



HAL
open science

A sufficient condition for the CLT in the space of nuclear operators - Application to covariance of random functions

Andre Mas

► To cite this version:

Andre Mas. A sufficient condition for the CLT in the space of nuclear operators - Application to covariance of random functions. 2018. hal-01820960

HAL Id: hal-01820960

<https://hal.science/hal-01820960>

Preprint submitted on 22 Jun 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A sufficient condition for the CLT in the space of nuclear operators - Application to covariance of random functions.

André MAS*
Université Montpellier 2

Abstract

We give a sufficient condition for the CLT to hold in the space of trace class (nuclear) operators. This condition turns out to be adapted to the asymptotic study of empirical covariance operators of Hilbert valued random variables.

Key words: Nuclear operators; Central limit theorem; Covariance operators; Random functions

1 About the CLT for covariance operators

This article aims at providing a sufficient condition for the central limit theorem to hold in a special vector space of operators. The reasons for this approach are motivated by their applications in a modern but quite promising area of statistics : the inference on random curves (or random functions). Indeed, let X_1, \dots, X_n be a sample of centered random curves. We will assume that X_i are random variables with values in a Hilbert space of functions H endowed with inner product $\langle \cdot, \cdot \rangle$ and norm $|\cdot|$. A basic tool for studying this sample is the empirical covariance operator Γ_n defined for all x in H by :

$$\Gamma_n(x) = \frac{1}{n} \sum_{k=1}^n \langle X_k, x \rangle X_k$$

Several statistical applications arise from Γ_n : functional principal component analysis, canonical covariance analysis, decomposition of gaussian curves, etc. In the case of an i.i.d. sample Dauxois, Pousse and Romain (1982) provided a complete asymptotic study of Γ_n and of its spectral elements. Silverman and Rice (1991) proposed estimation procedures for the mean and the covariance operator of random curves. Bosq (2000) in a book devoted to the generalization of ARMA models to the infinite-dimensional setting gave several asymptotic results for Γ_n when the X_k 's are dependent variables. Cardot, Ferraty, Sarda (1999) had to study such covariance operators in the framework of a functional linear regression model. Mas (2002) proved the asymptotic gaussianity of the vector of

*Mailing address : Laboratoire de Probabilités-Statistiques, CC051, Université Montpellier 2, place Eugène Bataillon, 34095 Montpellier Cedex 5, mas@math.univ-montp2.fr

covariance and cross-covariance of a generalized linear process with values in a Hilbert space.

If u and v are two vectors in H , by $u \otimes v$ we denote the rank one operator defined for all x in H by $(u \otimes v)(x) = \langle u, x \rangle v$.

Let us suppose that the sample X_1, \dots, X_n is strictly stationary and denote $\Gamma = \mathbb{E}(X_1 \otimes X_1)$. In the sequel we will study weak convergence for the random operator $\Gamma_n - \Gamma$. It is a well-known fact that spaces of operators on Hilbert spaces may be endowed with several topologies or norms that are not equivalent if H is infinite dimensional. It should be stressed that functional statisticians tend to match the regularity of underlying processes. Consequently the case $H = L^2[a, b]$ is not that common and Hilbert spaces of smoother functions such as Sobolev spaces are usually preferred.

In all the previous articles cited above, the space of operators is endowed with the Hilbert-Schmidt topology (defined below) for at least two reasons. First, computations are obviously simpler in this setting since Γ_n may be viewed as another Hilbert-valued random variable. Second, the central limit theorem is obtained under mild condition in this framework. This choice was made for convenience but cannot be justified otherwise. It turns out that, unfortunately, the natural space is in fact the space of trace class operators. We refer to the next section for definition and main properties of these operator spaces. The word "natural" could be explained here by the following well-known fact (see for example Vakhania, Tarieladze, Chobanyan (1987)) : **A random variable in a Hilbert space has strong moment of second order if and only if its covariance operator is a trace class operator.** With mathematical symbols, this means :

$$\mathbb{E}|X_1|^2 < +\infty \text{ iff } \Gamma \text{ is trace class.}$$

When only "weak" moments (Pettis integral) are finite the Hilbert-Schmidt assumption is no longer true since the covariance operator is just bounded then. In order to illustrate the situation let us switch operators to sequences. *This would mean that, up to the present day, l^1 sequences were dealt only with l^2 tools.* Consequently, the question of the central limit theorem in the space of trace class operators must be addressed.

As announced above, we will restrict ourselves to a sufficient condition for two reasons. It is probable that the necessary condition will not be obtained by usual probabilistic tools but requires very deep results related to functional analysis. Second, the condition, as enounced below, may be easily exploited in the special cases when the random operators are covariance (or cross-covariance operators).

