Maximal generalization of Baum-Katz theorem
and optimality of sequential tests

Didier Piau

Abstract
From Baum and Katz, the Cesàro means of i.i.d. increments distributed
like $X$ $r$-converge if and only if $|X|^{r+1}$ is integrable. We generalize this,
and we unify other results, by proving that the following equivalence holds,
if and only if $\Phi$ is moderate: the means $\Phi$-converge if and only if $\Phi(\ell_\varepsilon)$ is
integrable for every $\varepsilon$ if and only if $|X|\Phi(|X|)$ is integrable. Here, $\ell_\varepsilon$ is the
last time when the deviation of the mean from its limit is at least $\varepsilon$, and
$\Phi$-convergence is the analogue of $r$-convergence. This solves a question
about the asymptotic optimality of Wald's sequential tests.

Keywords and phrases. Baum-Katz theorem, $r$-convergence law of large
numbers, convergence rates, moderate functions, sequential analysis.


1 Introduction

In this paper, we prove a refinement of the strong law of large numbers and
we apply it to a problem of sequential analysis. Let $X$ and $(X_n)_{n \geq 1}$ be i.i.d.
random variables defined on the same probability space. The most standard
version of the strong law of large numbers asserts that $S_n/n$ converges a.s. if
and only if $X$ is integrable, where $S_n := X_1 + \cdots + X_n$. This amounts to two
statements: first, when $X$ is integrable, the last time $\ell_\varepsilon^X$ of a deviation of size
at least $\varepsilon$ of $S_n/n$ from $x = E[X]$ is a.s. finite, for any $\varepsilon$; and, when $X$ is not
integrable, no value of $x$ is such that $\ell_\varepsilon^X$ is a.s. finite for any $\varepsilon$. Strassen (1967)
introduced refined versions of this, see also Lai (1976).

Definition 1 (Strassen) Let $r > 0$ and $(Y_n)_n$ be random variables defined on
the same probability space. Set

$$\ell_\varepsilon^X := \sup\{0\} \cup \{n \geq 1; \|Y_n - x\| \geq \varepsilon\}.$$  

If $(\ell_\varepsilon^X)^r$ is integrable for all $\varepsilon > 0$, $(Y_n)_n$ $r$-converges to $x$.

Complete convergence to $x$, see Hsu and Robbins (1947) and Erdős (1949), is
the integrability of the total time spent outside of any ball around $x$. Since this

Date: October 4th, 2000.
time is at most $\ell^2$, 1-convergence implies complete convergence. The following characterizes the $r$-convergence of Cesàro means of i.i.d. increments.

**Theorem A (Baum and Katz (1965))** For any $r > 0$, $S_n/n$ $r$-converges if and only if $|X|^{r+1}$ is integrable. Then, $S_n/n$ $r$-converges to $E[X]$.

We generalize Theorem A to every function $\Phi$ such that an analogue of the result can hold. We prove that these are exactly the moderate functions. This allows to solve an open question of sequential analysis.

**Definition 2 (Feller (1969))** A positive function $\Phi$ defined on $t \geq 0$, is moderate if $\Phi$ is non decreasing, increases to infinity, and if $\Phi(2t)/\Phi(t)$ is bounded.

Bingham, Goldie and Teugels (1989) call the moderate functions, “increasing functions of dominated variation” (see their Section 2.1). For any deterministic non decreasing function $\Phi$, increasing to infinity, $(Y_n)_n$ $\Phi$-converges to $x$ if and only if $\Phi(\ell^2)$ is integrable for all $\varepsilon$. Here is our main result.

**Theorem B** Let $X$ and $(X_n)_n \geq 1$ be i.i.d. and defined on the same probability space. Let $\Phi$ be moderate. Then, $S_n/n$ $\Phi$-converges if and only if $|X| \Phi(|X|)$ is integrable. More precisely, assume that $X$ is integrable and centered. Then, assertions (i) to (iii) below are equivalent:

(i) $|X| \Phi(|X|)$ is integrable;

(ii) for all $\varepsilon > 0$, $\sum_{n \geq 1} n^{-1} \Phi(n) \mathbb{P}[|S_n/n| \geq \varepsilon] < \infty$;

(iii) for all $\varepsilon > 0$, $\Phi(\ell^2)$ is integrable.

Furthermore, if $\Phi$ is non decreasing, unbounded, and non moderate, assertions (i) and (iii) are not equivalent.

**Theorem C** Let $\Phi$ be a moderate function. Under an integrability assumption analogous to (i), Wald’s sequential test is asymptotically optimal, when the error probabilities goes to 0, with respect to the $\Phi$-moment of the observation time of the test.

See Section 2 for a precise statement of Theorem C. Baum and Katz (1965) prove Theorem B for $\Phi(t) = t^r$ with $r > 0$, as a part of their Theorem 3 and a similar statement for $\Phi(t) = t^\nu \log(t)$ with $r \geq 0$, as a part of their Theorem 2. All these functions are moderate. Baum and Katz (1965) use results of Katz (1963). Some methods of these two papers are from Erdős (1949). We use some methods of these three papers but none of their results.

