A Note on the Weak Law of Large Numbers for Free Random Variables

Raluca Balan* University of Ottawa

George Stoica†‡ University of New Brunswick

July 28, 2006

Abstract

In this article we prove that if $\{X_k\}_{k \geq 1}$ are free identically distributed random variables with common distribution μ, and $h(n), g(n)$ are positive constants such that the function $f(n) = g(n)h(n)$ can be extended to $(0, \infty)$ and satisfies

$$\lim_{t \to \infty} f^{-1}(t)\ \mu(\{x; |x| > t\}) = 0$$

then $g(n)^{-1} \sum_{k=1}^{n} X_k / h(k) - M_n \xrightarrow{d} \delta_0$ for some constants M_n. This is obtained under certain regularity conditions imposed on g, h.

Keywords: weak law of large numbers, noncommutative probability theory; regularly varying functions.

1 Introduction

A classical result in probability theory is Kolmogorov's weak law of large numbers (WLLN), which states that if $S_n = \sum_{k=1}^{n} X_k$ is the partial sum of a sequence of independent identically distributed random variables with common distribution μ (whose first order moment may not exist), then $S_n/n - M_n \xrightarrow{p} 0$ for suitable constants M_n, if and only if

$$\lim_{t \to \infty} t\ \mu(\{x; |x| > t\}) = 0.$$

*Corresponding Author. *Postal address*: University of Ottawa, Department of Mathematics and Statistics, 585 King Edward Avenue, Ottawa ON, K1N 6N5, Canada. *E-mail address*: rbalan@uottawa.ca

†*Postal address*: University of New Brunswick, Department of Mathematical Sciences, PO Box 5050, Saint John NB, E2L 4L5, Canada. *E-mail address*: stoica@unbsj.ca

‡Both authors were supported by grants from NSERC of Canada.

This fundamental limit theorem was extended by Bercovici and Pata (1996) and Pata (1996a) to the context of noncommutative probability theory, in which the classical probability space $(\Omega, \mathcal{F}, P)$ is replaced by a *noncommutative probability space* $(\mathcal{A}, \varphi)$, and the random variables $\{X_k\}_{k\geq 1}$ are self-adjoint operators affiliated with $\mathcal{A}$. More precisely, Pata's (1996a) WLLN states that if $S_n = \sum_{k=1}^n X_k$ is the partial sum of a sequence of *free* identically distributed random variables $\{X_k\}_{k\geq 1}$ with common distribution μ and $\beta > 1/2$, then $S_n/n^\beta - M_n \xrightarrow{d} \delta_0$ for suitable constants M_n, if and only if

$$\lim_{t\to\infty} t^{1/\beta}\ \mu(\{x; |x| > t\}) = 0.$$

The purpose of this paper is to prove that a similar WLLN continues to hold for a sequence of free identically distributed random variables X_k which are "weighted" by constants $h(k)$, and whose partial sum $S_n = \sum_{k=1}^n X_k/h(k)$ is normalized by another constant $g(n)$. This is obtained under the condition

$$(C) \qquad \lim_{t\to\infty} f^{-1}(t)\ \mu(\{x; |x| > t\}) = 0,$$

where $f(n) = g(n)h(n)$ and we assume that the function f can be extended to $(0, \infty)$ (and satisfies some additional conditions). This line of research was inspired by the result of Jaite (2003) in the context of classical probability theory, who proved that $S_n/g(n) - M_n \xrightarrow{\text{a.s.}} 0$ for suitable constants M_n, if and only if

$$(C^*) \qquad \int_{-\infty}^{\infty} f^{-1}(t) d\mu(t) < \infty.$$

We begin by introducing the terminology and notation specific to noncommutative probability theory.

