Moment Inequalities for Maxima of Partial Sums in Probability with Applications in the Theory of Orthogonal Series

Ferenc Móricz

Dedicated to the memory of Paul Erdős on the 100th anniversary of his birthday

Let \((X_k : k = 1, 2, \ldots)\) be a sequence of random variables. It is not assumed that these \(X_k\)'s are mutually independent or that they are identically distributed. We set

\[ S_{b,n} := \sum_{k=b+1}^{b+n} X_k \quad \text{and} \quad M_{b,n} := \max_{1 \leq k \leq n} |S_{b,k}|, \]

where \(b \geq 0\) and \(n \geq 1\) are integers. Thus, \(M_{b,n}\) is the largest magnitude for the \(n\) consecutive partial sums formed from the \(n\) consecutive \(X_k\)'s beginning with \(X_{b+1}\). Furthermore, for each vector \(X_{b,n} := (X_{b+1}, X_{b+2}, \ldots, X_{b+n})\) of \(n\) consecutive \(X_k\)'s, let \(F_{b,n}\) denote their joint distribution function.

The object of this paper is to provide bounds on \(E(M_{b,n}^p)\) in terms of given bounds on \(E|S_{b,n}|^p\), where \(p > 1\) and \(E\) is the symbol for the expected value of random variables. We emphasize that it is not assumed that these \(X_k\)'s are independent. The only restrictions on the dependence will be those imposed on the assumed bounds on \(E|S_{b,n}|^p\). These assumed bounds are guaranteed under a suitable dependence restriction, for example, martingale differences, weak multiplicativity of finite order, orthonormality, etc.

Bounds on \(E(M_{b,n}^p)\) are of use in deriving bounds on the tail distribution of the maximum of certain partial sums in order to study convergence properties of \(S_n := S_{0,n}\) as \(n \to \infty\). For development of such results under various dependence restrictions, the theorems of this paper reduce the problem of placing appropriate bounds on \(E(M_{b,n}^p)\) to the easier problem of placing appropriate bounds on \(E|S_{b,n}|^p\).

The problem posed above is treated essentially in a setting close to that of Serfling [26], whose results are contained as special cases in our Theorems 1 and 3 stated in the next two sections.

The Generalized Rademacher-Menshov Maximal Moment Inequality

In the following, the symbol \(g(F_{b,n})\) denotes a nonnegative functional depending on the joint distribution function \(F_{b,n}\) of the vector \(X_{b,n}\), where \(b \geq 0\) and \(n \geq 1\) are integers. Throughout the paper, we will assume that \(g(F_{b,n})\) possesses the following property of rather general nature:

\[ g(F_{b,k}) + g(F_{b+k,\ell}) \leq g(F_{b+k+\ell}) \]

for all integers \(b \geq 0\) and \(1 \leq k < k + \ell\). It may be called the superadditivity property of \(g(F_{b,n})\). In case the symbol “≤” is replaced by “=”, in (1), we may say that \(g(F_{b,n})\) possesses the commonly known additivity property. Examples are

\[ g(F_{b,n}) := n^\alpha \quad \text{or} \quad g(F_{b,n}) := \sum_{k=b+1}^{b+n} a_k^2, \]

where \(\alpha \geq 1\) or \((a_k)\) is a sequence of real numbers. In most cases, \(a_k^2 = \text{Var}(X_k)\), is the finite variance of \(X_k\), but this remark plays no role in the theorems stated in the sequel.
The next maximal moment inequality, indicated in the subtitle above, was proved by the present author [13, Theorem 3].

**Theorem 1.** Suppose that there exists a functional \(g(F_{b,n})\) satisfying condition (1). If
\[
E|S_{b,n}|^p \leq g(F_{b,n})
\]
for all \(b \geq 0\) and \(n \geq 1\), where \(p > 1\), then we have
\[
E(M_{b,n}^p) \leq (\log 2n)^p g(F_{b,n})
\]
for all \(b \geq 0\) and \(n \geq 1\).

Here and in the sequel, by “\(\log\)” we denote the logarithm to the base 2.

**Proof.** It goes by induction with respect to \(n\). Since \(M_{b,1} = |S_{b,1}|\), (3) is obvious for \(n = 1\).

Now, we assume as the induction hypothesis that (3) holds for each integer \(b\) and for each integer less than \(n\), where \(n > 1\), and we will prove it for \(n\) itself. To this effect, we make use of the so-called bisection technique. To this effect, let \(m\) be the integer part of \((n + 2)/2\). It is easy to check that \(n = 2m - 2\) or \(2m - 1\), while \(n - m = m - 2\) or \(m - 1\).

To sum up, in either case we have
\[
2(m - 1) \leq n \quad \text{and} \quad n - m \leq m - 1.
\]

Let \(b \geq 0\) be arbitrary. For \(m \leq k \leq n\), we clearly have
\[
|S_{b,k}| \leq |S_{b,m}| + |S_{b,m,k-m}|,
\]
whence for any such \(k\) it follows that
\[
|S_{b,k}| \leq |S_{b,m}| + M_{b,m,n-m}.
\]
Since, for \(1 \leq k < m\), by definition we have
\[
|S_{b,k}| \leq M_{b,m-1},
\]
thus, for any \(k\) between 1 and \(n\), we may estimate
\[
|S_{b,k}| \leq |S_{b,m}| + (M_{b,m-1} + M_{b,m,k-m})^{1/p},
\]
that is, we have
\[
M_{b,n} \leq |S_{b,m}| + (M_{b,m-1} + M_{b,m,n-m})^{1/p}.
\]
Applying the Minkowski inequality gives
\[
(E(M_{b,n}^p))^{1/p} \leq (E|S_{b,m}|^p)^{1/p} + (E(M_{b,m-1}^p)^{1/p} + E(M_{b,m,n-m}^p)^{1/p}).
\]
(5)

By the induction hypothesis, conclusion (3) of the theorem holds for any integer less than \(n\). Thus, by the choice of \(m\), we have
\[
[E(M_{b,m-1}^p)]^{1/p} \leq (\log 2(m - 1))^{1/p} g(F_{b,m-1})
\]
and
\[
[E(M_{b,m,n-m}^p)]^{1/p} \leq (\log 2(n - m))^{1/p} g(F_{b,m,n-m})
\]
\[
\leq (\log 2(m - 1))^{1/p} g(F_{b,m,n-m}),
\]

where the last inequality is due to (4). Putting these two inequalities together, by (1) we find that
\[
E(M_{b,m-1}^p) + E(M_{b,m,n-m}^p)
\]
\[
\leq (\log 2(m - 1))^{1/p} g(F_{b,m-1}) + g(F_{b,m,n-m})
\]
\[
\leq (\log 2(m - 1))^{1/p} g(F_{b,n}).
\]

Here we also took into account that, due to (1), the functional \(g(F_{b,k})\) is nondecreasing in \(k\).