Lanzinger (1998) studies a formally equivalent extension of Baum-Katz law. Basically, he shows that, in our Theorem B, (i) holds for $\Phi(t) = \exp(\alpha t)$ if and only if (ii) holds for $\Phi(t) = \exp(\beta t)$, and for $\varepsilon$ large enough (namely, for $\varepsilon > \beta/\alpha$). Thus, Lanzinger’s result applies to a law of large numbers if $\exp(\alpha |X|)$ is integrable for all $\alpha$. This condition is not equivalent to the integrability of $\Phi(|X|)$, for any single function $\Psi$. (If $\Psi$ is a candidate, $t = o(\log \Psi(t))$ when $t \to \infty$. Choose $\Omega$ such that $\Omega = o(\Psi)$ and $t = o(\log \Omega(t))$. Then, $\Omega(|X|)$ may be integrable while $\Psi(|X|)$ is not.)
The paper is organized as follows. Section 2 deals with the application of Theorem B to sequential analysis. In Section 3, we reduce Theorem B to the case of symmetric laws, and we state explicit universal bounds of (i), (ii) and (iii) in terms of (iii), (i) and (ii) respectively, for symmetric laws. The proofs of these explicit bounds are rather cumbersome, and we only sketch them very briefly, in Section 4. This section also proves Theorem B for non moderate functions. An appendix gives detailed proofs.

2 Sequential analysis

Theorem C answers a question of Koell (1995) about the asymptotic optimality of Wald’s sequential tests. Koell proves that (iii) implies (i) for $\varepsilon = 1/2$ when stronger assumptions hold (his result is a version of Proposition 3 below). The main result of Koell (1995) generalizes Lai (1981) and is as follows.

Assume that the unknown law of the process $(Y_n)_n$ is one of the finitely many distinct probability measures $P_i$, and call $H_i$ the hypothesis

$$H_i := \{\text{The law of } (Y_n)_n \text{ is } P_i\}.$$ 

Denote by $F_n$ the $\sigma$-field generated by $(Y_k)_{k \in \mathbb{N}}$, choose a probability $Q$ such that each $P_i$ is absolutely continuous with respect to $Q$, and a probability $P$ which is locally equivalent to each $P_i$. Call $p_{n,i}$ resp. $p_{n}$, the density of $P_i$ resp. $P$, with respect to $Q$, when restricted to $F_n$. The likelihood ratio of the sample $(Y_k)_{k \in \mathbb{N}}$ under hypothesis $H_i$ is

$$R_n^i := \frac{p_{n}(Y_1, \ldots, Y_n)}{p_{n,i}(Y_1, \ldots, Y_n)}.$$ 

A decision rule is a couple $(\tau, d)$, where the observation time $\tau$ is an integrable $(F_n)_n$ stopping time and the decision $d$ is a $\mathcal{F}_\tau$, r.v. with values in $\{H_i\}_i$. Given error probabilities $a = (a_i)_i$, a stopping time $\tau$ belongs to the class $T(a)$ if there exists a decision rule $(\tau, d)$ such that $P_i[d \neq H_i] \leq a_i$ for all $i$.

Wald’s sequential tests are commonly used decision rules, defined as follows. Fix a (multi)level $c = (c_i)_i$, and denote by $\rho_i$ the rejecting time of $H_i$ at level $c_i$, i.e.

$$\rho_i := \inf\{n \geq 1; R_n^i \geq c_i\}.$$ 

Wald’s sequential test at level $c$ is $\tau_c := \min_i \max_j c_{ij} \rho_j$ and $d_c := H_i$, where $\rho_i = \max_j \rho_j$. Theorem 2 of Koell (1995) proves the asymptotic optimality of $(\tau_c, d_c)$ with respect to $E[\tau^\tau]$. More precisely, assume that, for each $i$, $n^{-1} \log R_n^i$ $\tau$-converges under $P$. Then, for each (multi)probability of error $a = (a_i)_i$, there exists a level $c = (c_i)_i$ such that $(\tau_c, d_c)$ belongs to $T(a)$. Furthermore, $c$ can be chosen such that $\tau_c$ is asymptotically the smallest observation time in $T(a)$ when $a \rightarrow 0$, i.e.

$$E_a[\tau_c^\tau] = \inf\{E_a[\tau^\tau]; \tau \in T(a)\} \cdot (1 + o(1)),$$

for all $i$, when $a \rightarrow 0$. Lai (1981) showed this for two laws $P_1$ and $P_2$. Koell also proves this, if $t \mapsto \tau^\tau$ is replaced by an increasing function $\Phi \in \mathcal{A}$, where $\mathcal{A}$
is a strict subset of the space of continuous moderate functions. Our theorem
implies that the same result is valid for any moderate function.
As an example, assume that $\Phi$ is moderate and that the process $(Y_n)_n$ is i.i.d., and that
$$\log[p_n(Y_i)/p_0(Y_i)] \Phi(\log[p_n(Y_i)/p_0(Y_i)])$$
is integrable for every $i$. Then, $(\tau_n, d_n)$ is asymptotically optimal with respect
to $E[\Phi(\tau)]$, i.e.
$$E_1[\Phi(\tau_n)] = \inf \{ E_1[\Phi(\tau)] : \tau \in T(a) \} \cdot (1 + o(1)),$$
for every $i$, when $a \to 0$. The preceding was an incentive to prove Theorem B.

