ON KOLMOGOROFF'S INEQUALITY

by

A. RÉNYI

§ 0. Notations

Let \( S = [\Omega, \mathcal{A}, \mathbf{P}] \) be a probability space, i.e. \( \Omega \) a set (the set of elementary events), \( \mathcal{A} \) a \( \sigma \)-algebra of subsets of \( \Omega \), and \( \mathbf{P} \) a probability measure on \( \mathcal{A} \). We shall denote the elements of \( \mathcal{A} \) (called random events) by capital letters and we denote by \( \mathbf{P}(A) \) the probability of the event \( A \in \mathcal{A} \). Random variables (i.e. functions defined on \( \Omega \) and measurable with respect to \( \mathcal{A} \)) will be denoted by greek letters. We denote by \( \mathbf{M}(\xi) \) the mean value and by \( \mathbf{D}^2(\xi) \) the variance of the random variable \( \xi \). We denote by \( \mathbf{P}(A \mid B) \) the conditional probability of the event \( A \) with respect to the event \( B \).

§ 1. Introduction

In the present paper we deal with the celebrated inequality of A. N. KOLMOGOROFF ([1]) according to which if \( \xi_1, \xi_2, \ldots, \xi_n \) are independent random variables with mean value 0 and with finite variances \( d_k^2 = \mathbf{D}^2(\xi_k) \) \((k = 1, 2, \ldots, n)\) then putting

\[
\zeta_k = \xi_1 + \xi_2 + \ldots + \xi_k \quad (k = 1, 2, \ldots, n)
\]

and

\[
D_k^2 = d_1^2 + d_2^2 + \ldots + d_k^2 = \mathbf{D}^2(\zeta_k) \quad (k = 1, 2, \ldots, n)
\]

one has for any \( \lambda > 1 \)

\[
\mathbf{P}(\text{Max } 1 \leq k \leq n | \zeta_k | \geq \lambda D_n) \leq \frac{1}{\lambda^2}.
\]

As well known, this inequality is extremely useful in proving the strong law of large numbers, the law of the iterated logarithm and other related theorems.

In § 2 we generalize this inequality by considering instead of (3) the conditional probability of the inequality \( \text{Max } 1 \leq k \leq n | \zeta_k | \geq \lambda D_n \) with respect to some condition \( A \) having positive probability. We prove the following

**Theorem.** If the random variables \( \xi_k \) are independent, have zero means, finite variances \( d_k^2 \) and finite fourth moments \( f_k^4 = \mathbf{M}(\xi_k^4) \) \((k = 1, 2, \ldots, n)\), then if \( \zeta_k \) resp. \( D_k \) are defined by (1) resp. (2) and we put

\[
F_n^4 = f_1^4 + \ldots + f_n^4
\]

\[
\mathbf{P}(A) = \frac{1}{\lambda^2}
\]

for some constant \( \lambda > 1 \).
then one has for any \( \lambda > 1 \), and for any event \( A \) with \( P(A) > 0 \),

\[
P(\text{Max}_{1 \leq k \leq n} |\xi_k| \geq \lambda D_n | A) \leq \frac{2 + \sqrt{3 + \left(\frac{F_n}{D_n}\right)^4}}{\lambda^2 P(A)}.
\]

It is known that in the proof of Kolmogoroff’s inequality the supposition of independence of the random variables \( \xi_k \) can be replaced by the weaker supposition that the conditional mean value of \( \xi_k \) given \( \xi_1, \ldots, \xi_{k-1} \) is identically equal to 0, that is that the variables \( \xi_k \) form a martingale (see [2]). It will be seen from the proof that the same supposition is sufficient for the validity of our Theorem.

§ 2. Proof of the generalization of Kolmogoroff’s inequality

In this § we shall prove the Theorem formulated in § 1.

