Supplemental materials for this article are available through the TAS web page at http://www.amstat.org/publications.

Teacher's Corner

Understanding Convergence Concepts: A Visual-Minded and Graphical Simulation-Based Approach

Pierre LAFAYE DE MICHEAUX and Benoit LIQUET

This article describes the difficult concepts of convergence in probability, convergence almost surely, convergence in law, and convergence in rth mean using a visual-minded and graphical simulation-based approach. For this purpose, each probability of events is approximated by a frequency. An R package that reproduces all of the experiments cited in this article is available in CRAN. See the online Supplement for details.

KEY WORDS: Convergence almost surely; Convergence in law; Convergence in probability; Convergence in rth mean; Dynamic graphics; Monte Carlo simulation; R language; Visualization.

1. INTRODUCTION

Most departments of statistics teach at least one course on the difficult concepts of convergence in probability (P), almost sure convergence ($a.s.$), convergence in law (L), and convergence in rth mean (r) at the graduate level (see Sethuraman 1995). Indeed, as pointed out by Bryce et al. (2001), "statistical theory is an important part of the curriculum, and is particularly important for students headed for graduate school." Such knowledge is prescribed by learned statistics societies (e.g., the Accreditation of Statisticians by the Statistical Society of Canada and Curriculum Guidelines for Undergraduate Programs in Statistical Science by the American Statistical Association). The main textbooks (e.g., Chung 1974; Billingsley 1986; Ferguson 1996; Lehmann 2001; Serfling 2002) devote about 15 pages to defining these convergence concepts and their interrelations. Very often, these concepts are provided as definitions, and students are exposed only to some basic properties and to the universal implications displayed in Figure 1.

The aim of this article is to clarify these convergence concepts for master's students in mathematics and statistics, and

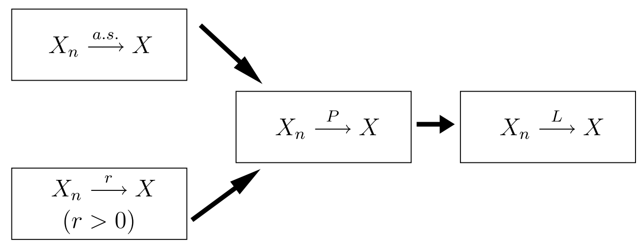

Figure 1. Universally valid implications of the four classical modes of convergence. (See Ferguson 1996 for proofs.)

also to provide software useful for learning these concepts. Each convergence notion provides an essential foundation for further work. For example, convergence in law is used to obtain asymptotic confidence intervals and hypothesis tests using the central limit theorem; convergence in probability is used to obtain the limiting distribution of the Z test replacing an unknown variance with its estimate (through Slutsky's theorem); quadratic mean convergence is used to obtain a mean squared error for point estimators; and almost sure convergence is a natural extension of deterministic uniform convergence. To explain these modes of convergence, we could follow Bryce et al.'s (2001) advice: "A modern statistical theory course might, for example, include more work on computer intensive methods." Dunn (1999) and Marasinghe et al. (1996) proposed interactive tools for understanding convergence in law. Mills (2002) proposed a review of statistical teaching based on simulation methods, and Chance and Rossman (2006) have written a book on this subject. In Section 2 we first define the convergence concepts and show how to visualize them and help form relevant mental images. We then use a graphical simulation-based approach to illustrate this perspective and to investigate some modes of convergence in practical situations. In Section 3 we point out subtler distinctions among the various modes through examples. We illustrate these differences through exercises and solutions that emphasize our visualization approach in an online Appendix (*http://www.biostatisticien.eu/ConvergenceConcepts*). We propose an R package (R Development Core Team 2008) package called `ConvergenceConcepts`. This package's interactive part provides an interesting pedagogic tool to facilitate visualization of the convergence concepts. The package also created all of the figures presented here and can be used to investigate

Pierre Lafaye de Micheaux is Associate Professor, Départment Mathématiques et de Statistique, Universitè de Montréal, Canada (E-mail: *lafaye@dms.umontreal.ca*). Benoit Liquet is Associate Professor, INSERM U897, ISPED, University of Bordeaux 2, Bordeaux, France (E-mail: *Benoit.Liquet@isped.u-bordeaux2.fr*). The authors thank R. Drouilhet, M. Lejeune, and the referees, the editor, and the associate editor for their many helpful comments.

 DOI: 10.1198/tas.2009.0032

the convergence of any random variable. This approach aims to help students develop intuition and logical thinking.

2. MODES OF CONVERGENCE

Probability theory is the body of knowledge that allows us to formally reason about any uncertain event A. A popular view of probability is the so-called "frequentist" approach (Fisher 1956): If an experiment is repeated M times "independently" under essentially identical conditions, and if event A occurs k times, then as M increases, the ratio k/M approaches a fixed limit, namely the probability $P(A)$ of A.

In our context we are interested mainly in the probability of events related to some random variables, namely $P[\omega \in \Omega; X_\omega \in E]$, where Ω is some arbitrary set. We use the following property:

$$P[\omega; X_\omega \in E] = \lim_{M\to\infty} \frac{\#\{j \in \{1,\ldots,M\}; x^j \in E\}}{M},$$

where x^j denotes the jth outcome of X independent of the others and $\#\{j \in \{1,\ldots,M\}; x^j \in E\} \equiv \#\{x^j \in E\}$ denotes the number of $j \in \{1,\ldots,M\}$ such that $x^j \in E$, for some set E.