The next paragraph outlines very basic but useful facts in operator theory.

2 Trace class operators

When looking for examples to illustrate general theorems on separable Banach spaces one commonly refers to spaces of functions (the $L^p(S, \mathcal{S}, \mu)$ with usual notations) amongst which the l_p 's or c_0 . Banach spaces of linear operators are rarely mentioned. They are however of much importance both in probability theory (especially in weak convergence theory, gaussian analysis, etc.) and in statistics where such random operators appear in quantum physics or when studying the covariance structure of random functions.

Here we will consider operators on an infinite dimensional separable Hilbert space H and not on a Banach space. The Hilbert space setting is quite general and allows to give conditions that are easily understood. Even if \mathcal{L} the space of bounded linear operators is not separable, it is possible to consider an important subspace : the space of compact

operators, \mathcal{C} , endowed with the usual operator norm. For the definition and the main algebraic and analytic properties of \mathcal{C} we refer to Dunford-Schwartz (1988).

We may now define the class $\mathcal{C}_p \subset \mathcal{C}$. Let T be a compact operator and $\lambda_k(T^*T)$ be the k^{th} positive eigenvalue of T^*T , (T^* denotes the adjoint operator of T). Set $\mu_k = \sqrt{\lambda_k}$, the k^{th} singular value of operator T , then fix $p \in \mathbb{N}$ and define :

$$\mathcal{C}_p = \left\{ T \in \mathcal{C} : \sum_k \mu_k^p < +\infty \right\}$$

The "natural" norm on this space is defined for all T and for $p \geq 1$ by $\|T\|_p = [\sum_k \mu_k^p]^{1/p}$. Then, $(\mathcal{C}_p, \|\cdot\|_p)$ is a separable Banach space and $\mathcal{C}_p \subset \mathcal{C}_q$ whenever $p < q$. Several other properties of \mathcal{C}_p are listed in Dunford-Schwartz (1988).

Note also that spaces of operators are non commutative algebras, whereas algebras of functions are usually commutative and that \mathcal{C}_2 is the space of Hilbert-Schmidt operators, which is itself a separable Hilbert space.

In the following we will focus on \mathcal{C}_1 the space of nuclear (or trace class) operators on H endowed with nuclear norm $\|\cdot\|_1$, which is the thinnest of all the $\|\cdot\|_p$ norm. In the case of a symmetric and positive operator, this norm is nothing but the sum of the eigenvalues. We will need the three following results. For further purpose we recall that the dual space of \mathcal{C}_1 is $\mathcal{C}_1^* = \mathcal{L}$. The duality bracket between $T \in \mathcal{C}_1$ and $U \in \mathcal{L}$ is $\langle\langle T, U \rangle\rangle = \text{tr} T^*U$.

Nicole Tomaszak-Jaegerman (1974) proved that \mathcal{C}_1 is a Banach space of cotype 2. This result is crucial as far as the central limit theorem is under concern. It is a well known fact that in such a Banach space, a random variable satisfies the CLT if and only if it is pregaussian. In other words, if its covariance operator is a gaussian covariance operator. We refer to Araujo and Giné (1980) for more information about the theory of types in Banach spaces. Unfortunately this approach that may lead us to a necessary and sufficient condition goes beyond the scope of this article. Note anyway that conversely to the other famous cotype 2 space l_1 , the CLT does not easily stems from the pre-gaussianity condition because \mathcal{C}_1 is not a commutative space.

We have to change our technique and to go back to the basements of weak convergence on Banach spaces : the Prokhorov theorem. In the sequel we will consequently prove that the sequence of measures is tight, i.e flatly concentrated and bounded (see de Acosta (1970))

3 Main results

3.1 The i.i.d. case

In all the following T_1, T_2, \dots, T_n is a sequence of i.i.d. centered nuclear operators from H to H . We set $S_n = \frac{1}{\sqrt{n}} \sum_{p=1}^n T_i$. We need the following assumption :

\mathcal{H} : There exists a basis of H , say $(e_p)_p$ for which

$$\sum_p \sqrt{\mathbb{E} |T_1(e_p)|^2} < +\infty \quad (1)$$

Proposition 1 *When \mathcal{H} holds, the sequence S_n is flatly concentrated and bounded in probability.*

