Journal of the Korean
Statistical Society
Vol. 22, No. 1, 1993

Random Elements in $L^1(R)$ and Kernel Density Estimators

Sungho Lee[1] and Robert Lee Taylor[2]

ABSTRACT

Random elements in $L^1(R)$ and some properties of $L^1(R)$ space are investigated with application to kernel density estimators. A weak law of large numbers for compact uniformly integrable random elements is introduced for further application.

1. INTRODUCTION

The consideration of stochastic processes, as function-valued random variables motivated the study of random elements (random variables with values in normed linear spaces) by Doob (1947), Mourier (1953), Prohorov (1956) and others. The laws of large numbers for random elements have been obtained, and a summary of many of these results was presented by Taylor (1978). In the space of real valued continuous functions which converge to zero at $\pm\infty$, Co(R), with the supremum norm, Taylor and Hu (1987) developed complete

[1] Department of Statistics, Taegu University, Kyungsan, 713-714, Korea
[2] Department of Statistics, University of Georgia, Athens, Ga30602, U.S.A.

convergence laws of large numbers which have direct applications in establishing the uniform strong consistency of the kernel density estimators. The kernel density estimator (cf : Rosenblatt (1956) and Parzen (1962)),

$$\hat{f}_n(t) = \frac{1}{nh_n} \sum_{k=1}^{n} K(\frac{t - X_k}{h_n}), \tag{1.1}$$

is a random function taking values in a function space which is determined by the choice of K. In (1.1) $\{X_k\}$ is a random sample from a distribution which has (unknown) pdf f in L^1 , $\{h_n\}$ is a sequence of constants (possibly random variables) tending to zero, and K is a weighting (or kernel) function which is selected by the experimenter. In particular, it is desired to show that

$$\left\|\hat{f}_n - f\right\| = \int \left|\hat{f}_n(t) - f(t)\right| dt \to 0 \qquad a.s.$$

when $h_n \to 0$ and K is a probability density function.

The L^1 consistency of kernel density estimators has been strongly supported by several authors, most notably by Devroye and Gyorfi (1985). Devroye (1983) showed that all types of L^1 consistency for the constant bandwidth kernel density estimators are equivalent in the general setting. This paper will give illustrative characterizations of random elements in $L^1(R)$ and develop the basic analytic and probabilistic properties of $L^1(R)$. Finally, a weak law of large numbers in $L^1(R)$ is introduced.

2. Random Elements in $L^1(R)$

Let $(\Omega,\mathcal{A},\mu)$ be a measure space. The space $L^1 = L^1(\Omega,\mathcal{A},\mu, R)$ is the set of all μ-equivalence classes of $\mathcal{A}$-measurable function x: $\Omega \to R$ such that $\int |x|\, d\mu < \infty$. The norm for L^1 is defined by $\|x\| = \int_\Omega |x|\, d\mu$. For the space $L^1(R)$ of this paper, $\Omega = R$, $\mathcal{A}$=Borel field and μ is the Lebesgue measure on R. Let $(\Omega,\mathcal{A},P)$ be a probability space and E be a separable Banach space with norm denoted by $\|\ \ \|$. A random element X in E is a function from Ω into E which is $\mathcal{A}$-measurable with respect to the Borel subsets of E.

Definition 2.1. A function X is said to be a random element in $L^1(R)$ if $[X \in B] = \{\omega : X(\omega) \in B\} \in \mathcal{A}, \quad \forall B \in B\ (L^1(R))$.

A random element is a generalization of a random variable. Since the σ-field generated by all intervals of the form $(-\infty, b]$ is the class of Borel subsets of R, X is a random element in R if and only if X is a random variable.

Lemma 2.1. (cf : Taylor (1978)) If E is a separable Banach space, then a function $X : \Omega \to E$ is a random element if and only if $f(X)$ is a random variable for each $f \in E^*$, the dual space of E.