A *W^*-probability space* is a pair $(\mathcal{A}, \varphi)$ be a probability space, where $\mathcal{A}$ is a complex unital von Neumann subalgebra of some $L(H)$ (the space of bounded linear operators on a Hilbert space H) and φ is a normal faithful trace. A *random variable* X is a self-adjoint operator affiliated with $\mathcal{A}$, i.e. $u(X) \in \mathcal{A}$ for any bounded Borel function u on $\mathbf{R}$. The *distribution* of a random variable X is a probability measure on $\mathbf{R}$ given by $\mu_X = \varphi \circ E_X$, where E_X is the spectral measure of X. We have

$$\varphi(u(X)) = \int_{-\infty}^{\infty} u(t) d\mu_X(t)$$

for every bounded Borel function u on $\mathbf{R}$. A sequence of random variables $\{X_n\}_{n\geq 1}$ *converges in distribution* to a probability measure ν (and we write $X_n \xrightarrow{d} \nu$) if μ_{X_n} converges to ν weakly. Finally, the random variables $\{X_n\}_{n\geq 1}$ are *independent* if $\varphi(A_j A_k) = \varphi(A_j)\varphi(A_k)$ for every $A_j \in W^*(X_j), A_k \in W^*(X_k)$, with $j \neq k$, where $W^*(X_i)$ denotes the smallest von Newmann to which each element of X_i is affiliated (see p. 592, Pata, 1996a).

Throughout this paper, we let $\{g(n)\}_n$ be a nondecreasing sequence of positive numbers with $g(n) \to \infty$, $\{h(n)\}_n$ a sequence of positive numbers and $f(n) = g(n)h(n)$. We denote by C a generic constant, which may be different from line to line.

2 Results and Examples

Our first theorem is the noncommutative analogue of the classical "degenerate convergence criterion" (see e.g. Theorem 5.2.3 of Chung, 2001, or p.278, Loève, 1963). Its proof uses the same "truncation" procedure as in the classical case which carries over to the noncommutative framework.

Theorem 2.1 *Let $\{X_k\}_{k\geq 1}$ be a sequence of independent random variables in a W^*-probability space $(\mathcal{A}, \varphi)$, and denote by μ_k the distribution of X_k. If*

$$(C1) \qquad \sum_{k=1}^{n} \mu_k(\{x; |x| > f(n)\}) \to 0$$

$$(C2) \qquad \frac{1}{g(n)^2} \sum_{k=1}^{n} \frac{1}{h(k)^2} \int_{-f(n)}^{f(n)} t^2 d\mu_k(t) \to 0$$

then

$$\frac{1}{g(n)} \sum_{k=1}^{n} \frac{1}{h(k)} \left(X_k - \int_{-f(n)}^{f(n)} t d\mu_k(t) \right) \xrightarrow{d} \delta_0.$$

Proof: For all $1 \leq k \leq n$, we define the "truncated" random variables $X^*_{k,n} = X_k E_{X_k}([-f(n), f(n)])$ (recalling that E_{X_k} is the spectral measure of the operator X_k), as well as the associated "first order moments" $m_{k,n} = \int_{-f(n)}^{f(n)} t d\mu_k(t)$. Let

$$S_n = \frac{1}{g(n)} \sum_{k=1}^{n} \frac{1}{h(k)} X_k, \quad S^*_n = \frac{1}{g(n)} \sum_{k=1}^{n} \frac{1}{h(k)} X^*_{k,n}, \quad M_n = \frac{1}{g(n)} \sum_{k=1}^{n} \frac{1}{h(k)} m_{k,n}.$$

We want to show that $S_n - M_n$ converges in distribution to δ_0, i.e.

$$\mu_{S_n - M_n}(\Delta_\varepsilon) \to 0 \quad \forall \varepsilon > 0 \tag{1}$$

where $\Delta_\varepsilon = \{x; |x| > \varepsilon\}$. By Proposition 3.1 (Pata, 1996a),

$$\mu_{S_n - M_n}(\Delta_\varepsilon) \leq \mu_{S_n - S^*_n}(\Delta_0) + \mu_{S^*_n - M_n}(\Delta_\varepsilon).$$

Using repeatedly Proposition 3.1 (Pata, 1996a) and condition $(C1)$, we get

$$\mu_{S_n - S^*_n}(\Delta_0) \leq \sum_{k=1}^{n} \mu_{h(k)^{-1}(X_k - X^*_{k,n})}(\Delta_0) = \sum_{k=1}^{n} \mu_{X_k - X^*_{k,n}}(\Delta_0)$$

$$= \sum_{k=1}^{n} \mu_{X_k E_{X_k}(\Delta_{f(n)})}(\Delta_0) = \sum_{k=1}^{n} \mu_k(\Delta_{f(n)}) \to 0.$$