3 Effective bounds

Theorem B is a consequence of Propositions 3 to 6 below. Propositions 3 to
5 provide effective bounds of (i), (ii) and (iii) in terms of (iii), (i) and (ii)
respectively, when the law of $X$ is symmetric. Proposition 6 reduces the general
case to the case of symmetric laws. We write $\ell_c$ for $\ell^0_c$.

**Proposition 3** Let $\Phi$ be moderate with $\Phi(2t) \leq c \Phi(t)$ for all $t$, and $X$ be an
integrable r.v. For $\alpha \in [0, 1]$, let $t$ be such that $E[|X|; |X| \geq t] \leq (1 - \alpha)$. Then,
$$E[|X| \Phi(|X|)] \leq 4c^2 \left\{ t \Phi(t) + \alpha^{-1} E[\Phi(\ell^1_\alpha)] \right\}.$$

**Proposition 4** Let $p \geq 1$ be an integer, and $\Phi$ be a non decreasing function
such that $\Phi(t)/t^{p+1}$ is integrable on $(1, +\infty)$ (call this the condition (1)). Let $\Psi$
be a non decreasing function such that, for any $n \geq 1$,
$$\Psi(n) \geq n^p \sum_{k \geq n} \Phi(k) k^{-(p+1)}.$$

Then, for any symmetric $X$, $\Sigma(X, \Phi, 1) \leq K_p(X)$, where
$$K_p(X) := E[|X| \Phi(|X|)] + 2^{-p} (2p)! E[1 + |X|^{p-1} E[|X| \Phi(|X|)]]$$
If $\Phi$ is moderate, (1) holds for $p$ large enough and $\Psi = c \Phi$ satisfies (2) for $c$
large enough.

**Proposition 5** Let $\Phi$ be a non decreasing function, and $X$ be a symmetric r.v.
Then,
$$E[\Phi(\ell_1)] \leq \Phi(0) + 12 \Sigma(X, \Phi, 1/8).$$

Denote by $X'$ an independent copy of $X$ and call $X^* := X - X'$ a symmetrized
version of $X$.

**Proposition 6** Let $\Phi$ be moderate. Each of the assertions (i) to (iii) of Theorem B holds for $X$ if and only if it holds for the symmetrized $X^*$.

4
From Proposition 3, (iii) implies (i) if \( \Phi \) is moderate. From Proposition 4, (i) implies (ii) if \( \Phi \) is moderate and if the law of \( X \) is symmetric. To see that the series in (ii) then converges for every \( \varepsilon \), note that, if (i) holds for \( X \), then (i) holds for every \( X/\varepsilon \) as well. From Proposition 5, (ii) implies (iii) if \( \Phi \) is non-decreasing and if the law of \( X \) is symmetric. To see that \( \Phi(\ell_\varepsilon) \) is then integrable for any \( \varepsilon \), note that \( \ell_1 \) for the sum of the random variables \( X_n/\varepsilon \) is \( \ell_\varepsilon \) for \( S_n \).

4 Sketches of proofs

The proof of Proposition 3 uses precise estimates of \( P[\sup_{k \geq n} |X_k/k| \geq 1] \) through Poincaré's formula for the probability of a union of sets, Abel's transformation for discrete sums, and the moderation of \( \Phi \).

In Proposition 3, \( T_n \) is the sum of \( n \) truncated variables \( X_{kn} := X_k I_{\{|X_k| \leq n\}} \). A multinomial expansion allows to estimate \( E[(T_n)^2] \). Since \( X_{nk} \) is symmetric, the terms with at least one odd power disappear. A bound of the remaining terms involves a decomposition along the first index \( i \) such that \( |X_{ni}| \) is maximal amongst \( \{|X_{kn}| \}_{k \leq n} \). Precise combinatorics then yield the result.

In Proposition 5, one decomposes the sum along the intervals \([2^i, 2^{i+1}]\). From a maximal inequality for symmetric random variables, due to Lévy, the events "\( \max |S_n|/n \) is large for \( n \) in an interval" and "\( |S_n|/n \) is large for the last \( n \) in this interval" have the same probability, up to a factor of 2. One then reverses the time on each interval, uses Lévy's inequality again, and another decomposition of a sum along the intervals \([2^i, 2^{i+1}]\) concludes the proof.

Proposition 6 follows from symmetrization techniques and estimations of the probability of a deviation from any median, as exposed in Section 18.1 of Loève (1977) for example. Furthermore, if \((X'_n)_n\) is an independent copy of \((X_n)_n\), and if \((X^*_n)_n\), where \(X^*_n := X_n - X'_n\), is a symmetrization of \((X_n)_n\), one can compare the events \( \{\ell \geq n\} \), \( \{\ell' \geq n\} \) and \( \{\ell^*_\varepsilon \geq n\} \), for different values of \( \varepsilon \) and \( n \), where \( \ell \), \( \ell' \) and \( \ell^*_\varepsilon \) are the functional \( \ell_\varepsilon \), when applied to \( X' \) and \( X^* \) instead of \( X \).