Since $f(X_1 + X_2) = f(X_1) + f(X_2)$ is a random variable whenever X_1 and X_2 are random elements in a separable Banach spaces E and $f \in E^*$, the sum of two random elements in a separable Banach space is a random element. However, the sum of two random elements in a nonseparable Banach space need not be a random element (cf : Taylor (1978)). The following characterization of random elements in $L^1(R)$ is illustrative and useful in obtaining laws of larger numbers.

Lemma 2.2 (cf : Taylor and Lee (1990))

(a) Let X be a function from $R \times \Omega$ into R such that

(i) $\forall t \in R$, $X(t, \cdot)$ is a random variable

(ii) $\forall \omega \in \Omega$, $X(\cdot, \omega)$ is a Riemann integrable function.

If for each $\omega \in \Omega$, $X(\cdot, \omega)$ is identified with $\tilde{X}(\cdot, \omega)$, the equivalence class of $X(\cdot, \omega)$, then X is a random element in $L^1(R)$.

(b) Let $\tilde{X}$ be a random element in $L^1(R)$. Then there exists a function $X : R \times \Omega \to R$ such that

(i) $\forall \omega \in R$, $X(\cdot, \omega)$ is a Lebesgue integrable function with $X(\cdot, \omega) \in \tilde{X}(\cdot, \omega)$,

(ii) $\forall t \in R$, $X(t, \cdot)$ is an extended random variable.

Lemma 2.2 (b) asserts the existence of a stochastic process $X(t, \cdot)$ taking values in $[-\infty, \infty]$ such that for each ω $X(\cdot, \omega)$ is Lebesgue integrable and the equivalence class containing $X(\cdot, \omega)$ coincides with a given equivalence class $\tilde{X}$. Thus we can say that Lemma 2.2 (b) asserts that for any $L^1(R)$-valued random element $\tilde{X}$ there exists a distinguished representative of $\tilde{X}(\cdot, \omega)$, say $g_{\tilde{X}}(t, \omega)$, such that $g_{\tilde{X}}(t, \cdot)$ is a stochastic process. $g_{\tilde{X}}(t, \omega)$ can be expressed as

$$g_{\tilde{X}}(t_0, \omega) = \overline{\lim} \frac{1}{2s_n} \int_{t_0 - s_n}^{t_0 + s_n} \tilde{X}(t, \omega)\, dt, \quad \forall t_0 \in R$$

where $\{s_n\}$ is a decreasing sequence of positive real numbers converging to zero.

Remark From now on, we'll denote a random element X in $L^1(R)$ as $\tilde{X}$ to distinguish that $X(\omega)$ is an equivalence class for each $\omega \in \Omega$.

Example 2.1. Let $\Omega = \{H, T\}$, $\mathcal{A} = \{\emptyset, \Omega\}$, and $P(\emptyset) = 0$, $P(\Omega) = 1$.

Define $X(t,\omega) = \mathbf{I}_{\{H\}}(\omega)\mathbf{I}_{[0,\frac{1}{2})}(t) + \mathbf{I}_{\{T\}}(\omega)\mathbf{I}_{[\frac{1}{2},1]}(t)$.

For $\omega = H$, identify $X(t,\omega)$ with the equivalence class $\mathbf{I}_{[0,\frac{1}{2})}(t)$.

For $\omega = T$, identify $X(t,\omega)$ with the equivalence class $\mathbf{I}_{[\frac{1}{2},1]}(t)$.

Hence, $\tilde{X} : \Omega \to L^1(R)$. But $\tilde{X}$ is not a random element in $L^1(R)$ since for $0 \le t < \frac{1}{2}$ $X(t)$ is not a random variable.

The expected value for a random element in a normed linear space is defined by the Pettis integral (cf : Taylor (1978)). That is, X has an expected value $EX \in E$ if $f(EX) = E(f(X))$ $\forall f \in E^*$. In a separable Banach space, the Pettis integral is equal to the Bochner integral when the Bochner integral exists. In particular a random element X in a separable Banach space E has a Bochner integral $EX \in E$ if and only if $E\|X\| < \infty$. The following lemma gives a characterization of expected values in $L^1(R)$.