Finally, by (2), we have
\[
E|S_{b,m}|^p \leq g(F_{b,m}) \leq g(F_{b,n}).
\]

Combining inequalities (5)–(7) and taking (4) into account gives
\[
\leq (1 + \log 2(m - 1))^{1/p} g(F_{b,n})
\]
\[
\leq (1 + \log n)^{1/p} g(F_{b,n}) = (\log 2n)^{1/p} g(F_{b,n}).
\]

The inequality just obtained is equivalent to (3) which was to be proved. Thus it follows by induction that (3) holds true for all \(b = 0, 1, 2, \ldots\) and \(n = 1, 2, \ldots\). The proof of Theorem 1 is complete.

We note that Theorem 1 is a special case of the following more general Theorem 2. Before formulating it, we introduce a recurrence definition. Let \(\Lambda(n)\) be a positive and nondecreasing function of the natural number \(n\). We set
\[
\Lambda(1) := \Lambda(1) \quad \text{and for} \quad n \geq 2,
\]
\[
\Lambda(n) := \Lambda(m) + \Lambda(n - m),
\]
(8)

where \(m\) denotes the integer part of \((n + 2)/2\) as in the proof of Theorem 1 before. Now, the proving method of Theorem 1 can be repeated with minor changes in notation (see the details in [13, Theorem 4]) and yields the following.

**Theorem 2.** Suppose that there exist a functional \(g(F_{b,n})\) satisfying condition (1) and a positive and nondecreasing function \(\Lambda(n)\) such that
\[
E|S_{b,n}|^p \leq \Lambda^p(n) g(F_{b,n})
\]
for all \(b \geq 0\) and \(n \geq 1\), where \(p > 1\). Let \(\Lambda(n)\) be defined in (8). Then we have
\[
E(M_{b,n}^p) \leq \Lambda^p(n) g(F_{b,n})
\]
for all \(b \geq 0\) and \(n \geq 1\).

Clearly, Theorem 1 is a special case of Theorem 2, where \(\Lambda(n)\) equals 1 for all \(n\), due to the fact that
\[
\Lambda(n) \leq \log 2n.
\]
Indeed, by (4) we have
\[
1 + \log 2(m - 1) \leq \log 2n.
\]

The subtitle “The Generalized Rademacher-Menshov Maximal Moment Inequality” of this section will be clear in the section “Pointwise Convergence of Orthogonal Series” below.
The Generalized Erdős-Stechkin Maximal Moment Inequality

We remind the reader that on the right-hand side of condition (2) in Theorem 1, the exponent of the functional $g(F_{b,n})$ equals 1. On the other hand, on the right-hand side of condition (9) in the next Theorem 3, the exponent $r$ of the functional $g(F_{b,n})$ will be greater than 1.

The next maximal moment inequality indicated in the subtitle above was also proved by the present author [13, Theorem 1].

**Theorem 3.** Suppose that there exists a functional $g(F_{b,n})$ satisfying condition (1) and

\[
E|S_{b,n}|^p \leq g^r(F_{b,n})
\]

for all $b \geq 0$ and $n \geq 1$, where $p > 1$ and $r > 1$. Then we have

\[
E(M_{b,n}^r) \leq C g^r(F_{b,n})
\]

for all $b \geq 0$ and $n \geq 1$, provided that the constant $C$ is large enough.

Although the specific value of $C$ will have no importance for us, the constant $C$ in (10) may be taken as

\[
C = (1 - 2^{(1-r)/p})^{-p}.
\]

**Proof.** We are to find a constant $C \geq 1$, depending only on $p$ and $r$ such that

\[
E(M_{b,n}^r) \leq C g^r(F_{b,n})
\]

for all $b \geq 0$ and $n \geq 1$. Analogous to the proof of Theorem 1, the proof of (12) will go by induction with respect to $n$. Since $M_{b,1} := |S_{b,1}|$ and $C \geq 1$, (12) is obvious for $n = 1$.

Now, we assume as induction hypothesis that (12) holds for each integer less than $n$, where $n > 1$, and we will prove it for $n$ itself. We make use of another variant of the so-called bisection technique. To this effect, let $m$ be an integer such that $1 \leq m \leq n$ and

\[
g(F_{b,m-1}) \leq \frac{1}{2} g(F_{b,n}) < g(F_{b,m}),
\]

with the agreement that $g(F_{b,m-1})$ on the left is taken to be zero if $m = 1$. Then (1) and (13) imply that

\[
g(F_{b,m-n-m}) \leq g(F_{b,n}) - g(F_{b,m}) < \frac{1}{2} g(F_{b,n}).
\]

Now, for $m \leq k \leq n$, we clearly have

$|S_{b,k}| \leq |S_{b,m}| + |S_{b,m-k,m-k}| \leq |S_{b,m}| + M_{b,m-k,m-k},$

with the agreement that $M_{b,m-k,m-k}$ on the right is taken to be zero if $m = n$. For $1 \leq k < m$, we clearly have

$|S_{b,k}| \leq M_{b,m-1}$,

and hence we get for all $1 \leq k \leq n$,

\[
|S_{b,k}| \leq |S_{b,m}| + (M_{b,m-1} + M_{b,m-k,m-k})^{1/p}.
\]