The proof of Theorem B for non moderate functions is as follows. If (i) and (iii) are equivalent, \( \Phi(\ell_\varepsilon) \) is integrable for all \( \varepsilon > 0 \) as soon as \( |X| \Phi(|X|) \) is integrable. But \( \ell_\varepsilon \) for \((2X_n)_n\) is \( \ell_\varepsilon \) for \((X_n)_n\), hence \( |X| \Phi(2|X|) \) should be integrable as soon as \( |X| \Phi(|X|) \) is.

Thus, let \( \Phi \) be a non decreasing, increasing to infinity, non moderate function. Then, \( \Phi(2t_n) \geq n \Phi(t_n) \) for an increasing sequence \((t_n)_n\). Let the law of \( X \) have as support the set \( \{\pm t_n; n \geq 1\} \), and give to \( t_n \) and \(-t_n \) the same weight of \( c/(n^2 t_n \Phi(t_n)) \), for a given constant \( c \). The sum of the weights converges, hence \( c \) can be chosen so as to get a probability measure. Then, \( E[|X| \Phi(|X|)] \) is finite and \( E[|X| \Phi(2|X|)] \) is infinite. Finally, (i) and (iii) are not equivalent for \( X \).

References


LaPCS – Laboratoire de Probabilités, Combinatoire et Statistique
Université Claude Bernard Lyon-1
Domaine de Gerland
50, avenue Tony-Garnier
69366 Lyon Cedex 07 (France)

Didier.Piau@univ-lyon1.fr
Appendix: Detailed proofs

5 Proof of Proposition 3

Let $U_n := S_n/n$. Since $\ell_\varepsilon$ is an a.s. finite nonnegative integer,

$$E[\Phi(\ell_\varepsilon)] = \Phi(0) + \sum_{n \geq 1} (\Phi(n) - \Phi(n-1)) P[\sup_{k \geq n} |U_k| \geq \varepsilon].$$

(3)

Starting from an interval of length $2\varepsilon$, for example $]-\varepsilon, +\varepsilon]$, one cannot make a jump whose length is larger than $2\varepsilon$ and still be in this interval after the jump. Assume from now on that $\varepsilon = 1/2$. The preceding remark gives:

$$\{\sup_{k \geq n} |U_k| \geq 1/2\} \supset E_n := \{\sup_{k \geq n} |X_k/k| \geq 1\}.$$

The idea of the proof is to estimate precisely $P[E_n]$ as $n \to \infty$. Denote by

$$\sigma_k := \sum_{i \geq k} P[|X_i| \geq i] = \sum_{i \geq k} P[|X| \geq i].$$

Since $E_n$ is a union of independent events, Poincaré's formula at the first and second orders gives

$$\sum_{k \geq n} P[|X_k| \geq k] \left(1 - \sum_{i \geq k} P[|X_i| \geq i]\right) \leq P[E_n] \leq \sum_{k \geq n} P[|X_k| \geq k],$$

hence $\sigma_n (1 - \sigma_{n+1}) \leq P[E_n] \leq \sigma_n$. Since $X$ is integrable, $\sigma_{n+1} \to 0$ and $P[E_n] \sim \sigma_n$ when $n \to \infty$. In a more quantitative way, assume that $n \geq n_X = [t]$, where $t$ is defined in the statement of Proposition 3. Then, $n+1 \geq t$ and

$$\sigma_{n+1} \leq E[|X| - n; |X| \geq n + 1] \leq E[|X|; |X| \geq t] \leq 1 - \alpha.$$

This yields $\alpha \sigma_n \leq P[E_n]$ for $n \geq n_X$. When $n < n_X$,

$$\alpha \sigma_n \leq P[E_n] + \alpha (n_X - n).$$

Going back to the equation (3), a lower bound of $E[\Phi(\ell_{1/2})]$ is

$$E[\Phi(\ell_{1/2})] \geq \Phi(0) + \sum_{n \geq 1} (\Phi(n) - \Phi(n-1)) \alpha (\sigma_n - (n_X - n)^+) .$$

An application of Abel's transformation to the right hand member gives:

$$\sum_{n \geq 1} \Phi(n) P[|X| \geq n] \leq \alpha^{-1} E[\Phi(\ell_{1/2})] + \sum_{n=1}^{n_X-1} \Phi(n) \leq \alpha^{-1} E[\Phi(\ell_{1/2})] + (n_X - 1) \Phi(n_X - 1) \leq \alpha^{-1} E[\Phi(\ell_{1/2})] + (t - 1) \Phi(t).$$

(4)
There should be a multiple of $\Phi(0)$ in the left hand side, but the coefficient of $\Phi(0)$ is $1 - \alpha(\sigma_1 - (n_X - 1))$, which is nonnegative because:

$$\sigma_1 \leq n_X + \sigma_{n_X+1} \leq n_X + 1 - \alpha \leq n_X + \alpha^{-1} - 1.$$  