Using Chebyshev's inequality, the independence of X_k's and condition $(C2)$,

$$\mu_{S_n^* - M_n}(\Delta_\varepsilon) \le \frac{1}{\varepsilon^2}\varphi((S_n^* - M_n)^2) \le \frac{1}{\varepsilon^2 g(n)^2}\sum_{k=1}^{n}\frac{1}{h(k)^2}\varphi((X_{k,n}^*)^2) =$$

$$\frac{1}{\varepsilon^2 g(n)^2}\sum_{k=1}^{n}\frac{1}{h(k)^2}\int_{-f(n)}^{f(n)} t^2 d\mu_k(t) \to 0$$

This concludes the proof of (1). □

When the random variables $\{X_k\}_{k\ge 1}$ have a common distribution μ, $(C1)$ becomes:

$$(C') \qquad \lim_{n\to\infty} n\ \mu(\{x : |x| > f(n)\}) = 0$$

(which is almost the same as (C)). Moreover, a result on p. 192 of Adler and Rosalsky (1991) says that (C') implies $(C2)$ (with $\mu_k = \mu$ for all k), if either one of the following sets of conditions hold:

$$(F1) \qquad f\uparrow, \quad \frac{f(n)}{n}\downarrow, \quad A(n) := \frac{1}{g(n)^2}\sum_{k=1}^{n}\frac{1}{h(k)^2} = o(1), \quad \sum_{k=1}^{n}\frac{f(k)^2}{k^2} = O(A_n^{-1})$$

$$(F2) \qquad \frac{f(n)}{n}\uparrow, \quad \sum_{k=1}^{n}\frac{1}{h(k)^2} = O\left(\frac{n}{h(n)^2}\right)$$

(Here the symbols $u_n\uparrow$ or $u_n\downarrow$ are used to indicate that the sequence $\{u_n\}_n$ is nondecreasing, respectively nonincreasing.) Hence, we obtain the following:

Corollary 2.2 *Let $\{X_k\}_{k\ge 1}$ be a sequence of independent random variables in a W^*-probability space, with common distribution μ. If $(F1)$-$(F2)$ and (C') hold, then*

$$\frac{1}{g(n)}\sum_{k=1}^{n}\frac{1}{h(k)}\left(X_k - \int_{-f(n)}^{f(n)} t\ d\mu(t)\right) \xrightarrow{d} \delta_0.$$

Example 2.3 Let $g(n) = n^a, a \ge 0$ and $h(n) = n^b, b > 1/2$. We see that $(F1)$, respectively $(F2)$ hold, depending on whether $a + b < 1$ or $a + b \ge 1$. Later in this paper, we will improve this example by requiring only $a, b \ge 0; a + b > 1/2$; see Example 2.10.

Example 2.4 If h is nonincreasing and $f(n)/n$ is nondecreasing, then $(F2)$ is satisfied. As an example we may take $g(n) = n^\rho, \rho > 1$ and $h(n) = 1/\log n$.

In order to enlarge the class of examples, we will use a variant of Theorem 2.1, which can be obtained for instance by applying Theorem 2.1 to the random variables $\tilde{X}_k = X_k/h(k)$ and the sequences $\tilde{g}(n) = g(n)$, $\tilde{h}(n) = 1$. Note that in this case $\tilde{X}_k$ has distribution $\tilde{\mu}_k$ defined by $\tilde{\mu}_k(B) := \mu_k(h(k)B)$ for any Borel set $B \subseteq \mathbf{R}$.