Thus, we conclude that

\[
M_{b,n} \leq |S_{b,m}| + (M_{b,m-1} + M_{b,m-n-m})^{1/p},
\]

and by the Minkowski inequality, we have

\[
(E(M_{b,n}^p))^{1/p} \leq (E(|S_{b,m}|^p))^{1/p} + (E(M_{b,m-1}^p + E(M_{b,m-n-m}^p))^{1/p}.
\]

Since $m-1 < n$, we may apply the induction hypothesis, and by (12) and (13) we conclude that

\[
E(M_{b,n}^p) \leq C g^r(F_{b,m-1}) \leq \frac{C}{2} g^r(F_{b,n}).
\]

If the indices in (10) are restricted to $b + m$ and $0 \leq k \leq n - m$, then only the random variables $X_{b+m+1}, \ldots, X_{b+n}$ are involved. Since $n - m < n$, the induction hypothesis applies again, and by (12) we obtain

\[
E(M_{b+m,n-n}^p) \leq C g^r(F_{b+m,n-m}) \leq \frac{C}{2} g^r(F_{b,n}),
\]

the last inequality following from (14). In case $m = n$, inequality (17) is trivial.

Finally, by (9) and the monotonicity of $g(F_{b,n})$ in $n$, we get

\[
E|S_{b,n}|^p \leq g^r(F_{b,n}) \leq g^r(F_{b,m}).
\]

Collecting the inequalities in (15)–(18) yields

\[
(E(M_{b,n}^p))^{1/p} \leq \left(1 + \frac{C^{1/p}}{2^{(r-1)/p}}\right) g^r(F_{b,n}).
\]

Taking into account that $r > 1$, hence it follows that if $C$ is large enough, then

\[
(E(M_{b,n}^p))^{1/p} \leq C^{1/p} g^r(F_{b,n}),
\]

which is equivalent to (10) to be proved. The smallest $C$ satisfying the inequality

\[
1 + \frac{C^{1/p}}{2^{(r-1)/p}} \leq C^{1/p}
\]

is given in (11). This completes the proof of Theorem 3. \(\square\)

The subtitle “The Generalized Erdős-Stechkin Maximal Moment Inequality” of this section will be explained in the section “Pointwise Convergence of Fourier Series” below.

**Pointwise Convergence of Orthogonal Series**

Instead of random variables, one may also consider real-valued functions $(f_k(t) : k = 1, 2, \ldots)$ over some bounded interval $I := [a_1, a_2]$. We assume that these functions $f_k(t)$ are integrable in the Lebesgue sense. To investigate pointwise convergence of the infinite series $\sum_{k=1}^{\infty} f_k(t)$, we introduce such notation that is analogous to that introduced in the preceding sections. Namely, this time we set

\[
S_{b,n}(t) := \sum_{k=b+1}^{b+n} f_k(t) \quad \text{and} \quad M_{b,n}(t) := \max_{1 \leq k \leq n} |S_{b,k}(t)|,
\]

578 Notices of the AMS Volume 61, Number 6
where \( b \geq 0 \) and \( n \geq 1 \) are integers. In statements about \( S_{0,b}(t) \) only, the abbreviated notation \( S_b(t) \) will be used.

Furthermore, we denote by \( g(b,n) \) a nonnegative functional depending only on the functions \( f_{b,1}(t), f_{b,2}(t), \ldots, f_{b,n}(t) \), which possesses the subadditivity property (cf. (1))

\[
(20) \quad g(b,k) + g(b+k,\ell) \leq g(b,k+\ell)
\]

for all \( b \geq 0 \) and \( 1 \leq k < k + \ell \).

The analogue of Theorem 1 reads as follows.

**Theorem 4.** Suppose that there exists a functional \( g(b,n) \) satisfying condition (20) and

\[
\int_I |S_{b,n}(t)|^p dt \leq g(b,n)
\]

for all \( b \geq 0 \) and \( n \geq 1 \), where \( p > 1 \). Then we have

\[
\int_I M_{b,n}^p(dt) \leq (\log 2n)^p g(b,n)
\]

for all \( b \geq 0 \) and \( n \geq 1 \).

The proof of Theorem 4 runs along the same lines as that of Theorem 1.

Next, we consider the special case, where

\[
f_k(t) := a_k \varphi_k(t), \quad k = 1, 2, \ldots
\]

with \( (a_k) \) a sequence of real numbers (the so-called coefficients), while \( \varphi_k(t) \) is an orthonormal system (in abbreviation: ONS) on the interval \( I \); that is, the functions \( \varphi_k(t) \) are square integrable on \( I \) and

\[
\int_I \varphi_k(t) \varphi_\ell(t) dt = \left\{ \begin{array}{ll}
0 & \text{if } k \neq \ell, \\
1 & \text{if } k = \ell; \ k, \ell = 1, 2, \ldots
\end{array} \right.
\]

By orthonormality, we clearly have (cf. notation in (19))

\[
(21) \quad \int_I S_{b,n}^2(t) dt := \int_I \left( \sum_{k=1}^{b+n} a_k \varphi_k(t) \right)^2 dt
\]

\[
= \sum_{k=b+1}^{b+n} \sum_{\ell=b+1}^{b+n} a_k a_\ell \int_I \varphi_k(t) \varphi_\ell(t) dt = \sum_{k=b+1}^{b+n} a_k^2
\]

This means that by choosing \( p := 2 \) and

\[
g(b,n) := \sum_{k=b+1}^{b+n} a_k^2, \quad b \geq 0 \quad \text{and} \quad n \geq 1,
\]

condition (20) is trivially satisfied, even with the sign “=” in place of “≤” (the so-called additive case).