Decomposing the sum of (4) along the intervals $I(i) := [2^i, 2^{i+1}]$, one gets:

$$\sum_{n \geq 1} \Phi(n) P[|X| \geq n] \geq \sum_{i \geq 0} 2^i \Phi(2^i) P[|X| \geq 2^{i+1}]. \quad (5)$$

On the other hand,

$$E[|X| \Phi(|X|)] \leq \sum_{n \geq 0} ((n + 1) \Phi(n + 1) - n \Phi(n)) P[|X| \geq n]$$

$$\leq \Phi(1) + \sum_{i \geq 0} (2^{i+1} \Phi(2^{i+1}) - 2^i \Phi(2^i)) P[|X| \geq 2^i].$$

The last sum runs over integers $i \geq 0$. We replace the term $i = 0$ by its value and for any $i \geq 1$, we bound the coefficient of $P[|X| \geq 2^i]$ by $2^{i+1} \Phi(2^{i+1})$. This yields

$$E[|X| \Phi(|X|)] \leq 2 \Phi(2) + \sum_{i \geq 1} 2^{i+1} \Phi(2^{i+1}) P[|X| \geq 2^i]$$

$$\leq 2 \Phi(2) + 4c^2 \sum_{i \geq 0} 2^i \Phi(2^i) P[|X| \geq 2^{i+1}],$$

by an application of the moderation property of $\Phi$ between $2^{i+1}$ and $2^i$. Composing this with (4) and (5), one gets the first assertion of Proposition 3, up to an additional term in the right hand side. This extra term reads $2 \Phi(2) - 4c^2 \Phi(t)$, which is nonpositive and therefore may be omitted.

6 Proof of Proposition 4

In this section, the law of $X$ is symmetric. For $k \in [1, n]$, set

$$X_k := X_k \mathbb{1}_{|X_k| \leq n}, \quad T_n := \sum_{k=1}^n X_k.$$ 

If $|X_k| \leq n$ for all $k \in [1, n]$, then $S_n = T_n$. Hence:

$$P[|S_n| \geq n] \leq n P[|X| \geq n] + P[|T_n| \geq n].$$

We write $P[|X| \geq n]$ as a sum of $P[k \leq |X| < k + 1]$ and we use the fact that $\Phi$ is non decreasing. This yields

$$\sum_{n \geq 1} n^{-1} \Phi(n) n P[|X| \geq n] \leq E[|X| \Phi(|X|)].$$

For $P[|T_n| \geq n]$, we use the Markov inequality with an even integer $2p$, whose value will be specified later:

$$P[|T_n| \geq n] \leq n^{-2p} E[(T_n)^{2p}].$$
In the expansion of \((T_n)^{2p}\), all the terms with at least one odd power disappear because \(X\) is symmetric, hence

\[
E[(T_n)^{2p}] = \sum_c M(2p, 2c) E[X_{1n}^{2c_1} X_{2n}^{2c_2} \cdots X_{mn}^{2c_m}],
\]

where \(c = (c_i)_{i \in [1, n]}\) runs over all the nonnegative integer valued sequences of length \(n\) and of sum \(p\), and where \(M(2p, 2c)\) is the multinomial coefficient associated to the sequence \(2c\). Recall that for any \(p \geq 0\) and any nonnegative integer valued sequence \(c = (c_i)_{i \in [1, n]}\) of sum \(c_1 + c_2 + \cdots + c_n = p\), one has

\[
M(p, c) := \frac{p!}{c_1! c_2! \cdots c_n!}.
\]

If \(q \in [0, p]\), we write \(M(p, q) := M(p, (q, p-q))\) for the usual binomial coefficient.

The value of the expectation \(E[X_{1n}^{2c_1} X_{2n}^{2c_2} \cdots X_{mn}^{2c_m}]\) is entirely determined by the sequence \(d = (d_i)\) which is made by listing the non zero terms of \(c\) and by writing them in the same order. This value is:

\[
\epsilon(n, d) := E[X_1^{2d_1} X_2^{2d_2} \cdots X_q^{2d_q}; \forall i \in [1, q], |X_i| \leq n]. \tag{6}
\]

In (6), the sum of \(d\) is \(p\), the length of \(d\) is \(q\), and all the terms of \(d\) are non zero. Exactly \(M(n, q)\) sequences \(c\) give the same sequence \(d\) of length \(q\). Since \(M(n, q) \leq n^q (q!)^{-1}\), this yields

\[
E[(T_n)^{2p}] \leq \sum_d M(2p, 2d) n^q (q!)^{-1} \epsilon(n, d),
\]

where \(q\) is the length of the sequence \(d\). Summing up over \(n\), one gets

\[
\sum_{n \geq 1} n^{-1} \Phi(n) P[|T_n| \geq n] \leq \sum_d M(2p, 2d) (q!)^{-1} \epsilon(d),
\]

where \(\epsilon(d)\) is the sum

\[
\epsilon(d) := \sum_{n \geq 1} \Phi(n) n^{q-2p-1} \epsilon(n, d).
\]

