ON SOME INEQUALITIES CONNECTED TO THE STRONG LAW
OF LARGE NUMBERS*

A. N. KOLMOGOROV

(Translated by A. Zakharov)

DOI. 10.1137/S0040585X9798035X

The note that is published here was written by A. N. Kolmogorov more than 40 years ago
(the author dates it April 1962). At that time I was a graduate student of Andrei Nikolaevich
[Kolmogorov] and was investigating the possibility of generalizing and amplifying the well-
known Chebyshev inequality. Some of my results I presented at Kolmogorov’s seminar at
the Moscow State University. In 1962, on Bernstein’s request, I wrote a commentary on
his work “On some modifications of the Chebyshev inequality” (this paper can be found in
the fourth volume of Bernstein’s collected works). Andrei Nikolaevich approached my work
(which was published by the MphTI press the same year) with interest. During my next
visit to Komarovka, when I was reporting the work in progress on my doctoral thesis, he
gave me a short manuscript and asked me to read it. The idea of that note was close to the
one contained in my published work and in my commentary on Bernstein’s work. After some
time I asked Andrei Nikolaevich whether he planned to prepare that note for publication.
He said that he did not plan to do so in the near future. The manuscript remained in my
archive.

This note does not contain a fundamental result, as was usually the case with most
of Kolmogorov’s other works. However, it presents an opportunity of learning what this
great scientist thought and worked on during a fruitful period of his career. In this respect,
this short work is certainly valuable to both experts and new practitioners in the field of
probability theory.

The manuscript contains formula (9), where \( \varepsilon > 0 \) and \( p \in (0, 1) \) and \( \mu_n \) denotes the
number of successes in \( n \) Bernoulli trials with the probability of success \( p \). For \( p = \frac{1}{2} \) the
manuscript contains the more precise inequality (8).

It should be mentioned that similar inequalities appear in several textbooks published
at later dates: A. A. Borovkov, Probability Theory, Gordon and Breach, United Kingdom,

In particular, the first textbook gives the following inequalities:

\[
\mathbb{P} (\mu_n - np \geq \varepsilon) \leq e^{-nH(p+\varepsilon/n)}, \quad \mathbb{P} (\mu_n - np \leq -\varepsilon) \leq e^{-nH(p-\varepsilon/n)},
\]

where \( H \) is some function that satisfies \( H(x) \geq 2x^2 \). The second textbook gives the inequality
\( \mathbb{P} (|\mu_n/n - p| \geq \varepsilon) \leq 2e^{-2n\varepsilon^2} \). A more careful analysis of Kolmogorov’s technique may lead
to the inequality \( \mathbb{P} (\sup_{k \geq n} |\mu_k/k - p| \geq \varepsilon) \leq 2e^{-2n\varepsilon^2} \) for all \( p \in (0, 1) \).

Finally, I express gratitude to V. Yu. Korolev and V. M. Kruglov for their help in
preparing this manuscript for publication.

V. M. Zolotarev

*From archival materials.
1. Как будем считать два числа равными:

a) Если при $k = 1, 2, \ldots, n$

$$M(\xi_k | \xi_1, \xi_2, \ldots, \xi_{k-1}) \leq \theta_k,$$

где $\theta_k$ контаненты,

$$M(\xi_1, \xi_2, \ldots, \xi_n) \leq \theta_1 \theta_2 \cdots \theta_n.$$

b) $P(\xi \geq E) \leq e^{-a \xi} M \leq a^\xi$

при выбрано $a > 0$.

**Теорема.** Пусть для последовательности случайных

величин

$$\xi_1, \xi_2, \ldots, \xi_n, \ldots$$

в контаненты $a > 0$ при каждом $n = 1, 2, \ldots$

$$M(e^{a \xi_n | \xi_{n-1}, \xi_{n-2}, \ldots}) \leq \theta_n,$$

где $\theta_n$ некоторая константа. Тогда

$$\log P\left(\frac{\xi_1 + \xi_2 + \cdots + \xi_n}{N} > c\right) \leq N \log \theta - ax$$

при любых $c > 0$.