The best-known example of an orthogonal system of functions is the trigonometric system:

1. \( \cos t, \sin t, \cos 2t, \sin 2t, \ldots, \cos kt, \sin kt, \ldots \)

on the interval \( I := [0, 2\pi] \). Taking into account the periodicity of these functions, they are orthogonal on any interval of length \( 2\pi \). The orthogonality is immediately clear if we apply the trigonometric identities

\[
(22) \quad \cos \alpha \cos \beta = \frac{1}{2} (\cos(\alpha + \beta) + \cos(\alpha - \beta)),
\]

\[
\sin \alpha \cos \beta = \frac{1}{2} (\sin(\alpha + \beta) + \sin(\alpha - \beta)),
\]

\[
(23) \quad \sin \alpha \sin \beta = \frac{1}{2} (\cos(\alpha - \beta) - \cos(\alpha + \beta)).
\]

However, the trigonometric system is not a normal one, since

\[
\int_0^{2\pi} 1 dt = 2\pi, \int_0^{2\pi} \cos^2 ktdt = \int_0^{2\pi} \sin^2 ktdt = \frac{\pi}{2}, \quad k = 1, 2, \ldots
\]

Thus, the orthonormal trigonometric system is the following one:

\[
\frac{1}{\sqrt{2\pi}}, \quad \frac{\cos kt}{\sqrt{\pi}}, \quad \frac{\sin kt}{\sqrt{\pi}}, \quad k = 1, 2, \ldots
\]

The fundamental result in the theory of orthogonal series is the Rademacher-Menshov inequality (see [10], [24], and also [1, p. 80]). The following Corollary 5 is a special case of Theorem 4, where \( p = 2 \) and \( g(b,n) \) is given in (21).

**Corollary 5.** If \( \varphi_k(t) \) is an arbitrary ONS on a bounded interval \( I \) and \( (a_k) \) is a sequence of real numbers, then we have

\[
\int_I M_{b,n}^2(dt) \leq (\log 2n)^2 \sum_{k=b+1}^{b+n} a_k^2
\]

for all \( b \geq 0 \) and \( n \geq 1 \), where \( M_{b,n}(t) \) is defined in (19).

We recall that the next Rademacher-Menshov theorem on the pointwise convergence of the orthogonal series \( \sum_{k=1}^{\infty} a_k \varphi_k(t) \) (see [10], [24] and also [1, Theorem 2.4.2, p. 88]) is proved by means of Corollary 5.

**Theorem 6.** If \( \varphi_k(t) \) is an arbitrary ONS on a bounded interval \( I \) and \( (a_k) \) is a sequence of real numbers satisfying the condition

\[
(24) \quad \sum_{k=1}^{\infty} a_k^2 (\log k)^2 < \infty,
\]

then the orthogonal series

\[
(25) \quad \sum_{k=1}^{\infty} a_k \varphi_k(t)
\]

converges almost everywhere (a.e.) on \( I \).

**Proof.** We begin with proving that condition (24) implies the a.e. convergence of the subsequence
(S_2^n(t)). Indeed, on account of the orthonormality and conditions (21) and (24), we have
\[\sum_{n=1}^{\infty} \int_I n^2 S_{2^n,2^n}^2(t) dt = \sum_{n=1}^{\infty} n^2 \sum_{k=2^n+1}^{2^{n+1}} a_k^2 \]
\[\leq \sum_{n=1}^{\infty} \sum_{k=2^n+1}^{2^{n+1}} a_k^2 (\log k)^2 < \infty.\]

By the monotone convergence theorem (called Beppo Levi’s theorem in [25, pp. 35-36]), it follows that the series in the integrands above converges a.e.; that is, we have
\[\sum_{n=1}^{\infty} n^2 S_{2^n,2^n}^2(t) < \infty \quad \text{a.e.}\]

Next, we apply the familiar Cauchy inequality for sequences to obtain for \(\nu \geq 2\) and \(\ell \geq 1\) that
\[\left( S_{2^n,2^n}^2(t) - S_{2^n}^2(t) \right)^2 = \left( \sum_{\nu=1}^{\nu+q-1} S_{2^n,2^n}^2(t) \right)^2 \]
\[\leq \sum_{\nu=1}^{\nu+q-1} n^2 S_{2^n,2^n}^2(t) \sum_{n=\nu}^{\nu+q-1} 1/n^2 \]
\[= a\left( \frac{1}{\nu} \right) \to 0 \quad \text{a.e. as} \ \nu \to \infty.\]

Therefore, the a.e. convergence of the subsequence \((S_{2^n}(t))\) is proved.

It remains to prove that the relation (26)
\[M_{2^n,2^n}(t) := \max_{2^n \leq \nu \leq 2^{n+1}} |S_{\nu}(t) - S_{2^n}(t)| \to 0 \quad \text{as} \ \nu \to \infty\]
also holds a.e. By Corollary 5 and condition (24) we estimate as follows:
\[\sum_{n=3}^{\infty} \int_I M_{2^n,2^n}^2(t) dt \leq \sum_{n=3}^{\infty} (n+1)^2 \sum_{k=2^n+1}^{2^{n+1}} a_k^2 \]
\[\leq \sum_{n=3}^{\infty} \sum_{k=2^n+1}^{2^{n+1}} a_k^2 (\log k)^2 = 2 \sum_{k=9}^{\infty} a_k^2 (\log k)^2 < \infty,\]
since \((n+1)^2 \leq 2n^2\) if \(n \geq 3\). Again, by the monotone convergence theorem, the series
\[\sum_{n=3}^{\infty} M_{2^n,2^n}^2(t)\]
converges a.e., whence (26) follows. This completes the proof of Theorem 6. \(\Box\)

A leading expert of the Hungarian orthogonal school, Károly Tandori [29] proved (see in Theorem 8 below) that in certain cases condition (24) is not only a sufficient, but also a necessary, condition for the a.e. convergence of an arbitrary orthogonal series (25).

**Theorem 7.** If a sequence \((a_k)\) of real numbers is such that
\[|a_1| \geq |a_2| \geq \cdots \geq |a_k| \geq \cdots\]
and
\[\sum_{k=1}^{\infty} a_k^2 (\log k)^2 = \infty,\]
then on any bounded interval \(I\) one can construct an ONS \((\phi_k(t))\) so that the orthogonal series (25) diverges at each point of \(I\).

We note that analogous theorems have been proved in the case of orthogonal random variables. As an example, we present the following.