We now bound each \(\epsilon(d)\). Summing up the equations (6) yields

\[
\epsilon(d) = E \left[ \prod_{i=1}^{q} |X_i|^{2d_i} \cdot \sum_{n \geq |X_0|} \Phi(n) n^{q-2p-1} \right],
\]

where \(|X_0| := \max\{|X_i|; i \in [1, q]\}\). Let \(p\) be an integer such that the condition of Proposition 4 holds. Since \(q \leq p\) for every \(d\), one has

\[
\sum_{n \geq |X_0|} \Phi(n) n^{q-2p-1} \leq |X_0|^{q-p} \sum_{n \geq |X_0|} \Phi(n) n^{-(p+1)} \leq |X_0|^{q-2p} \Psi(|X_0|).
\]

Since \(|X_i|^{2d_i-1} \leq |X_0|^{2d_i-1}\) for every \(i \in [1, q]\), one has

\[
\epsilon(d) \leq E \left[ \Psi(|X_0|) \prod_{i=1}^{q} |X_i| \right].
\]
Write $i_0$ for one of the integers such that $|X_{i_0}| = |X_0|$, for example the smallest one. The expectation of the right hand side, when restricted to the event $\{i_0 = j\}$, is at most

$$E \left[ \Psi(\{X_j\}) |X_j| \prod_{i \neq j} |X_i| : i_0 = j \right] \leq E[|X|^q] E[|X| | \Psi(\{X\})].$$

There are at most $q$ possible values of $i_0$ and, for a fixed length $q$, there are $M(p-1,q-1)$ possible values of the $q$-plet $d$. (For completeness, we prove this last assertion in Lemma 7 at the end of this section.) For a given $q$, $M(2p,2d)$ is maximal when each $d_i$ is as close as possible to $p/q$. For $q \in [1,p]$, this upper bound is maximal for $q = p$. Hence,

$$M(2p,2d) \leq c(p) := (2p)!/2^p.$$

This shows that Proposition 4 holds with the bound $E[|X| \Phi(|X|)] + c_p(X)$, where

$$c_p(X) := c(p) \sum_{q=1}^{p} M(p-1,q-1) q (q!)^{-1} E[|X|^q] E[|X| \Psi(\{X\})]$$

$$\leq c(p) E[1 + |X|^p] E[|X| \Psi(\{X\})].$$

The last assertions of Proposition 4 about moderate functions are direct consequences of the inequality $\Phi(2t) \leq c \Phi(t)$. The combinatorial lemma used during the proof is the following.

**Lemma 7** For $q \in [1,p]$, there are $M(p-1,q-1)$ sequences $d = (d_i)_{i \in [1,q]}$, integer valued, of length $q$ and sum $p$, such that $d_i \geq 1$ for every $i \in [1,q]$.

Proof of Lemma 7] We count the integer valued sequences $b = (b_i)_{i \in [1,q]}$ of sum $(p-q)$ such that $b_i \geq 0$ for every $i \in [1,q]$. The cardinal of this set is the coefficient of $z^{p-q}$ in the expansion:

$$\prod_{i=1}^{q} \left( \sum_{n=0}^{\infty} z^n \right) = (1 - z)^{-q},$$

and this coefficient is $M(p-1,q-1)$. 

7 Proof of Proposition 5

In this section, the law of $X$ is symmetric. In Section 5, (3) provides $E[\Phi(\ell_i)]$ as the sum of a series. The decomposition of this series along the intervals $I(i) := [2^i, 2^{i+1}[$ yields

$$E[\Phi(\ell_i)] \leq \Phi(0) + \sum_{\ell \geq 0} P[\ell_i \geq 2^\ell] \sum_{n \in I(\ell)} (\Phi(n) - \Phi(n-1))$$