Proof. In order to prove the first point it suffices to prove that, for all $\varepsilon > 0$,

$$\lim_{m \rightarrow +\infty} \sup_n \mathbb{P} (\|S_n - \Pi_m S_n \Pi_m\|_1 > \varepsilon) = 0$$

where Π_m denotes the projection on the vector space spanned by (e_1, \dots, e_m) , $e_i \in H$. Now we refer to Lemma 32 p.1116 of Dunford Schwartz, which provides :

$$\|T\|_1 \leq \sum_p |T_1(e_p)|. \quad (2)$$

The previous inequality together with Markov and Jensen's inequalities imply :

$$\mathbb{P} (\|S_n - \Pi_m S_n \Pi_m\|_1 > \varepsilon) \leq \frac{1}{\varepsilon} \left(\sum_{p \leq m} \sqrt{\mathbb{E} |(I - \Pi_m) T_1(e_p)|^2} + \sum_{p \geq m+1} \sqrt{\mathbb{E} |T_1(e_p)|^2} \right)$$

The second term clearly decays and Lebesgue's dominated convergence implies that the first term also converges to zero since for fixed p each real sequence $\mathbb{E} |(I - \Pi_m) T_1(e_p)|^2$ tends to zero as m tends to infinity.

The proof of the boundedness in probability is really similar. For all strictly positive M ,

$$\mathbb{P} (\|S_n\|_1 > M) \leq \frac{1}{M} \sum_p \sqrt{\mathbb{E} |T_1(e_p)|^2}$$

and the term on the right tends to zero when M goes to infinity. ■

Theorem 2 *When \mathcal{H} holds, the sequence S_n satisfies the central limit theorem in \mathcal{C}_1 .*

Proof. The Prokhorov Theorem implies that the sequence S_n is weakly relatively compact by Proposition 1. It suffices now to prove that all the limits of the finite dimensional distributions are gaussian. Lets us take T in the dual space $\mathcal{L} = \mathcal{C}_1^*$. The central limit theorem on the line implies that $\langle T, S_n \rangle = \text{tr} T^* S_n$ converges weakly for all T to a gaussian centered random variable with variance $\mathbb{E} (\text{tr} T^* X_1)^2 = \mathbb{E} \langle T, X_1 \rangle^2$. Finally S_n satisfies the central limit theorem in \mathcal{C}_1 . The covariance operator of the gaussian limiting distribution is the covariance operator of X_1 . ■

Remark 3 *The assumption \mathcal{H} implies the finiteness of $\mathbb{E} \|T_1\|_1$ and of $\mathbb{E} \|T_1\|_2^2$. Note that \mathcal{H} is also really close to the condition that insures that a random element with values in the sequence space l_1 is pregaussian.*

3.2 The m-dependent case

The result of this paragraph may be viewed as a by-product of the preceding. It seems that anyway, the development of functional techniques in statistics, especially in the analysis of time dependent processes, makes it necessary to go beyond the simple i.i.d. setting. The technique of proof is quite similar to what is done in the finite dimensional or even the scalar case.

Let T and S denote two nuclear operators then $T \tilde{\otimes} S$ stands for the linear operator from \mathcal{L} to \mathcal{C}_1 defined by

$$T \tilde{\otimes} S(U) = \ll T, U \gg S$$

Theorem 4 Let X_t be a strictly stationary m -dependent sequence of centered random trace class operators. For all $p = 1, \dots, m$ set

$$\Gamma(p) = \mathbb{E} [X_1 \tilde{\otimes} X_p] \quad \text{and} \quad \Gamma^*(p) = \mathbb{E} [X_p \tilde{\otimes} X_1].$$

Suppose that X_1 satisfies \mathcal{H} , then

$$\frac{1}{\sqrt{n}} (X_1 + \dots + X_n) \xrightarrow[n \rightarrow +\infty]{w} G_m$$

where G_m is a centered Gaussian element in \mathcal{C}_1 with covariance operator

$$\Theta_m = \Gamma(0) + \sum_{j=1}^m (\Gamma(j) + \Gamma^*(j))$$

Proof :

The scheme of the proof is quite classical. In spite of the unusual framework we will go quickly through it. In order to alleviate the computations we will also prove Theorem 4 in the case $m = 1$ (consequently G_m becomes G_1). Generalization is straightforward.

The latent idea consists in using Theorem 4.2 p.25 in Billingsley (1968) about weak convergence for arrays of random variables in metric spaces. Take an integer $k > 2$ and let

$$\begin{aligned} Y_{nk} &= n^{-1/2} (X_1 + \dots + X_{k-1}) + (X_{k+1} + \dots + X_{2k-1}) \\ &\quad + \dots + (X_{(r-1)k+1} + \dots + X_{rk-1}) \end{aligned}$$

with $r = [n/k]$.