**Theorem 8.** Let \((X_k : k \geq 1)\) be a sequence of orthogonal random variables with finite second moments; that is,
\[E(X_k X_\ell) = 0 \quad \text{and} \quad \sigma_k^2 := E(X_k^2) < \infty \quad \text{for} \ k \neq \ell; k, \ell = 1, 2, \ldots.\]

If
\[\sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2} (\log k)^2 < \infty,
\]
then the strong law of large numbers holds true:
\[\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} X_k = 0 \quad \text{almost surely.}\]

The proof of Theorem 8 hinges on Theorem 1 in the case \(p = 2\) and on the next lemma (see, e.g., [1, Theorem 2.2.2, p. 72]).

**Kronecker lemma.** Suppose the sequence \((\lambda_k : k \geq 1)\) of positive numbers is nondecreasing and tends to \(\infty\) as \(k \to \infty\). If a sequence \((a_k)\) of real numbers is such that the series
\[\sum_{k=1}^{\infty} \frac{a_k}{\lambda_k}\]
converges, then
\[\lim_{n \to \infty} \frac{1}{\lambda_n} \sum_{k=1}^{n} a_k = 0.\]

Further strong convergence as well as complete convergence theorems were proved in [13, pp. 308-313].

**Pointwise Convergence of Fourier Series**

We consider the trigonometric series
\[\sum_{k=1}^{\infty} (a_k \cos m_k t + b_k \sin m_k t),\]
where \((a_k)\) and \((b_k)\) are sequences of real numbers and \((m_k)\) is a strictly increasing sequence of natural numbers. We recall that the series (27) is said to be lacunary (see, e.g., [30, p. 202]) if the inequality
\[\frac{m_{k+1}}{m_k} \geq \lambda > 1\]
is satisfied for all \(k\).
Theorem 9 (Kolmogorov [7]). If
\[ \sum_{k=1}^{\infty} (a_k^2 + b_k^2) < \infty, \]
then the lacunary series (27) converges a.e.

We note that, by condition (29), the series (27) is the Fourier series of a square integrable function (see [25, p. 70] or [30, (1.1) Theorem, p. 127]).

We recall that a sequence \((m_k)\) of natural numbers is said to satisfy condition \((B_2)\) (see [2, p. 233]) if every natural number \(m\) can be represented only by a limited number of means in the form
\[ m = m_k + m_\ell \quad \text{or} \quad m = m_k - m_\ell, \]
where \(m_k\) and \(m_\ell\) are terms of the given sequence \((m_k : k = 1, 2, \ldots)\).

Theorem 10 (Erdős [4], see also [2, p. 233]). If a sequence \((m_k)\) of natural numbers satisfies condition \((B_2)\) and the sequences \((a_k)\) and \((b_k)\) of real numbers satisfy condition (29), then the series (27) converges a.e.

In the next section, we will prove that from the lacunary condition (28) it follows that \((m_k)\) satisfies condition \((B_2)\), while the converse implication is not true in general. Thus, Erdős’s theorem is an essential generalization of Kolmogorov’s theorem.

In his proof, Erdős made use of the following lemma of Sidon [28] (see also [2, p. 258]).

Lemma 11. If a sequence \((m_k)\) of natural numbers satisfies condition \((B_2)\), then for any trigonometric polynomials
\[ T_n(t) := \sum_{k=1}^{n} a_k \cos m_k t \quad \text{or} \quad T_n(t) := \sum_{k=1}^{n} a_k \sin m_k t, \]
the following inequality holds true:
\[ \int_0^{2\pi} T_n^4(t) dt \leq C \left( \int_0^{2\pi} T_n^2(t) dt \right)^2 = C \left( \pi \sum_{k=1}^{n} a_k^2 \right)^2, \]
where the constant \(C\) is independent of the order \(n\) of the polynomial \(T_n(t)\) and its coefficients \((a_k)\).

Clearly, inequality (30) is a particular case of inequality (9), where \(p = 4\) and \(r = 2\),
\[ S_{b,n}(t) := \sum_{k=b+1}^{b+n} a_k \cos m_k t \quad \text{or} \quad \sum_{k=b+1}^{b+n} a_k \sin m_k t, \]
and
\[ g(b, n) := \sqrt{C \pi} \sum_{k=b+1}^{b+n} a_k^2. \]

Analyzing Erdős’s proof in [4] (see also [2, pp. 259–263]) gave rise to the next so-called Erdős-Stechkin maximal moment inequality.

Theorem 12. Suppose \(p > 2\) and there exists a sequence \((a_k)\) of real numbers such that
\[ \int_0^1 |S_{b,n}(t)|^p dt \leq C \left( \sum_{k=b+1}^{b+n} a_k^2 \right)^{p/2}, \]
where \(C\) is a constant. Then we also have
\[ \int_0^1 \left( \max_{1 \leq k \leq n} |S_{b,k}(t)| \right)^p dt \leq C \varepsilon \left( \sum_{k=b+1}^{b+n} a_k^2 \right)^{p/2}, \]
where the constant \(C_\varepsilon\) does not depend on \(b, n, \) the sequence \((a_k)\), and the exponent \(p\) for \(p \geq 2 + \varepsilon, \varepsilon > 0\).

One can easily see that Theorem 12 was actually proved by Erdős [4] in the special case of such trigonometric polynomials, where the indices of the nonzero coefficients satisfy condition \((B_2)\) and \(p = 4\). The general form as stated in Theorem 12 is proved along the analogous lines of Erdős’s proof. The extension of Erdős’s Theorem 10 was suggested by Stechkin, whose oral communication was later elaborated by Gaposhkin [6, pp. 29–31].

