{2}\ ]$$

$$\leq 4L^2\ \delta^{-1}.$$

Hence

$$(5) \qquad Q_Z(L) \leq \sqrt{2}C\ L\ \delta^{-\frac{1}{2}} + 4L^2\ \delta^{-1}.$$

The inequality $Q_Z(L) \leq (\sqrt{2}C + 4)\ L\ \delta^{-\frac{1}{2}}$ now follows from (5) if $L\ \delta^{-\frac{1}{2}} \leq 1$ and is trivially satisfied if $L\ \delta^{-\frac{1}{2}} \geq 1$. This completes the proof of the theorem.

References

1. Paul Lévy, *Théorie de l'addition des variables aléatoires*, Gauthier-Villars, Paris, 1937.

2. A. N. Kolmogorov, Two uniform limit theorems for sums of independent random variables, *Teoriya Veroyatnosteii i ee Primeneniya*, 1 (1956), pp. 426-436.

3. A. N. Kolmogorov, Sur les propriétés des fonctions de concentrations de M. P. Lévy, *Annales de l'Institut de Henri Poincaré*, 16 (1958), pp. 27-34.

4. B. A. Rogozin, On the increase of dispersion of sums of independent random variables, *Teoriya Veroyatnosteii i ee Primeneniya*, 6 (1961), pp. 105-108.