Note that $n^{1/2} Y_{nk}$ is a sum of r i.i.d. random variables each with mean zero and the covariance operator of $(X_1 + \dots + X_{k-1})$ that is indeed :

$$(k-1)\Gamma(0) + (k-2)(\Gamma(1) + \Gamma^*(1))$$

The first elementary steps consists in proving that, for fixed k , \mathcal{H} is fulfilled for the sum $X_1 + \dots + X_{k-1}$. But, stationarity of the sequence X_n yields

$$\mathbb{E} |(X_1 + \dots + X_{k-1})(e_p)|^2 \leq (k-1) \mathbb{E} |X_1(e_p)|^2.$$

Since X_1 satisfies \mathcal{H} , the central limit theorem in the space of nuclear operators gives :

$$(n/r)^{1/2} Y_{nk} \xrightarrow{w} G_{k,1}$$

where $G_{k,1}$ is a centered Gaussian element in \mathcal{C}_1 with the same covariance operator as above. This may be rewritten $Y_{nk} \xrightarrow{w} k^{-1/2} G_{k,1}$ as n goes to infinity. Now we have to study the weak convergence, as k tends to infinity, of the gaussian random operator $k^{-1/2} G_{k,1}$. Since we are just interested in its distribution, this random variable may be written the following way :

$$k^{-1/2} G_{k,1} = \frac{k-2}{k} (G_1 + N_k).$$

where N_k is a centered gaussian element, independent from G_1 and with covariance operator $\frac{1}{k} \Gamma(0)$. The conclusion follows $k^{-1/2} G_{k,1} \xrightarrow{w} G_1$.

We are now in position to apply Theorem 4.2 p.25 in Billingsley (1968) and if we prove that

$$\lim_{k \rightarrow +\infty} \limsup_n P \left(\left\| \sqrt{n} \bar{X}_n - Y_{nk} \right\|_1 > \varepsilon \right) = 0$$

the conclusion of Theorem 4 will follow. Computations are similar to those leading to the tightness in the i.i.d. case.

$$\begin{aligned} & P \left(\left\| \sqrt{n} \bar{X}_n - Y_{nk} \right\|_1 > \varepsilon \right) \\ & \leq P \left[n^{-1/2} \left\| \sum_{j=1}^r X_{jk} \right\|_1 > \varepsilon/2 \right] + P \left[n^{-1/2} \|X_1 + \dots + X_{n-rk}\|_1 > \varepsilon/2 \right] \\ & \leq \sqrt{\frac{r}{n}} \sum_{p=1}^{+\infty} \sqrt{\mathbb{E} |X_1(e_p)|^2} + \sum_{p=1}^{+\infty} \sqrt{\frac{k}{n} \mathbb{E} |X_1(e_p)|^2 + 2 \frac{k-1}{n} \mathbb{E} |X_1(e_p)|^2} \end{aligned}$$

We used the stationarity of the process and the fact that $n - rk \leq k$. Since $r/n \leq 1/k$, the first term (that does not depend on n any more) will tend to zero as k tends to infinity and the second term as n does. ■

4 Application to empirical covariance operators

Obviously Theorem 2 may be applied to obtain a central limit theorem in trace class norm for the empirical covariance operators of i.i.d. Hilbertian random elements. The covariance operator of a random variable with strong (i.e. Bochner) second order moment is a nuclear operator. The \mathcal{C}_1 -norm is consequently the thinnest and the more natural of all standard topologies to study these special operators.

4.1 Covariance operators

Let X_1, \dots, X_n be i.i.d. centered random variables with values in H . The sequence under concern is now

$$S_n = \frac{1}{\sqrt{n}} \sum_{k=1}^n (X_k \otimes X_k - \Gamma)$$

where $\Gamma = \mathbb{E} (X_1 \otimes X_1)$.

Proposition 5 *Let $\lambda_p = \mathbb{E} \langle X_1, e_p \rangle^2$ be the sequence of the eigenvalues of Γ , associated to the eigenvector e_p . Suppose that*

$$\sum_p \sqrt{\lambda_p} < +\infty \text{ and } \mathbb{E} \langle X_1, e_p \rangle^4 = O(\lambda_p^2), \quad (3)$$

then S_n converges weakly to a centered gaussian trace class random operator G with the same covariance operator as $X_1 \otimes X_1 - \Gamma$.