Proof of Sidon’s Lemma. For the reader’s convenience, we present the proof of Lemma 11 in the case of the cosine polynomial \(T_n(t)\). Since
\[ T_n^2(t) := \left( \sum_{k=1}^{n} a_k \cos m_k t \right)^2 = \sum_{k=1}^{n} a_k^2 \cos^2 m_k t + \sum_{1 \leq k \neq \ell \leq n} a_k a_\ell \cos (m_k + m_\ell) t \cos m_\ell t, \]
applying (22) and its particular case
\[ \cos^2 \alpha = \frac{1 - \cos 2\alpha}{2}, \]
gives that
\[ T_n^4(t) = \frac{1}{2} \sum_{k=1}^{n} a_k^2 + \frac{1}{2} \sum_{k=1}^{n} a_k^2 \cos 2m_k t + \sum_{1 \leq k \neq \ell \leq n} a_k a_\ell \cos (m_k + m_\ell) t \cos m_\ell t \]
\[ + \frac{1}{2} \sum_{1 \leq k \neq \ell \leq n} a_k a_\ell \cos (m_k - m_\ell) t \cos m_\ell t \]
\[ = \sum_{j=0}^{2m_n} A_j \cos jt \]
with the coefficients
\[ A_0 := \frac{1}{2} \sum_{k=1}^{n} a_k^2, \]
\[ A_j := \frac{1}{2} \sum_{1 \leq k \neq \ell \leq n} a_k a_\ell, \quad j = 1, 2, \ldots, 2m_n, \]
where the double summation is extended to those pairs \((k, \ell)\) for which
\[ |m_k \pm m_\ell| = j \quad \text{and} \quad k \neq \ell. \]
It is obvious that
\[
A_0^2 = \frac{1}{4} \left( \sum_{k=1}^{n} a_k^2 + \sum_{1 \leq k \neq \ell \leq n} a_k a_\ell \right).
\]

Due to condition (B2), the representation \(|m_k \pm m_\ell| = j\) can hold only in finitely many cases, which we denote by \(s\). This natural number \(s\) depends upon neither \(j\) nor \(n\). Applying the familiar Cauchy inequality for sequences gives
\[
A_j^2 = \frac{1}{4} \left( \sum_{1 \leq k \neq \ell \leq n} a_\ell a_\ell \right) \leq \frac{s}{4} \left( \sum_{k=1}^{n} a_k^2 \right).
\]

Hence it follows immediately that
\[
\sum_{j=1}^{2m} A_j^2 \leq \frac{s}{4} \left( \sum_{k=1}^{n} a_k^2 \right)^2 \leq \frac{s}{4} \left( \sum_{k=1}^{n} a_k^2 \right)^2.
\]

Combining (31)–(33) yields (30) as follows:
\[
\begin{align*}
\int_0^{2\pi} T_n(t) \, dt &= \int_0^{2\pi} \left( \sum_{j=1}^{2m} A_j \cos j t \right) \, dt \\
&= \pi \left( 2 A_0^2 + \sum_{j=0}^{2m} A_j^2 \right) \\
&= \pi \left( 2 A_0^2 + \sum_{j=1}^{2m} A_j^2 \right) \\
&= \pi \left( 2 + s \right) \left( \sum_{k=1}^{n} a_k^2 \right)^2 \\
&\leq C \left( \pi \sum_{k=1}^{n} a_k^2 \right)^2,
\end{align*}
\]

where \(C := \frac{2 + s}{4\pi}\).

In the case of sine polynomials \(T_n(t)\), the proof goes along the same lines as above, except that this time (23) is used instead of (22). The proof of Sidon's Lemma 11 is complete. \(\square\)

Next, we will prove the assertion mentioned above that if a sequence \((m_k : k \geq 1)\) of natural numbers is lacunary, that is, for some \(\lambda > 1\),
\[
m_k+1 \geq \lambda \cdot m_k, \quad k = 1, 2, \ldots,
\]
then a natural number \(s\) depending only on \(\lambda\) can be given such that for any natural number \(m\), the representations
\[
m = m_\ell + m_k \quad \text{and} \quad m = m_\ell - m_k, \quad \ell > k \geq 1,
\]
can hold true for at most \(s\) pairs \((\ell, k)\).

First, we justify this statement in the case where \(m = m_\ell + m_k, \ell > k \geq 1\). It follows from (34) that
\[
\frac{m_\ell}{m_{\ell-1}} \geq \lambda, \quad \text{and consequently,} \quad m_k \leq m_{\ell-1} \leq \frac{1}{\lambda} m_\ell;
\]
whence we conclude that
\[
m_\ell < m = m_\ell + m_k \leq m_\ell + m_{\ell-1} \leq m_\ell \left( 1 + \frac{1}{\lambda} \right) = \frac{\lambda + 1}{\lambda} m_\ell.
\]
Consequently, any number that can be written in the form \(m = m_\ell + m_k\) satisfies the inequality
\[
\frac{\lambda}{\lambda + 1} m \leq m_\ell < m.
\]

Second, we consider the case where \(m = m_\ell - m_k\), \(\ell > k \geq 1\). Again, it follows from (34) that
\[
\frac{m_{\ell-1}}{m_\ell} \leq \frac{1}{\lambda}, \quad \text{and consequently,} \quad m_{\ell-1} \leq \frac{1}{\lambda} m_\ell;
\]
whence we can conclude that
\[
m_\ell > m = m_\ell - m_k \geq m_{\ell-1} \geq m_\ell - \frac{1}{\lambda} m_\ell = \frac{\lambda - 1}{\lambda} m_\ell.
\]

Consequently, any number that can be written in the form \(m = m_\ell - m_k\) satisfies the inequality
\[
\frac{\lambda}{\lambda - 1} m \geq m_\ell > m.
\]

Given an integer \(m\), denote by \(\ell_1\) the smallest integer and by \(\ell_2\) the largest integer of those indices \(\ell\) that occur in the representations \(m = m_\ell \pm m_k\), where \(\ell > k \geq 1\). It follows from (35) and (36) that
\[
m_{\ell_1} \geq \frac{\lambda}{\lambda + 1} m \quad \text{and} \quad m_{\ell_2} \leq \frac{\lambda}{\lambda - 1} m,
\]
whence it follows that
\[
\frac{m_{\ell_1}}{m_{\ell_2}} \leq \frac{\lambda}{\lambda - 1}. \frac{\lambda + 1}{\lambda + 1} = \frac{\lambda + 1}{\lambda - 1}.
\]

On the other hand, it follows from (34) that
\[
\frac{m_{\ell_1}}{m_{\ell_2}} \geq m_{\ell_{i-1}} \geq \frac{m_{\ell_i}}{m_{\ell_{i-1}}}, \quad i = 1, 2, \ldots, i \geq 2.
\]

Combining the last two inequalities results in the inequality
\[
\lambda^i \leq \frac{m_{\ell_i}}{m_{\ell_{i-1}}} \leq \frac{\lambda + 1}{\lambda - 1}.
\]

Since \(\lambda > 1\), we have
\[
\lim_{s \to \infty} \lambda^s = \infty,
\]
and this means that \(s\) must be bounded in (37).