Lemma 2.3 (cf : Taylor and Lee (1990)) Let $\tilde{X}$ be a random element in $L^1(R)$ such that $E\left\|\tilde{X}\right\| < \infty$. Then there exists a unique $EX \in L^1(R)$ such that

(i) $f(\widetilde{EX}) = E(f(\widetilde{X}))$ $\forall f \in L^1(R)^*$ and

(ii) $\widetilde{EX} = \widetilde{E[X(t,\cdot)]}$

Part (ii) of Lemma 2.3 indicates that expected values in $L^1(R)$, in certain situations, can be identified with the mean value function of a stochastic process.

3. Some Analytic and Probabilistic Properties of $L^1(R)$

In this section $L^1(R)$ space is generated from standard kernel functions and some probabilistic properties of $L^1(R)$ are considered. Our attention is restricted to the concept of type p and tightness properties relating to laws of large numbers for separable Banach spaces.

Definition 3.1. (cf : Taylor and Hu (1987)) A standard kernel function $K(t)$ is an even probability density function which satisfies

(i) $K(t)$ is strictly decreasing in $[0,1]$,

(ii) $K(t) = 0$ when $t \notin [-1,1]$, and

(iii) $|K(t) - K(s)| \le c \mid t-s \mid^{\alpha}$ for some $\alpha > 0$ and $c > 0$.

A kernel function with support $[a-b, a+b]$ is defined as $K_b^a(t) = K(\frac{t-a}{b})$. Let $F = \{K_b^a(t),\ a \in R,\ b > 0\}$ and let E be the space of all finite linear combination of elements in F, that is,

$$E = \{\sum_{i=1}^{n} c_i K_{b_i}^{a_i}(t) : a_i, b_i, c_i \in R \text{ with } b_i > 0\}.$$

Let $E_c \subset E$ be the subspace consisting of all $f \in E$ such that supp $(f) \subset C$.

Theorem 3.1. (cf : Taylor and Hu (1987)) If K is a standard kernel function, then $C_o(R)$ is the completion of E, relative to the metric defined by the supremum norm.

Lemma 3.1. For each compact subset D of R, $C_D \subset \hat{E}_{D'}^{\infty}$ where $D' \supset D$, $\hat{E}_{D'}^{\infty}$ denotes the completion of $E_{D'}$ in $L^{\infty}(R)$, and C_D is the set of continuous functions with support D.

Proof. Since for $f \in E_{D'}$, $\|f\|_{\infty} = \sup_x |f(x)|$, by the same procedure in the proof of Theorem 3.1 the result follows.

Theorem 3.2. If K is a standard kernel function, then $L^1(R)$ is the completion of E, relative to the metric defined by the L^1 norm.

Proof. By Lemma 3.1, for any compact set D there exists a compact set $D' \supset D$ and $C_D \subset \hat{E}_{D'}^{\infty}$. Thus, if $f \in C_D$, then there exists $f_n \in E_{D'}$ such that $\|f_n - f\|_{\infty} = o(1)$ as $n \to \infty$. Hence, $\int |f_n - f| = \int_{D'} |f_n - f| \leq \|f_n - f\|_{\infty} \int_{D'} dx = o(1)$ as $n \to \infty$ and $C_D \subset \hat{E}_{D'}^1 \subset \hat{E}^1$. This implies $(\widehat{\cup C_D})^1 \subset \hat{E}^1$, where the union is over all compact sets D. Notice that $\hat{C}_D^1 = L^1(D)$, the space of all L^1 functions with support D. Thus, we have $\cup \hat{C}_D^1 = \cup L^1(D) \subset (\widehat{\cup C_D})^1 \subset \hat{E}^1$ so that the completion of $(\cup L^1(D))$ in $L^1(R)$ is a subset of $\hat{E}^1$ and is obviously $L^1(R)$.