3. В заключение рассмотрим схему формулы: последовательность неизвестных значений $a$ постоянной величины $\theta$ некоторых значений, когда

$$\xi_1 = 1$$ в случае независимого значения $N$-го независимого и

$$\xi_1 = 0$$ в случае отрицательного,

$$M_{\xi_1} = \xi_1 + \xi_2 + \cdots + \xi_n.$$

В этой системе теорема имеет следующий вид:

$$\log P\left(\frac{\xi_1 + \xi_2 + \cdots + \xi_n}{N} > \rho, \epsilon\right) \leq N\left[-\log(p \rho^N + q) - a(p + \epsilon)\right].$$

Выбор наиболее выгодного следует при данных $\rho, \epsilon$, и $N$ значения $a$ приводит к довольно большим вычислениям.
1. Our argument will be based on two well-known facts:
   a) If for \( k = 1, 2, \ldots, n \) we have
      \[
      M(\xi_{k-1}, \xi_{k-2}, \ldots, \xi_1) \leq b_k,
      \]
      where \( b_k \) are constants, then
      \[
      M(\xi_1, \xi_2, \ldots, \xi_n) \leq b_1 b_2 \cdots b_n,
      \]
   b) \( P(\xi \geq z) \leq e^{-az} M e^{ac} \)

   for every \( a \geq 0 \).

THEOREM. Suppose that for a sequence of random variables \( \xi_1, \xi_2, \ldots, \xi_n, \ldots \) and a constant \( a \geq 0 \), for each \( n = 1, 2, \ldots \), we have
\[
M(e^{a\xi_n} | \xi_{n-1}, \xi_{n-2}, \ldots, \xi_1) \leq b,
\]
where \( b \) is some constant. Then
\[
\log P\left( \sup_{n \geq N} \frac{\xi_1 + \xi_2 + \cdots + \xi_n}{n} \geq c \right) \leq N \log b - ac
\]
for any \( c \) and \( N \).

We shall prove (1) for a finite sequence
\[
\xi_1, \xi_2, \ldots, \xi_n, \ldots, \xi_s.
\]

The limit pass to the case of an infinite sequence is not too difficult. In the finite sequence case the upper bound in (1) is achieved and (1) can be rewritten as
\[
\log P\{E(n): N \leq n \leq s, \xi_1 + \xi_2 + \cdots + \xi_n - nc \geq 0\} \leq N \log b - ac.
\]

If \( \log b - ac \geq 0 \), then inequality (1) is trivial since \( \log P \leq 0 \) always holds. For that reason it will be assumed that
\[
\log b - ac \leq 0.
\]

Let \( \nu \) be the least \( n \), \( N \leq n \leq s \), such that
\[
\xi_1 + \xi_2 + \cdots + \xi_n - nc \geq 0.
\]

If it does not exist, let \( \nu = s \). It is worth noting that knowing \( \xi_{n-1}, \xi_{n-2}, \ldots, \xi_1 \) is enough to see whether \( n \leq \nu \) or not.

Put
\[
\xi_n' = \begin{cases} 
\xi_n - c & \text{if } n \leq \nu, \\
0 & \text{if } n > \nu.
\end{cases}
\]

Then
\[
\Delta_n = M(e^{a\xi_n'} | \xi_{n-1}', \ldots, \xi_1') \begin{cases} 
\leq be^{-ac} & \text{if } n \leq \nu, \\
1 & \text{if } n > \nu.
\end{cases}
\]

\[\text{— V. Zolotarev.}\]
From (2) it follows that
\[ \Delta_n \leq 1. \]

However, since \( \nu \geq N \), for \( n \leq N \), we have
\[ \Delta_n \leq be^{-ac}. \]

Therefore, for
\[ \zeta = \xi'_1 + \xi'_2 + \cdots + \xi'_s = \xi_1 + \xi_2 + \cdots + \xi_{\nu} - \nu c \]
we get
\[ M e^{ac} \leq (be^{-ac})^N, \quad \log P(\zeta \geq 0) \leq N(\log b - ac). \]

But
\[ P(\zeta \geq 0) = P(\xi_1 + \xi_2 + \cdots + \xi_{\nu} - \nu c \geq 0) \]
is the probability estimated in the theorem. The proof is complete.