To sum up, the representation \(m = |m_k \pm m|\), where \(\ell > k \geq 1\), can occur only in those cases where the index \(\ell\) of \(m_k\) equals in turn \(\ell_1, \ell_2, \ldots, \ell_s\). This completes the proof of our assertion made just after (34).

To see that Erdős's Theorem 10 is an essential generalization of Kolmogorov's Theorem 9, we note that it is possible to construct such a sequence \((m_k : k \geq 1)\) of natural numbers that satisfies condition (B2) and
\[
\frac{m_\ell}{\ell^2} \leq C, \quad \ell = 1, 2, \ldots,
\]
where \(C\) is a constant (see, e.g., [2, p. 263]). This sequence \((m_k)\) cannot be lacunary, since otherwise one gets a contradiction. Indeed, suppose the existence of some \(\lambda > 1\) such that (34) holds true. Then we would have
\[
m_{\ell+1} \geq \lambda m_k \geq \lambda^2 m_{\ell-1} \geq \cdots \geq \lambda^{\ell-1} m_1, \quad \ell = 1, 2, \ldots.
\]
Taking into account (39), we would conclude hence that
\[
1 \geq \frac{m_{\ell+1}}{m_{\ell+1}} \geq \frac{\lambda^{\ell-1} m_1}{(\ell + 1)^3}, \quad \ell = 1, 2, \ldots.
\]
But this is a contradiction, since for any \( \lambda > 1 \), we have
\[
\lim_{\ell \to \infty} \frac{\lambda^\ell}{(\ell+1)^3} = \infty.
\]
This completes the proof of our remark made just before (39).

The Upper Part of the Law of the Iterated Logarithm

Let \( (\varphi_k(t) : k = 1, 2, \ldots) \) be a system of real-valued functions that are measurable in the Lebesgue sense on the unit interval \([0,1]\). The following definition was introduced by Alexits [1, pp. 87–88].

The system \( (\varphi_k(t)) \) is said to be an **equinormed, strongly multiplicative system (ESMS)** if
\[
\int_0^1 \varphi_k(t) dt = 0, \quad \int_0^1 \varphi_k^2(t) dt = 1
\]
for all \( k = 1, 2, \ldots \), and
\[
\int_0^1 \varphi_{k_1}^{\alpha_1}(t) \varphi_{k_2}^{\alpha_2}(t) \cdots \varphi_{k_\ell}^{\alpha_\ell}(t) dt = \int_0^1 \varphi_{k_1}^{\alpha_1}(t) dt \int_0^1 \varphi_{k_2}^{\alpha_2}(t) dt \cdots \int_0^1 \varphi_{k_\ell}^{\alpha_\ell}(t) dt
\]
for all \( 1 \leq k_1 < k_2 < \cdots < k_\ell \), each of the exponents \( \alpha_1, \alpha_2, \ldots, \alpha_\ell \) may equal 1 or 2, and \( \ell = 2, 3, \ldots \).

We note (see, e.g., [9]) that any system of stochastically independent functions on the unit interval \([0,1]\) with mean value 0 and variance 1 is an ESMS. Another example is a **strongly lacunary** system of the trigonometric functions of the form (cf. (28))
\[
\sqrt{2} \sin 2\pi m_k t, \quad \text{where} \quad \frac{m_{k+1}}{m_k} \geq 3, \quad k = 1, 2, \ldots.
\]

We note that Alexits proved (see [1, Theorem 3.2.4, p. 189 and Theorem 3.2.7, p. 194]) that the **zero-one law** holds for an ESMS; that is, the series
\[
\sum_{k=1}^{\infty} a_k \varphi_k(t)
\]
converges a.e. if \( \sum_{k=1}^{\infty} a_k^2 < \infty \), and diverges a.e. if \( \sum_{k=1}^{\infty} a_k^2 = \infty \).

Clearly, the functions of an ESMS can be considered as random variables on the unit interval \([0,1]\) endowed with the Lebesgue measure. Given a sequence \( (a_k) \) of real numbers and an ESMS \( (\varphi_k(t)) \), we set
\[
S_n(t) := \sum_{k=1}^{n} a_k \varphi_k(t)
\]
and
\[
A_n := \left( \sum_{k=1}^{n} a_k^2 \right)^{1/2}, \quad n = 1, 2, \ldots
\]
By the Erdős-Stechkin maximal moment inequality, we have for any \( p > 2 \),
\[
\int_0^1 \left( \max_{1 \leq n \leq N} |S_n(t)|^p \right) dt \leq C_p A_n^p, \quad N = 1, 2, \ldots,
\]
where the constant \( C_p \) does not depend on the sequence \( (a_k) \) and the system \( (\varphi_k(t)) \). Now, repeating our proof of [12, Theorem 5] while making use of the maximal moment inequality (40), one can conclude the upper part of the law of the iterated logarithm (LIL) for any uniformly bounded ESMS, as stated in our next

**Theorem 13.** Let \( (\varphi_k(t)) \) be a uniformly bounded ESMS. If
\[
A_n \to \infty \quad \text{and} \quad \max_{1 \leq k \leq n} |a_k| = o \left( \frac{A_n^2}{\ln \ln A_n} \right),
\]
where by “\( \ln \)” we denote the logarithm to the natural base \( e \), then
\[
\text{meas} \{ t \in [0,1] : \limsup_{n \to \infty} \frac{S_n(t)}{\sqrt{2 A_n \ln \ln A_n}} \leq 1 \} = 1,
\]
where by “\( \text{meas} \)” we mean the Lebesgue measure on the real line.

Indeed, taking into account [12, Remark 3, p. 119], the proof of [12, Theorem 5, pp. 130–132] supplemented by the maximal moment inequality (40) completes the proof of Theorem 13.

We note that only a weaker form of the upper part of the LIL was proved in [12, Theorem 5] for a uniformly bounded ESMS, owing to the unfortunate fact that the present author was unaware of the Erdős-Stechkin maximal moment inequality at that time.