$$\leq \Phi(0) P[\ell_i = 0] + \sum_{\ell \geq 0} (\Phi(2^{i+1} - 1) - \Phi(2^i - 1)) P[\ell_i \geq 2^\ell].$$
The probabilities written in the right hand side are bounded by
\[
P[\ell_1 \geq 2^i] = P[\sup_{n \geq 2^i} |U_n| \geq 1] \leq \sum_{j \geq i} P[\max_{n \in I(j)} |U_n| \geq 1],
\]
where \(U_n := S_n/n\). Furthermore,
\[
\{ \max_{n \in I(j)} |U_n| \geq 1 \} \subset D_j := \{ \max_{n \in I(j)} |S_n| \geq 2^j \}
\]
An interversion of the order of the summations on \(i\) and \(j\) yields
\[
E[\Phi(\ell_1)] \leq \Phi(0) + \sum_{i \geq 0} \Phi(2^{i+1}) P[D_i]. \tag{7}
\]
Since \(X\) is symmetric, zero is a median of every \(S_n\) and a maximal inequality due to Lévy, see Loève (1977), Section 18.1, states that
\[
P[\max_{n \leq m} |S_n| \geq t] \leq 2 P[|S_m| \geq t]
\]
for any \(m \geq 1\) and \(t \geq 0\). This yields an upper bound of \(P[D_i]\) as follows. First, \(D_i\) can only occur if \(|S_{2^{i+1}}| \geq 2^i/2\), or if \(D'_i\) is realized, where
\[
D'_i := \{ \exists n \in I(i) : |S_n - S_{2^i}| \geq 2^i/2 \}.
\]
We now bound the probability \(P[D'_i]\). The process
\[
(S_{2^{i+1}} - S_n)_{n \in I(i)}
\]
follows the law of the time reversal of \((S_n)_{n \in [1,2^i]}\), hence \(P[D''_i] = P[D'_i]\) with
\[
D''_i := \{ \exists n \leq 2^i, |S_n| \geq 2^i/2 \}.
\]
Lévy's inequality yields
\[
P[D''_i] \leq P[\max_{n \leq 2^{i+1}} |S_n| \geq 2^i/2] \leq 2 P[|S_{2^{i+1}}| \geq 2^i/2].
\]
Coming back to the event \(D_i\), we proved that
\[
P[D_i] \leq 3 P[|S_{2^{i+1}}| \geq 2^i/2]. \tag{8}
\]
This yields
\[
E[\Phi(\ell_1)] \leq \Phi(0) + \sum_{i \geq 0} 3 \Phi(2^{i+1}) P[|U_{2^{i+1}}| \geq 1/4]
\]
\[
= \Phi(0) + \sum_{i \geq 1} 3 \Phi(2^i) P[|U_{2^i}| \geq 1/4].
\]
On the other hand, we have to estimate
\[
s := \sum_{n \geq 1} n^{-1} \Phi(n) P[|U_n| \geq 1/8].
\]
A decomposition of \( s \) along the powers of 2 gives
\[
s \geq \sum_{i \geq 0} 2^{-(i+1)} \Phi(2^i) \sum_{n \in I(i)} P[|U_n| \geq 1/8].
\]
Recall that \( X \) is symmetric and let \( n \in I(i) \). Since the law of \( S_k \) is symmetric,
\[
P[|S_{2^i} - 2^i| \geq 2^{i-2}] \leq P[\max_{j \leq n} |S_j| \geq 2^{i-2}]
\]
\[
\leq 2P[|S_n| \geq 2^{i-2}] \leq 2P[|S_n| \geq n/8].
\]
Finally,
\[
4s \geq \sum_{i \geq 0} \Phi(2^i) P[|U_{2^i}| \geq 1/4]. \tag{9}
\]
The inequalities (7), (8) and (9) prove the assertion of Proposition 5.

8 Proof of Proposition 6

We recall some basic facts about the symmetrization of a random variable around one of its medians. A median of a real random variable \( Y \) is any \( \mu(Y) \) such that
\[
P[Y \leq \mu(Y)] \geq 1/2 \quad \text{and} \quad P[Y \geq \mu(Y)] \geq 1/2.
\]
Let \( Y' \) be an independent copy of \( Y \), independent of all the other random variables. Then, the law of \( Y^* := Y - Y' \) is symmetric and
\[
P[|Y - \mu(Y)| \geq 2t] \leq 2P[|Y^*| \geq 2t] \leq 4P[|Y| \geq t] \tag{10}
\]
for any \( t \), see Loève (1977), Section 18.1. If \( \Psi \) is non decreasing, equation (10) yields
\[
E[\Psi(|Y - \mu(Y)|)] \leq 2E[\Psi(|Y^*|)] \leq 4E[\Psi(2|Y|)]. \tag{11}
\]
In order to apply this to \( X_n \), we introduce
\[
\sum_{k=1}^{n} X^*_k = S_n - S'_n,
\]
which is a symmetrized version \( S^*_n \) of \( S_n \). Set \( U^*_n := S^*_n/n \) and let \( \Phi \) be a moderate function. Then
\[
E[|X^*| \Phi(|X^*|)] \leq 4E[|X| \Phi(2|X|)] \leq 4cE[|X| \Phi(|X|)].
\]
Hence, (i) for \( X \) implies (i) for \( X^* \). For the reverse implication, (11) yields
\[
E[|X - m| \Phi(|X - m|)] \leq 2E[|X^*| \Phi(|X^*|)],
\]
where \( m := \mu(X) \) is a median of \( X \). The expectation of the left hand side is finite, when it is restricted to the set \( \{|X - m| \geq m\} \), and, on this set, \( |X - m| \geq |X|/2 \). Hence,
\[
E[|X| \Phi(|X|/2) ; |X - m| \geq m]
\]

12
is finite. The complete expectation $E[|X| \Phi(|X|/2)]$ is finite as well. Finally, $\Phi(t) \leq e^{\Phi(t/2)}$ for any $t$, hence (i) for $X^*$ implies (i) for $X$.

The equivalence of (ii) for $X$ and (ii) for $X^*$ stems from the fact that, if $\mu(\xi_n)$ is a median of $\xi_n$ and if $\xi_n$ converges to zero in probability, then $\mu(\xi_n)$ converges to zero. (For any $\alpha > 0$, $P[|\xi_n| \geq \alpha] < \frac{1}{\alpha}$ for $n$ large enough. This means that $|\mu(\xi_n)| \leq \alpha$ for $n$ large enough. Since $\alpha$ is arbitrary, $\mu(\xi_n)$ converges to zero.)