Lemma 3.2. Let K be a compact subset of $L^1(R)$. Then for each $\epsilon > 0$ there exists a constant m_k such that $\sup_{\tilde{x} \in K} \left\| \tilde{x} - \tilde{x}\mathbf{I}_{[|t| \leq m_k]} \right\| < \epsilon$.

Proof. Since K is compact, for given $\epsilon > 0$ there exist $\{\tilde{x}_1, \cdots, \tilde{x}_s\} \subset K$ such that $\bigcup_{i=1}^{s} \{\tilde{y} : \|\tilde{y} - \tilde{x}_i\| < \frac{\epsilon}{3}\} \supset K$. Since $\tilde{x}_i \in L^1(R)$ implies $\left\| \tilde{x}_i - \tilde{x}_i \mathbf{I}_{[|t| \leq n]} \right\| \to 0$ as $n \to \infty$, we can choose m_k such that $\sup_{1 \leq i \leq s} \left\| \tilde{x}_i - \tilde{x}_i \mathbf{I}_{[|t| \leq m_k]} \right\| < \frac{\epsilon}{3}$. Thus, $\forall \tilde{x} \in K$ there exists $\tilde{x}_i$ such that $\|\tilde{x} - \tilde{x}_i\| < \frac{\epsilon}{3}$. Hence,

$$\left\| \tilde{x} - \tilde{x}\mathbf{I}_{[|t| \leq m_k]} \right\| \leq \|\tilde{x} - \tilde{x}_i\| + \left\| \tilde{x}_i - \tilde{x}_i \mathbf{I}_{[|t| \leq m_k]} \right\| + \left\| \tilde{x}_i \mathbf{I}_{[|t| \leq m_k]} - \tilde{x}\mathbf{I}_{[|t| \leq m_k]} \right\| < \epsilon.$$

Definition 3.2. A separable Banach space E is said to be of type p, $1 \le p \le 2$, if for any independent random elements $X_1, \cdots, X_n$ in E with $EX_i = 0$ and $E\|X_i\|^p < \infty$ for each i,

$$E\left(\left\|\sum_{i=1}^{n} X_i\right\|^p\right) \le C \sum_{i=1}^{n} E\|X_i\|^p .$$

By the triangular inequality, every separable Banach space is at least type 1. However, $L^1(R)$ is only of type 1 as the following theorem shows.

Theorem 3.3. $L^1(R)$ is only of type 1.

Proof. Let $\tilde{x}_n, n \ge 1$, be the equivalence class containing $2^n \mathbf{I}_{[1-2^{1-n}, 1-2^{-n})}(t)$ and $\{Y_n\}$ be i.i.d random variables such that $P[Y_n = 1] = P[Y_n = -1] = \frac{1}{2}$. For each n, let $\tilde{X}_n = \tilde{x}_n Y_n$. Then $\{\tilde{X}_n\}$ is a sequence of independent random elements in $L^1(R)$. Since $\left\|\tilde{X}_n\right\| = 1$ and $\widetilde{EX}_n = \tilde{0}$ for each n, $E\left\|\sum_{k=1}^{n} \tilde{X}_k\right\|^p = n^p$ and $C\sum_{k=1}^{n} E\left\|\tilde{X}_k\right\|^p = n \cdot C$ thus, $n^p \le n \cdot C$, for each n, implies $p = 1$.

Before closing this section we will investigate compact uniform integrability of random elements in $L^1(R)$ which is motivated from kernel density estimation problems. Let $\{X_n\}$ be a sequence of random elements defined on a probability space $(\Omega, \mathcal{A}, P)$ taking values in a separable normed linear space E. Then $\{X_n\}$ is said to be compact uniformly integrable if for every $\epsilon > 0$ there exists a compact subset K of E such that $\sup_n E[\left\|X_n \mathbf{I}_{[X \notin K]}\right\|] < \epsilon$. The following two lemmas are useful for later results.