Proof. First note that the condition $\mathbb{E} \langle X_1, e_p \rangle^4 = O(\lambda_p^2)$ implies that $\mathbb{E} |X_1|_H^4 < +\infty$. We just have to prove that \mathcal{H} holds, namely

$$\sum_p \sqrt{\mathbb{E} |(X_1 \otimes X_1 - \Gamma)(e_p)|^2} < +\infty \quad (4)$$

Let us take for (e_p) an eigenbasis for Γ . We have :

$$\begin{aligned} \mathbb{E} |(X_1 \otimes X_1 - \Gamma)(e_p)|^2 &\leq \mathbb{E} \langle X_1, e_p \rangle^2 |X_1|^2 \\ &\leq \sqrt{\mathbb{E} \langle X_1, e_p \rangle^4 \mathbb{E} |X_1|^4} = O(\lambda_p) \end{aligned}$$

hence the desired result.

Remark 6 *If the random variable X_1 is gaussian the second assumption in (3) holds. Indeed we have $\langle X_1, e_p \rangle = \sqrt{\lambda_p} \eta_p$ where η_p are i.i.d gaussian centered and $\mathbb{E} \langle X_1, e_p \rangle^4 = 3\lambda_p^2$. Also note that assumption \mathcal{H} is easily checked since it suffices to compute the projections of the process on a given basis of H (spline, sinusoidal, etc.)*

Remark 7 *Theorem 4 could be applied to the covariance operator of the autoregressive process in Bosq (2000) :*

$$X_t = \rho(X_{t-1}) + \varepsilon_t$$

where the X_t 's and the ε_t 's are H -valued and ρ is linear from H to H . The same technique reveals fruitful in Mas (2002). Limiting distributions of the covariance operators (for the Hilbert-Schmidt norm) were obtained for the more general linear model :

$$X_t = \sum_k a_k(\varepsilon_{t-k})$$

where the a_k 's are linear operators.

4.2 Cross-covariance operators

Suppose now $X_1, \dots, X_n, Y_1, \dots, Y_n$ are all independent centered random variables with values in H . Now we set

$$\tilde{S}_n = \frac{1}{\sqrt{n}} \sum_{k=1}^n X_k \otimes Y_k$$

The random trace class operator \tilde{S}_n is the cross-covariance operator of the couple (X, Y) and appears in problems connected to functional statistics (see Cardot, Ferraty, Sarda (1999), Bosq (2000)). The following is an easy consequence of Theorem 2.

Proposition 8 *Once more let λ_p be the sequence of the eigenvalues of $\Gamma = \mathbb{E}(X_1 \otimes X_1)$, associated to the eigenvector e_p . Assume that :*

$$\sum_p \sqrt{\lambda_p} < +\infty \text{ and } \mathbb{E} |Y_1|^2 < +\infty$$

then \tilde{S}_n converges weakly to a centered gaussian trace class random operator \tilde{N} with the same covariance operator as $X_1 \otimes Y_1$.

References

- [1] de Acosta, A. (1970) : Existence and convergence of probability measures in Banach spaces, *Trans. Amer. Math. Soc.* **152**, 273-298.

- [2] Araujo A., Gine E. (1980) : *The central limit theorem for real and Banach valued random variables* (Wiley Series in Probability and Mathematical Statistics).
- [3] Billingsley, P. (1968) : *Convergence of probability measures* (Wiley Series in Probability and Mathematical Statistics).
- [4] Bosq, D. (2000) : *Linear processes in function spaces*. Lectures notes in statistics (Springer Verlag).
- [5] Brockwell, P.J., Davis, R.A. (1991) : *Times Series : Theory and methods* (Springer Series in Statistics. 2nd ed. Springer -Verlag).
- [6] Cardot, H., Ferraty, F., Sarda, P. (1999) : Functional linear model, *Stat. Proba. Letters*, **45**, 11-22.
- [7] Dauxois, J., Pousse, A., Romain, Y. (1982) : Asymptotic theory for the principal component analysis of a vector random function : some applications to statistical inference, *J. Multivar. Anal.* **12** 136-154.
- [8] Dunford, N., Schwartz, J.T. (1988) : *Linear Operators Vol I,II,III* (Wiley Classics Library).
- [9] Mas, A. (2002) : Weak convergence for the covariance operators of a Hilbertian linear process, *Stoch. Process. Appl.* **99**,117-135.
- [10] Rice, J.A., Silverman, B.W. (1991) : Estimating the mean and covariance structure nonparametrically when the data are curves. *J.R.S.S. Ser. B* **53**, 233-243.
- [11] Tomczak-Jaegerman, N. (1974) : The moduli of smoothness and convexity of trace class \mathcal{S}_p , *Studia Math.* **50**, 163-182.
- [12] Vakhania, N.N., Tarieladze, V.I., Chobanyan, S.M. (1987) : *Probability Distributions on Banach Spaces*. Mathematics and its Applications. (D. Reidel Publishing).