2. Suppose that the random variables of the sequence
\[ \xi_1, \xi_2, \ldots, \xi_n, \ldots \]
satisfy the conditions
\[ \begin{align*}
M(\xi_n \mid \xi_{n-1}, \xi_{n-2}, \ldots, \xi_1) &\leq 0, \\
M(\xi_n^2 \mid \xi_{n-1}, \xi_{n-2}, \ldots, \xi_1) &\leq \sigma^2, \\
\xi_n &\leq l.
\end{align*} \]

Then (\( \xi_{n-1}, \ldots, \xi_1 \) were omitted from the computation) we have
\[ M(e^{a\xi_n} \mid \xi_{n-1}, \ldots, \xi_1) = M\left(1 + a\xi_n + \frac{a^2\xi_n^2}{2} + \frac{a^3\xi_n^3}{6} e^{\theta l}\right), \]
where \( \theta \leq 1 \), and for \( al \leq 1 \) we have
\[ Me^{a\xi_n} \leq 1 + \frac{a^2\sigma^2}{2} + \frac{a^3\sigma^2l}{6} e \leq 1 + \frac{a^2\sigma^2}{2} (1 + al), \]
\[ \log Me^{a\xi_n} \leq \frac{a^2\sigma^2}{2} (1 + al). \]

Using the theorem proven in section 1 and (4), for \( al \leq 1 \) we obtain
\[ \log P\left(\sup_{n \geq N} \frac{\xi_1 + \cdots + \xi_n}{n} \geq \varepsilon\right) \leq N\left[\frac{a^2\sigma^2}{2} (1 + al) - a\varepsilon\right]. \]

Setting
\[ a = \frac{\varepsilon}{\sigma^2}, \]
we can see that

\[ \log P \left( \sup_{n \geq N} \frac{\xi_1 + \cdots + \xi_n}{n} \geq p + \varepsilon \right) \leq -\frac{N \varepsilon^2}{2 \sigma^2} \left( 1 - \frac{\varepsilon l}{\sigma^2} \right) \text{ for } \varepsilon l \leq \sigma^2. \]  

(6)

3. Finally, consider the Bernoulli scheme: a sequence of random trials with a constant probability of success \( p \). As usual, set \( \xi_n = 1 \) if the \( n \)th trial is successful and \( \xi_n = 0 \) if it is not. Denote

\[ \mu_n = \xi_1 + \xi_2 + \cdots + \xi_n. \]

In this setting the theorem leads to the inequality\(^2\)

\[ \log P \left( \sup_{n \geq N} \frac{\mu_n}{n} \geq p + \varepsilon \right) \leq N \left[ \log(pe^a + q) - a(p + \varepsilon) \right]. \]  

(7)

Choosing the optimal value of \( a \) for given \( p, \varepsilon \), and \( N \) requires complicated computations. However, one can consider special cases of inequality (7). For \( p = \frac{1}{2} \) one can obtain the very simple inequality

\[ \log 2 \left( \sup_{n \geq N} \frac{\mu_n}{n} \geq p + \varepsilon \right) \leq -2N \varepsilon^2. \]  

(8)

The following inequality is true for any \( p \), \( 0 \leq p \leq 1 \):

\[ \log 2 \left( \sup_{n \geq N} \frac{\mu_n}{n} \geq p + \varepsilon \right) \leq -2N \varepsilon^2(1 - \varepsilon). \]  

(9)

\(^2\)Where \( q = 1 - p \). — V. Zolotarev.