Theorem 2.5 *Let $\{X_k\}_{k\geq 1}$ be a sequence of independent random variables in a W^*-probability space, and denote by μ_k the distribution of X_k. If*

$$(\widetilde{C}1) \qquad \sum_{k=1}^{n} \mu_k(\{x; |x| > g(n)h(k)\}) \to 0$$

$$(\widetilde{C}2) \qquad \frac{1}{g(n)^2} \sum_{k=1}^{n} \frac{1}{h(k)^2} \int_{-g(n)h(k)}^{g(n)h(k)} t^2 d\mu_k(t) \to 0$$

then

$$\frac{1}{g(n)} \sum_{k=1}^{n} \frac{1}{h(k)} \left(X_k - \int_{-g(n)h(k)}^{g(n)h(k)} t \; d\mu_k(t) \right) \xrightarrow{d} \delta_0.$$

Suppose now that the random variables X_k have common distribution μ. Our goal is to show that in this case, (C) implies both $(\widetilde{C}1)$ and $(\widetilde{C}2)$.

For this we will assume that $g(n), h(n)$ satisfy

$$(F3) \qquad m := \inf_{n\geq 1} h(n) > 0$$

$$(F4) \qquad \sum_{k=1}^{n} \frac{1}{f^{-1}(g(n)h(k))} \leq C, \quad \text{for } n \text{ large enough}$$

and the function f can be extended to $(0, \infty)$ and satisfies:

$(F5)$ the inverse f^{-1} of f exists and satisfies $\lim_{t\to\infty} f^{-1}(t) = \infty$

$(F6)$ f is regularly varying at ∞ with index $\rho > 1/2$, i.e. for every $\lambda > 0$ $\lim_{x\to\infty} f(\lambda x)/f(x) = \lambda^{\rho}$

Let $d_{k,n} = g(n)h(k)$ for all $k \leq n$ and $\tau(t) = f^{-1}(t)\; \mu(\{x; |x| > t\})$ for $t > 0$.

Proposition 2.6 *Under* $(F3)$-$(F5)$, (C) *implies* $(\widetilde{C}1)$.

Proof: Let $\varepsilon > 0$ be arbitrary. By (C), there exists $T = T_\varepsilon$ such that $\tau(t) < \varepsilon$ for all $t \geq T$. Since $g(n) \to \infty$ and $(F3)$ holds, there exists $N = N_\varepsilon$ such that $d_{k,n} \geq mg(n) \geq T$ for all $n \geq N, k \leq n$. Therefore, for every $n \geq N$

$$\sum_{k=1}^{n} \mu(\{x; |x| > d_{k,n}\}) = \sum_{k=1}^{n} \frac{1}{f^{-1}(d_{k,n})} \tau(d_{k,n}) \leq \varepsilon \sum_{k=1}^{n} \frac{1}{f^{-1}(d_{k,n})} \leq C\varepsilon,$$

where we used $(F4)$ for the last inequality. This concludes the proof. □

The following lemma generalizes Lemma 4.(iii) (Bercovici and Pata, 1996) to the case of invertible, regularly varying functions f.

Lemma 2.7 *Under $(F5)$-$(F6)$, (C) implies that*

$$\lim_{y\to\infty} \frac{f^{-1}(y)}{y^k} \int_{-y}^{y} |t|^k d\mu(t) = 0, \quad \forall k \geq 2.$$

Proof: Using integration by parts, we have

$$\int_{-y}^{y} |t|^k d\mu(t) = -y^k \mu(\{x; |x| > y\}) + k \int_0^y t^{k-1} \mu(\{x; |x| > t\}) dt.$$

The result will follow once we prove that

$$\lim_{y\to\infty} \frac{f^{-1}(y)}{y^k} \int_0^y t^{k-1} \mu(x; |x| > t) dt = 0. \tag{2}$$

Note that f^{-1} is regularly varying at ∞ with index $1/\rho$. Using Karamata's Representation Theorem (see Theorems 1.3.1, 1.4.1, Bingham, Goldie and Teugels, 1987), respectively Potter's Theorem (see Theorem 1.5.6.(iii), Bingham, Goldie and Teugels, 1987), we know that for every $\delta > 0, A > 1$ there exists $C = C_\delta > 0, Y = Y_{\delta,A} > 0$ such that

$$f^{-1}(y) \leq C y^{(1/\rho)+\delta}, \quad \forall y \geq Y \tag{3}$$

$$\frac{f^{-1}(y)}{f^{-1}(t)} \leq A \left(\frac{y}{t}\right)^{(1/\rho)+\delta}, \forall y \geq t \geq Y \tag{4}$$