Lemma 3.3. (cf : Wang and Rao (1987)) Let $\{X_n\}$ be a sequence of random elements in a separable Banach space E. Then (i) and (ii) are equivalent:

(i) $\{X_n\}$ is uniformly tight and $\{\|X_n\|\}$ is uniformly absolutely continuous.
(ii) $\{X_n\}$ is compact uniformly integrable.

Furthermore, if $\{X_n\}$ is compact uniformly integrable, then $\{X_n - EX_n\}$ is compact uniformly integrable.

Theorem 3.4. Let $K(\cdot)$ be a Riemann integrable function with $\int |K(t)|\, dt = 1$ and let $\{X_n\}$ be a sequence of i.i.d. random variables. Suppose that $H_n = H_n(X_1, \cdots, X_n)$, $n = 1, 2, \cdots$, are uniformly bounded, positive random variables such that $H_n \to 0$ in probability. If for each $\omega \in \Omega$, $K(\frac{t-X_k}{H_n})$ is identified with $\tilde{K}(\frac{t-X_k}{H_n})$, then

$$\{\tilde{X}_{n_k}(t,\omega) = \tilde{K}(\frac{t-X_k}{H_n}) - \widetilde{EK}(\frac{t-X_k}{H_n}) \ : 1 \le k \le n,\ n \ge 1\}$$

is a triangular array of random elements in $L^1(R)$ which are compact uniformly integrable.

Proof. By Lemma 2.2, $\{\tilde{X}_{n_k} : 1 \le k \le n,\ n \ge 1\}$ is a triangular array of random elements in $L^1(R)$ and $\left\|\tilde{X}_{n_k}\right\| \le H_n + E(H_n)$. Let $\epsilon > 0$ be given. Since $\tilde{X}_{n_k}$ is tight and $\tilde{X}_{n_1}, \cdots, \tilde{X}_{n_n}$ are identically distributed for each n, there exists a compact subset $C_n{}'$ of $L^1(R)$ for each n such that

$$\inf_k P[\tilde{X}_{n_k} \in C_n{}'] = P[\tilde{X}_{n_1} \in C_n{}'] > 1 - \frac{\epsilon}{2}.$$

Since $H_n \to 0$ in probability, there exists $\{\delta_n\}$ such that $\delta_n \to 0$ and $P[H_n + E(H_n) > \delta_n] < \frac{\epsilon}{2}$. Let $C_n = C_n{}' \cap \{\tilde{x} : \|\tilde{x}\| \le \delta_n\}$. Then C_n is compact and $P[\tilde{X}_{n_k} \notin C_n] \le P[\tilde{X}_{n_k} \notin C_n{}'] + P[\left\|\tilde{X}_{n_k}\right\| > \delta_n] < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$. Set $A = \bigcup_{n=1}^{\infty} C_n$. For each $\tilde{x} \in C_n$, $\|\tilde{x}\| \le \delta_n \to 0$ as $n \to \infty$. Hence, for each $\delta > 0$ there exits N such that

$$\{\tilde{x} : \|\tilde{x}\| < \delta\} \supset \bigcup_{n=N+1}^{\infty} C_n.$$

Since $\bigcup_{n=1}^{N} C_n$ is compact, there exits a δ-net for A. Thus, $\bar{A}$ is compact and $\inf_{n,k} P[\tilde{X}_{n_k} \in \bar{A}] = \inf_n P[\tilde{X}_{n_1} \in \bar{A}] > 1 - \epsilon$. Since for each $B \in \mathcal{A}$,

$$\begin{aligned}\sup_{n,k} E[\left\|\tilde{X}_{n_k}\right\| \mathbf{I}_B] &= \sup_{n,k} \int_B \|X_{n_k}\|\, dP \\ &\le \sup_n \int_B (H_n + EH_n) dP \\ &\le \sup_n \int_B H_n dP + P(B) \sup_n EH_n \to 0\end{aligned}$$

as $P(B) \to 0$, $\{\left\|\tilde{X}_{n_k}\right\|\}$ is uniformly absolutely continuous, and hence the result follows by Lemma 3.3.