Let \( A \) be an arbitrary event, having positive probability \( P(A) > 0 \). Let \( a \) denote the indicator of \( A \), i.e. a random variable, which is equal to 1 on the set \( A \) (i.e. if the event \( A \) takes place) and equal to 0 on the complementary set \( \Omega - A \) (i.e. if the event \( A \) does not take place). Let \( B_k \) \((k = 1, 2, \ldots, n)\) denote the event that \( |\xi_1|, |\xi_2|, \ldots, |\xi_k| \) which is not less than \( \lambda D_n \), i.e. \( B_k \) takes place if \( |\xi_1| < \lambda D_n, \ldots, |\xi_k| < \lambda D_n \) and \( |\xi_k| \geq \lambda D_n \). Let \( \beta_k \) denote the indicator of \( B_k \). Then clearly

\[
0 \leq \sum_{k=1}^{n} \beta_k \leq 1, \text{ further } \beta_k \beta_l = 0 \text{ if } k < l
\]

and \( \beta_k \) depends only on \( \xi_1, \ldots, \xi_k \), and thus is independent of \( \xi_{k+1}, \ldots, \xi_n \). Let finally \( C_n \) denote the event \( \text{Max}_{1 \leq k \leq n} |\xi_k| \geq \lambda D_n \), that is \( C_n \) is the union of the sets \( B_1, \ldots, B_n \). We have clearly

\[
M(\xi_n^2 a) \geq \sum_{k=1}^{n} M(\xi_k^2 a \beta_k) = \sum_{k=1}^{n} M(\xi_k^2 \beta_k a) + 2 \sum_{k=1}^{n} M(\xi_k \beta_k (\xi_n - \xi_k) a) + \sum_{k=1}^{n} M((\xi_n - \xi_k)^2 a \beta_k) \geq \sum_{k=1}^{n} M(\xi_k^2 \beta_k a) + 2 \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} M(\xi_k \beta_k \xi_j a).
\]

Now put

\[
\eta_{kj} = \xi_k \beta_k \xi_j \quad (1 \leq k \leq n - 1; \; k + 1 \leq j \leq n).
\]

Clearly we have, if \( 1 \leq k < j < h \leq n \)

\[
M(\eta_{kj} \eta_{kh}) = M(\xi_k^2 \beta_k \xi_j \xi_h) = M(\xi_k^2 \beta_k) M(\xi_j) M(\xi_h) = 0,
\]

further if \( k < l, k + 1 \leq j, l + 1 \leq h \) then owing to \( \beta_k \beta_l = 0 \) one has

\[
M(\eta_{kj} \eta_{lh}) = 0.
\]
Further

(9c) \[ M(\eta_{kj}^2) = M(\zeta_k^2 \beta_k^2) \, d_j^2. \]

Thus the system

(10) \[ \eta_{kj}^* = \frac{\eta_{kj}}{d_j \sqrt{M(\zeta_k^2 \beta_k)}}, \]

is orthonormal. It follows by Bessel's inequality that

(11) \[ \left| \sum_{k=1}^{n} \sum_{j=k+1}^{n} M(\eta_{kj} \alpha) \right| = \left| \sum_{k=1}^{n} \sum_{j=k+1}^{n} d_j \sqrt{M(\zeta_k^2 \beta_k)} \cdot M(\eta_{kj} \alpha) \right| \leq \sqrt{M(\alpha^2) \cdot \left( \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} d_j^2 \right)}. \]

Taking into account that

(12) \[ M(\zeta_n^2 \beta_k) - M(\zeta_k^2 \beta_k) = M((\xi_n - \xi_k)^2 \beta_k) + 2 M(\xi_k \beta_k (\xi_n - \xi_k)) \]

and \[ M(\xi_k \beta_k (\xi_n - \xi_k)) = 0, \]

it follows that

(13) \[ M(\zeta_k^2 \beta_k) \leq M(\zeta_n^2 \beta_k). \]

Thus

(14) \[ \sum_{k=1}^{n-1} M(\zeta_k^2 \beta_k) \left( \sum_{j=k+1}^{n} d_j^2 \right) \leq D_n^2 \cdot \sum_{k=1}^{n-1} M(\zeta_k^2 \beta_k) \leq D_n^2 M(\zeta_n^2) = D_n^4. \]

Thus we obtain finally, taking into account that \[ M(\alpha^2) = P(A), \]

that

(15) \[ M(\zeta_n^2 \alpha) \geq \sum_{k=1}^{n} M(\zeta_k^2 \beta_k \alpha) - 2 D_n^2 \sqrt{P(A)}. \]

On the other hand, if \( \beta_k = 1 \), one has \( \zeta_k^2 \geq \lambda^2 D_n^2. \)