Since $X$ is integrable and centered, $U_n$ converges almost surely to zero and $\mu(U_n)$ converges to zero. For a fixed $\varepsilon > 0$ and for $n$ large enough,

$$P[|U_n| \geq \varepsilon] \leq P[|U_n - \mu(U_n)| \geq \varepsilon/2] \leq 2P[|U_n^*| \geq \varepsilon/2] \leq 4P[|U_n| \geq \varepsilon/4],$$

where the convergence of $\mu(U_n)$ to zero yields the first inequality and (10) at the beginning of this section yields the two other inequalities. The equivalence for (ii) holds.

For (iii), we start with the remark that

$$\{\ell \geq n\} = \{\exists k \geq n, |U_k| \geq \varepsilon\} = \{\sup_{k \geq n} |U_k| \geq \varepsilon\}.$$

The first equality is the definition of $\ell$, the second equality is a consequence of the almost sure convergence of $U_n$ to zero. (Hence, the supremum is in fact almost surely a maximum.) Write $\ell^c_\varepsilon$, resp. $\ell^c_\varepsilon$, for the functional $\ell_\varepsilon$ associated to $(X_n)_n$, resp. to $(X^*_n)_n$, rather than to $(X_n)_n$. We rely on the following inclusions:

$$\{\ell^c_\varepsilon \geq n\} \subset \{\ell_\varepsilon \geq n\} \cup \{\ell' \geq n\},$$

and

$$\{\ell_\varepsilon \geq n\} \cap \{\ell' \leq n\} \subset \{\ell^c_\varepsilon \geq n\}.$$

The first inclusion yields $P[\ell^c_\varepsilon \geq n] \leq 2P[\ell_\varepsilon \geq n]$, and

$$E[\Phi(\ell^c_\varepsilon)] \leq 2E[\Phi(\ell_\varepsilon)].$$

In the second inclusion, $\ell_\varepsilon$ and $\ell'$ are independent and $\ell'_\varepsilon$ is a.s. finite, hence $P[\ell'_\varepsilon \leq n]$ converges to 1 and $P[\ell_\varepsilon \geq n]$ is equivalent to $P[\ell^c_\varepsilon \geq n]$ when $n \to \infty$. This proves that $\Phi(\ell^c_\varepsilon)$ is integrable as soon as $\Phi(\ell_\varepsilon)$ is. Finally, (iii) for $X^*$ holds iff (iii) for $X$ does. This ends the proof of Proposition 6.

9 The non moderate case

The proof of Theorem B for non moderate functions is as follows. If (i) and (iii) are equivalent, $\Phi(\ell_\varepsilon)$ is integrable for all $\varepsilon > 0$ as soon as $|X| \Phi(|X|)$ is integrable. But $\ell^c_\varepsilon$ for $(2X_k)_n$ is $\ell_\varepsilon$ for $(X_n)_n$, hence $|X| \Phi(2|X|)$ should be integrable as soon as $|X| \Phi(|X|)$ is.

Let $\Phi$ be a non decreasing, increasing to infinity, non moderate function. Then, $\Phi(2t_n) \geq n\Phi(t_n)$ for an increasing sequence $(t_n)_n$. Let the law of $X$ have as support the set $\{\pm t_n; n \geq 1\}$, and give to $t_n$ and $-t_n$ the same weight of $c/(n^2 t_n \Phi(t_n))$, for a given constant $c$. The sum of the weights converges, hence
can be chosen so as to get a probability measure. Then, $E[X \mid \Phi(|X|)]$ is finite and $E[X \mid \Phi(2X)]$ is infinite. Thus, (i) and (iii) are not equivalent for $X$.

The following properties of moderate functions are direct consequences of the definition and we state them without proof.

Power and logarithmic functions are moderate, exponential functions are not. The function $\Phi(t) = t^c \log(t)^c \log \log(t)^d \ldots$ is moderate if the first non-zero exponent is positive. If $\Phi(t) \sim at^c$ when $t \to \infty$, with $c$ nonnegative, then $\Phi$ is moderate. If $\Phi$ is moderate, then $\Phi(t) \leq ct^c$, for $c$ large enough and $t \geq 1$. There exists a non moderate $\Phi$, such that $\Phi(t) \leq ct^c$, for any positive $c$. There exists non moderate $\Phi$, which are differentiable and such that $\Phi'(t) \leq t^c$, for any positive $c$. For any moderate $\Phi$, there exists a smooth moderate $\Psi$ and $c \geq 1$, such that $c^{-1} \Psi \leq \Phi \leq c\Psi$. One can choose $\Psi$ such that $\Psi(t+1)/\Psi(t) \leq 1 + a/(t+1)$, for $a$ large enough. Finally, there exists moderate $\Phi$, such that the limit set of $\log \Phi(t)/\log(t)$ as $t \to \infty$ is the interval $[0,c]$, for any positive $c$. 