Finally, convergence in mean will be shown for compact uniformly integrable random elements in $L^1(R)$. Recall that for a random element $\tilde{X}$ in $L^1(R)$ $X \in \tilde{X}$ is defined in Lemma 2.2. The following theorem is from Taylor and Lee (1990).

Theorem 3.5. (cf: Taylor and Lee (1990)) Let $\{\tilde{X}_{n_k} : 1 \leq k \leq n,\ n \geq 1\}$ be a triangular array of random elements in $L^1(R)$ which are compact uniformly integrable with $\widetilde{EX}_{n_k} = \tilde{0}$ for each n and k such that $|X_{n_k}(t)| \leq M$ for each t, k, and n. Let $\{a_{n_k}\}$ be an array of real numbers such that $\sum_{k=1}^{n} |a_{n_k}| \leq \Gamma < \infty$ for each n. If $\sum_{k=1}^{n} a_{n_k} X_{n_k}(t) \to 0$ in probability for each $t \in R$, then $E\left\|\sum_{k=1}^{n} a_{n_k} X_{n_k}\right\| \to 0$.

Corollary 3.1. Let $\{\tilde{X}_{n_k} : 1 \leq k \leq n,\ n \geq 1\}$ be a triangular array of random elements in $L^1(R)$ which are compact uniformly integrable with $\widetilde{EX}_{n_k} = \tilde{0}$ for each n and k such that $|X_{n_k}(t)| \leq M$ for each t, k, and n. Let $\{a_{n_k}\}$ be an array of real numbers such that $\sum_{k=1}^{n} |a_{n_k}| \leq \Gamma < \infty$ for each n.

If $\sum_{k=1}^{n} a_{n_k} X_{n_k}(t) \to 0$ in probability for each $t \in R$, then $\left\|\sum_{k=1}^{n} a_{n_k} X_{n_k}\right\| \to 0$ in probability.

Proof. Since $E\left\|\sum_{k=1}^{n} a_{n_k} \tilde{X}_{n_k}\right\| \to 0$ implies $\left\|\sum_{k=1}^{n} a_{n_k} X_{n_k}\right\| \to 0$ in probability, the result follows.

REFERENCES

(1) Devroye, L. (1983). The equivalence of weak, strong and complete convergence in L_1 for kernel density estimates. *Annals of Statistics*, 11, 896-404.

(2) Devroye, L. and Gyorfi, L. (1985). *Nonparametric Density Estimation. The L_1 View*, John Wiley, New York.

(3) Doob, J. L. (1947). Probability in function space. *Bulletin of the American Mathematical Society*, 53, 15-30.

(4) Mourier, E. (1953). Elements aleatories dan un espace de Banach. *Annales de I'Institut Henri Poincare*, 13, 159-244.

(5) Parzen, E. (1962). On estimation of a probability density and mode. *Annals of Mathematical Statistics*, 33, 1065-1076.

(6) Prohorov, Y. V. (1956). Convergence of random processes and limit theorems in probability theory. *Theory of Probability and its Applications*, 1, 157-214.

(7) Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27, 832-837.

(8) Taylor, R. L. (1978). *Stochastic Convergence of Weighted Sums of Random Elements in Linear Spaces*, Lecture Notes in Mathematics, V. 672, Springer-Verlag, New York.

(9) Taylor, R. L. and Hu, T. C. (1987). Consistency of kernel density estimators and laws of large numbers in Co(R). In : *Mathematical Statistics and Probability Theory*, Volume A of the Proceedings of the 6th Pannonian Symposium on Mathematical Statistics, D. Reidel publ. Comp., Dordrecht, 253-266.

(10) Taylor, R. L. and Lee, S. (1990). Laws of large numbers in $L^1(R)$, University of Georgia Statistics Technical Report ♯133.

(11) Wang, X. C. and Rao, M. B. (1987). Some results on the convergence of weighted sums of random elements in separable Banach spaces. *Studia Mathematica*, *T.L*XXXVi, 131-153.