Let $0 < \delta < k - (1/\rho)$ and $A > 1$ be fixed. Let $\varepsilon > 0$ be arbitrary. By (C), there exists $N = N_\varepsilon > Y$ such that $\tau(t) < \varepsilon,\ \forall t > N$. Using (3),

$$\frac{f^{-1}(y)}{y^k} \int_0^N t^{k-1} \mu(x; |x| > t) dt \leq N^{k-1} \frac{f^{-1}(y)}{y^k} \leq N^{k-1} C y^{(1/\rho)+\delta-k} \leq \varepsilon \tag{5}$$

for y large enough. Using (4),

$$\frac{f^{-1}(y)}{y^k} \int_N^y t^{k-1} \mu(x; |x| > t) dt \leq \varepsilon \frac{f^{-1}(y)}{y^k} \int_N^y \frac{t^{k-1}}{f^{-1}(t)} dt \leq \frac{A\epsilon}{y^k} \int_N^y t^{k-1} \left(\frac{y}{t}\right)^{(1/\rho)+\delta} dt$$

$$= \frac{A\varepsilon}{y^\alpha} \int_N^y t^{\alpha-1} dt = \frac{A\varepsilon}{\alpha} \left[1 - \left(\frac{N}{y}\right)^\alpha\right] \leq \frac{A\varepsilon}{\alpha} \tag{6}$$

where $\alpha := k - (1/\rho) - \delta > 0$. The proof of (2) is complete by (5)- (6).
□

Using the previous lemma, we obtain the following result.

Proposition 2.8 *Under* $(F3)$-$(F6)$, (C) *implies* $(\widetilde{C}2)$.

Proof: Let $\varepsilon > 0$ be arbitrary. By Lemma 2.7, there exists $Y = Y_\varepsilon$ such that

$$v(y) := \frac{f^{-1}(y)}{y^2} \int_{-y}^{y} |t|^2 d\mu(t) \leq \varepsilon, \quad \forall y \geq Y.$$

Since $g(n) \to \infty$ and $(F3)$ holds, there exists $N = N_\epsilon$ such that $d_{k,n} > Y$ for all $n \geq N, k \leq n$. Hence, using $(F4)$ we have

$$\sum_{k=1}^{n} \frac{1}{d_{k,n}^2} \int_{-d_{k,n}}^{d_{k,n}} t^2 d\mu(t) = \sum_{k=1}^{n} \frac{v(d_{k,n})}{f^{-1}(d_{k,n})} \leq \varepsilon \sum_{k=1}^{n} \frac{1}{f^{-1}(d_{k,n})} \leq C\varepsilon$$

for all $n \geq N$, which concludes the proof. □

In summary, Theorem 2.5, Proposition 2.6 and Proposition 2.8, lead us to the following result.

Corollary 2.9 *Let* $\{X_k\}_{k\geq 1}$ *be a sequence of independent random variables in a* W^**-probability space, with common distribution* μ*. If* $(F3)$-$(F6)$ *and* (C) *hold, then*

$$\frac{1}{g(n)} \sum_{k=1}^{n} \frac{1}{h(k)} \left(X_k - \int_{-g(n)h(k)}^{g(n)h(k)} t \; d\mu(t) \right) \xrightarrow{d} \delta_0.$$

Example 2.10 Let $g(n) = n^a$ and $h(n) = n^b$, where $a, b \geq 0,\ a + b > 1/2$. In this case $f(n) = n^{a+b}$ and $(F3)$-$(F6)$ hold: $(F3), (F5)$ and $(F6)$ can be checked directly, whereas for $(F4)$ we have:

$$\sum_{k=1}^{n} (n^a k^b)^{-1/(a+b)} = n^{-a/(a+b)} \sum_{k=1}^{n} k^{-b/(a+b)} \leq n^{-a/(a+b)} \cdot C n^{1-b/(a+b)} = C.$$