We also note that it was Kolmogorov [8], who proved even the full part of the LIL (that is, (41) with “probability \( P \)” instead of “\( \text{meas} \)” and with “=1” in place of “\( \leq 1 \)” for any sequence of uniformly bounded, stochastically independent random variables over a probability space \( (\Omega, \mathcal{F}, P) \).

**Historical Remarks**

As we emphasized in the section “Pointwise Convergence of Fourier Series”, the ingenious proof of Theorem 10 by Paul Erdős in 1943 exhibited the basic ideas that were applied by a number of mathematicians in proving maximal moment inequalities so far.

The superadditivity condition (1) is due to Serfling [26] in 1970. Our Theorems 1 and 3, proved in [13, Theorems 1 and 3], differ from those in [26, Theorems A and B] in the exponents on the right-hand sides in assumptions (2) and (9), and accordingly in conclusions (3) and (10), respectively. Furthermore, our proofs in [13] are more straightforward than the proofs in [26].
Assuming additivity instead of superadditivity, analogous inequalities were proved by Billingsley in his book [3, Theorem 12.2, p. 94, and Problem 5, p. 102], which appeared in 1968. We note that Billingsley also proved maximal inequalities, where the assumptions were not on the moments of sums $S_{b,n}$ but on product moments or joint tails of two nonoverlapping sums $S_{b,n}$ and $S_{a,m}$. The latter inequalities can be applied in such cases where the moments do not exist. It should also be noted that Billingsley’s proofs work, with trivial changes, in the superadditive case as well.

We mention that Billingsley’s theorems in [3, see Theorem 12.1 on p. 89, Theorem 12.2 on p. 94, Theorem 12.5 on p. 98, and Theorem 12.6 on p. 101] were generalized in our paper [21, see Theorems 1 and 2, and Corollaries 1 and 2, respectively] in 1982.

In a joint paper [20] with Serfling and Stout, we introduced a new notion called $Q$-superadditivity as follows. A nonnegative function $g(i,j)$ is said to be $Q$-superadditive if for all integers $1 \leq b \leq n < m$,

\[
g(b, n) \leq g(b, n + 1), \\
g(b, n) + g(n + 1, m) \leq Qg(b, m),
\]

where $1 \leq Q < 2$. Clearly, the case $Q = 1$ corresponds to the usual notion of superadditivity. A number of moment as well as probability bounds were proved in [20] for the maximum of partial sums which gave back the previously proved inequalities in the case $Q = 1$.

We also proved moment inequalities for the maxima of the rectangular partial sums of random fields in [16] and [18]; probability inequalities of exponential type in [14], [17], and [19]; and maximal moment as well as probability inequalities for stochastic processes in [22] and [23].

Applications of maximal inequalities to obtain strong laws of large numbers can also be found in [13] and [27] for sequences of random variables; furthermore, in [15], and [18] for quasi-stationary sequences and random fields, respectively.

Finally, we mention that in recent times maximal inequalities are also of use in Numerical Analysis in the study of the convergence of wavelets (see, for example, [11]) as well as in Functional Data Analysis, (see, for example, [5]).

**Acknowledgments**

I express my thanks to Professor John Horváth for his encouragement and valuable comments which helped improve the presentation of this paper. I am also grateful to the referees whose remarks have corrected typos and improved the presentation.

**References**


**Author’s Note**

Ferenc Móricz was born and educated in Hungary, received his BSc, MSc and Ph.D. at the Bolyai Institute of the University of Szeged; the Bolyai Institute was founded by F. Riesz and A. Haar just after the First World War. In 1983 Ferenc Móricz received the Erdős Prize, awarded each year to one of the outstanding Hungarian mathematicians. He earned a Keldysh scholarship and spent one and a half years at the Steklov Mathematical Institute in Moscow; spent a “Sommer Semester” at the University of Ulm, Germany, as a Visiting Professor; and altogether six academic years at the Johns Hopkins University at Baltimore, Indiana University at Bloomington, University of Wisconsin at Madison, Syracuse University, University of Tennessee at Knoxville, and Texas A&M University at College Station, as a Visiting Professor between the years 1981 and 2005.

---

**ICERM**

Institute for Computational and Experimental Research in Mathematics

ICERM welcomes applications for long- and short-term visitors. Support for local expenses may be provided. Full consideration will be given to applications received by March 17, 2014. Decisions about online workshop applications are typically made 1-3 months before each program, as space and funding permit. ICERM encourages women and members of underrepresented minorities to apply.

**SEMESTER PROGRAM: SPRING 2015**

**Phase Transitions and Emergent Properties**

*February 2 - May 8, 2015*

**Organizing Committee:**

Mark Bowick, Syracuse University
Beatrice de Tiliere, Université Pierre et Marie Curie, Paris
Richard Kenyon, Brown University
Charles Radin, University of Texas at Austin
Peter Winkler, Dartmouth College

**Program Description:**

Emergent phenomena are properties of a system of many components which are only evident or even meaningful for the collection as a whole. A typical example is a system of many molecules, whose bulk properties may change from those of a fluid to those of a solid in response to changes in temperature or pressure. The basic mathematical tool for understanding emergent phenomena is the variational principle, most often employed via entropy maximization. The difficulty of analyzing emergent phenomena, however, makes empirical work essential; computations generate conjectures and their results are often our best judge of the truth. The semester will concentrate on different aspects of current interest, including unusual settings such as complex networks and quasicrystals, the onset of emergence as small systems grow, and the emergence of structure and shape as limits in probabilistic models.

**Workshops:**

- Crystals, Quasicrystals and Random Networks  
  *February 9-13, 2015*
- Small Clusters, Polymer Vesicles and Unusual Minima  
  *March 16-20, 2015*
- Limit Shapes  
  *April 13-17, 2015*

**Program details:**

[http://icerm.brown.edu](http://icerm.brown.edu)

About ICERM: The Institute for Computational and Experimental Research in Mathematics is a National Science Foundation Mathematics Institute at Brown University in Providence, Rhode Island. Its mission is to broaden the relationship between mathematics and computation.