Thus

(16) \[ \sum_{k=1}^{n} M(\zeta_k^2 \beta_k \alpha) \geq \lambda^2 D_n^2 M \left( \alpha \left( \sum_{k=1}^{n} \beta_k \right) \right) = \lambda^2 D_n^2 P(AC_n) \]

where \( C_n \) stands for the event \( \max_{1 \leq k \leq n} |\xi_k| \geq \lambda D_n \). We obtain from (15) and (16)

(17) \[ P(AC_n) \lambda^2 D_n^2 \leq M(\zeta_n^2 \alpha) + 2 D_n^2 \sqrt{P(A)}. \]

On the other hand,

(18) \[ M(\zeta_n^2 \alpha) \leq \sqrt{P(A)} M(\zeta_n^4) . \]

As clearly

(19) \[ M(\zeta_n^4) \leq F_n^4 + 3 D_n^4. \]
we obtain from (17), (18) and (19)

\begin{equation}
\mathbf{P}(C_n | A) = \frac{\mathbf{P}(AC_n)}{\mathbf{P}(A)} \leq \frac{1}{\lambda^2 \sqrt{\mathbf{P}(A)}} \left( 2 + \sqrt{3 + \frac{F_n^4}{D_n^4}} \right)
\end{equation}

Thus (5) is proved.

Our theorem may e.g. be used to obtain an estimate for

\[ \mathbf{P}( \text{Max}_{1 \leq k \leq v_n} | \xi_k | \geq l_n) \]

where \( v_n \) is a random variable, which may depend on the variables \( \xi_k \). Let \( v_n \) take on the values \( n + 1, n + 2, \ldots, n + s \) with the corresponding probabilities \( p_1, p_2, \ldots, p_s \). If \( A_i \) denotes the event \( v_n = n + l \) (\( l = 1, 2, \ldots, s \)) one has by Theorem 1, in the case \( |z_k| \leq 1 \) (\( k = 1, 2, \ldots, n \))

\begin{equation}
\mathbf{P}( \text{Max}_{1 \leq k \leq v_n} | \xi_k | \geq l_n) = \sum_{i=1}^{s} \mathbf{P}(\text{Max}_{1 \leq k \leq n+l} | \xi_k | > l_n | A_i) \mathbf{P}(A_i) \leq \frac{4}{l_n^2} \sum_{i=1}^{s} \mathbf{P}(A_i) D_{n+l}^2 \leq \frac{4 D_{n+s}^2}{l_n^2} \sqrt{s}.
\end{equation}

Thus we obtain, putting \( t_n = \lambda D_{n+s} \), the following

**Corollary.** If \( \xi_1, \ldots, \xi_n \) are independent random variables, with mean value zero and satisfying \( |\xi_k| \leq 1 \), further if \( v_n \) is a random variable capable of the values \( n + 1, \ldots, n + s \) and if \( D_k^2 \) denotes the variance of \( \xi_k = \xi_1 + \xi_2 + + \ldots + \xi_k \), we have for \( \lambda < 2 \sqrt{s} \)

\begin{equation}
\mathbf{P}(\text{Max}_{1 \leq k \leq v_n} | \xi_k | > \lambda D_{n+s}) < \frac{4 \sqrt{s}}{\lambda^2}.
\end{equation}

(Received June 18, 1961.)

**REFERENCES**


**О НЕРАВЕНСТВЕ А. Н. КОЛМОГОРОВА**

**A. RÉNYI**

**Резюме**

Доказывается следующее обобщение известного неравенства А. Н. Колмогорова. Пусть \( \xi_k \) (\( k = 1, 2, \ldots \)) независимые случайные величины, имеющие математическое ожидание 0, конечные дисперсии \( d_k \) и четьерые моменты \( f_k^4 \). Положим \( \xi_k = \xi_1 + \xi_2 + \ldots + \xi_k \), \( D_k^2 = d_1^2 + d_2^2 + \ldots + d_k^2 \),
\( F_n = f_1 + \ldots + f_n \). Пусть \( A \) произвольное событие с положительной вероятностью \( P(A) > 0 \). Тогда имеет место для всех \( \lambda > 1 \)

\[
P(\max_{1 \leq k \leq n} |z_k| \geq \lambda D_n | A) \leq \frac{2 + \sqrt{3 + \left(\frac{F_n}{D_n}\right)^4}}{\lambda^2 \sqrt{P(A)}}.
\]