Example 2.11 If h is nonincreasing and f is nondecreasing (for x large enough), then $(F4)$ holds:

$$\sum_{k=1}^{n} \frac{1}{f^{-1}(g(n)h(k))} \leq \sum_{k=1}^{n} \frac{1}{f^{-1}(f(n))} = \sum_{k=1}^{n} \frac{1}{n} = 1.$$

As an example, we may take

$$g(n) = n^a l_1(n) \quad \text{and} \quad h(n) = 1 + \frac{1}{n^b l_2(n)}$$

where $a, b \geq 0,\ a - b > 1/2$ and $l_1(x), l_2(x)$ are slowly varying functions such that $\lim_{x\to\infty} l_1(x) = \lim_{x\to\infty} l_2(x) = \infty$ and $l_1(x)/l_2(x)$ is nondecreasing for

x large (e.g. $l_1(x) = \log x$, $l_2(x) = \log\log x$). In this case $f(n) = n^a l_1(n) + n^{a-b} l_1(n)/l_2(n)$ and $(F3)$-$(F6)$ hold. In particular, by taking

$$g(n) = n^\rho l(n) \quad \text{and} \quad h(n) = 1$$

where $\rho > 1/2$ and $l(x)$ is a slowly varying function, we enlarge the class of examples considered by Gut (2004), where $\rho > 1$.

Concluding Remarks: (a) Corollary 2.9 extends several results in noncommutative probability theory, such as the Kolmogorov WLLN (cf. Bercovici and Pata, 1996) and the Marcinkiewicz WLLN (cf. Pata, 1996a), and in addition gives new such WLLN's considering for instance regularly varying weights.

(b) Propositions 2.6, 2.8 show that under $(F3) - (F6)$, (C) is a sufficient condition for the WLLN with regularly varying normalizing constants even in the classical sense, which seems to be a new result. (See Theorem 1.3 of Gut, 2004 which treats the case $h(n) = 1$.)

(c) In view of the the central limit theorem for free random variables (cf. Pata, 1996b), the index ρ in $(F6)$ has to be strictly larger than 1/2.

(d) The case of logarithmic averages (i.e. $g(n) = \log n, h(n) = n$) is not covered either by conditions $(F1) - (F2)$, nor by conditions $(F3) - (F6)$. In the classical theory, condition (C^*) (with $f(t) := t \log t$) is known to be necessary and sufficient for the *strong* LLN (see Jaite, 2003). However, even in this classical setting, it is not clear whether the WLLN for logarithmic averages holds, under (C) alone.

(e) The necessity of condition (C) for the WLLN (in Corrolary 2.9) remains an open problem. In order to tackle this problem, one would have to make use of the complex function

$$\phi_\mu(z) = F_\mu^{-1}(z) - z$$

which is the noncommutative analogue of the classical log-characteristic function. Here $F_\mu(z) = 1/G_\mu(z)$ with $G_\mu(z) = \int_{-\infty}^{\infty} 1/(z - x) d\mu(x)$ and F_μ^{-1} is the inverse with respect to composition.

References

[1] Adler, A. and Rosalsky, A. (1991), On the weak law of large numbers for normed weighted sums of i.i.d. random variables, *Intern. J. Math. & Math. Sci.* **14**, 191-202.

[2] Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987), *Regular Variation* (Cambridge University Press, Cambridge).

[3] Bercovici, H. and Pata, V. (1996), The law of large numbers for free identically distributed random variables, *Ann. Probab.* **24**, 453-465.

[4] Chung, K. L. (2001), *A Course in Probability Theory*, Third Edition (Academic Press, San Diego).

[5] Gut, A. (2004), An extension of the Kolmogorov-Feller weak law of large numbers with an application to the St. Petersburg game, *J. Theoret. Probab.* **17**, 769-779.

[6] Jaite, R. (2003), On the strong law of large numbers, *Ann. Probab.* **31**, 409-412.

[7] Loève, M. (1963), *Probability Theory*, Third Edition (D. van Nostrand Company, Princeton, New Jersey).

[8] Pata, V. (1996a), A generalized weak law of large numbers for noncommuting random variables, *Indiana Univ. Math. J.* **45**, 591-601.

[9] Pata, V. (1996b), The central limit theorem for free additive convolution, *J. Funct. Anal.* **140**, 359-380.