In the sequel, we will study the convergence (in some sense, to be defined later) of sequences of random variables X_n to X.

We note that $(x_n^j - x^j)_{n\in\mathbb{N}} = (x_1^j - x^j, x_2^j - x^j, \ldots, x_n^j - x^j, \ldots)$, the jth sample path of $(X_n - X)_{n\in\mathbb{N}}$.

2.1 Convergence in Probability

We write $X_n \xrightarrow{P} X$ and say that the sequence $(X_n)_{n\in\mathbb{N}}$ converges in probability to X if

$$\forall\epsilon > 0, \qquad p_n = P[\omega; |X_{n,\omega} - X_\omega| > \epsilon] \underset{n\to\infty}{\longrightarrow} 0. \qquad (1)$$

The index ω can be seen as a labeling of each sample path. To understand this notion of convergence, we use the aforementioned frequentist approach to approximate the probability $p_n = P[\omega; |X_{n,\omega} - X_\omega| > \epsilon]$ by the frequency $\hat{p}_n = \frac{1}{M} \times \#\{|x_n^j - x^j| > \epsilon\}$.

Mind visualization approach. We can mentally visualize the M sample paths of the stochastic process $(X_n - X)_{n=1,\ldots,n_{max}}$. Each sample path is made up of a sequence of points indexed by the integers. For each successively increasing value of n, we can then evaluate the proportion $\hat{p}_n$ of the sample paths that are out of a horizontal band $[-\epsilon, +\epsilon]$. This band can be chosen to be arbitrarily narrow. The sample paths should be observed only at each fixed position n, for example, by mentally sliding a highlighting vertical bar along the n values axis. This is illustrated in Figure 2, which can be considered a static example of our dynamic mental images. The evolution of $\hat{p}_n$ toward 0 informs us about the convergence (or not) in probability of X_n toward X.

To better understand how Figure 2 describes the idea of convergence in probability, students are invited to manipulate the interactive version of it provided in our package, as demonstrated in Example 1.

Example 1. Figure 3 shows the convergence in probability $X_n = \overline{Y}_n = \frac{1}{n}\sum_{i=1}^n Y_i \xrightarrow{P} X = 0$, where the random variables Y_i are iid $N(0,1)$. We use $M = 500$ realizations, consider $\epsilon = 0.05$, and take $n_{max} = 2000$. Using our package, the user can move the vertical bar on the left side of Figure 3, and thus see the sample paths lying outside the horizontal band, as indicated by small horizontal (red) marks, and simultaneously observe their proportion $\hat{p}_n$ decreasing to 0 on right side of Figure 3, as indicated by a sliding (blue) circle.

Remark 1. Note that $X_n \xrightarrow{P} X \Leftrightarrow X_n - X \xrightarrow{P} 0$. Therefore, to study the convergence in probability of a random vari-

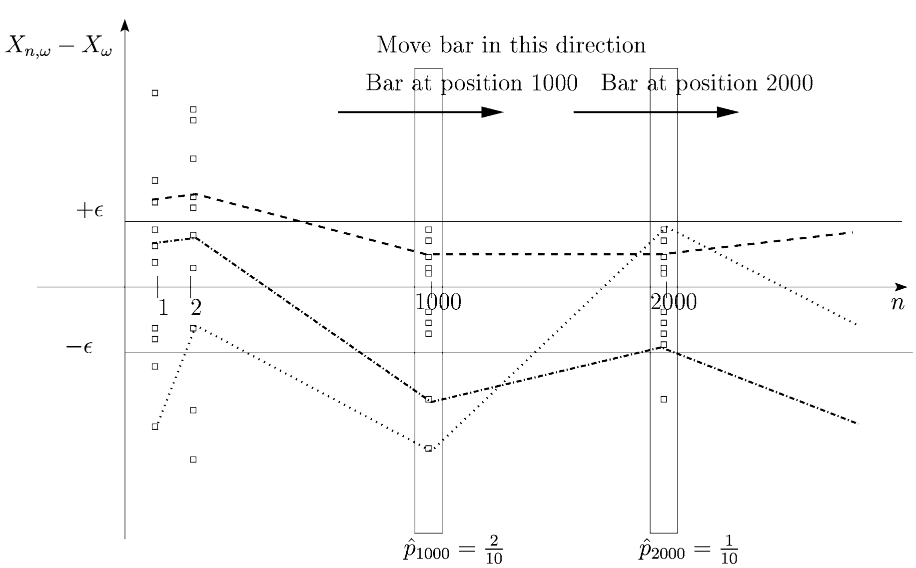

Figure 2. Seeing convergence in probability with $M = 10$ fictitious realizations. For $n = 1000$, $\hat{p}_n = 2/10$, because we can see two sample paths lying outside the band $[-\epsilon, +\epsilon]$ in the bar at position 1,000. For $n = 2000$, $\hat{p}_n = 1/10$, because we can see one sample path lying outside the band $[-\epsilon, +\epsilon]$ in the bar at position 2,000.

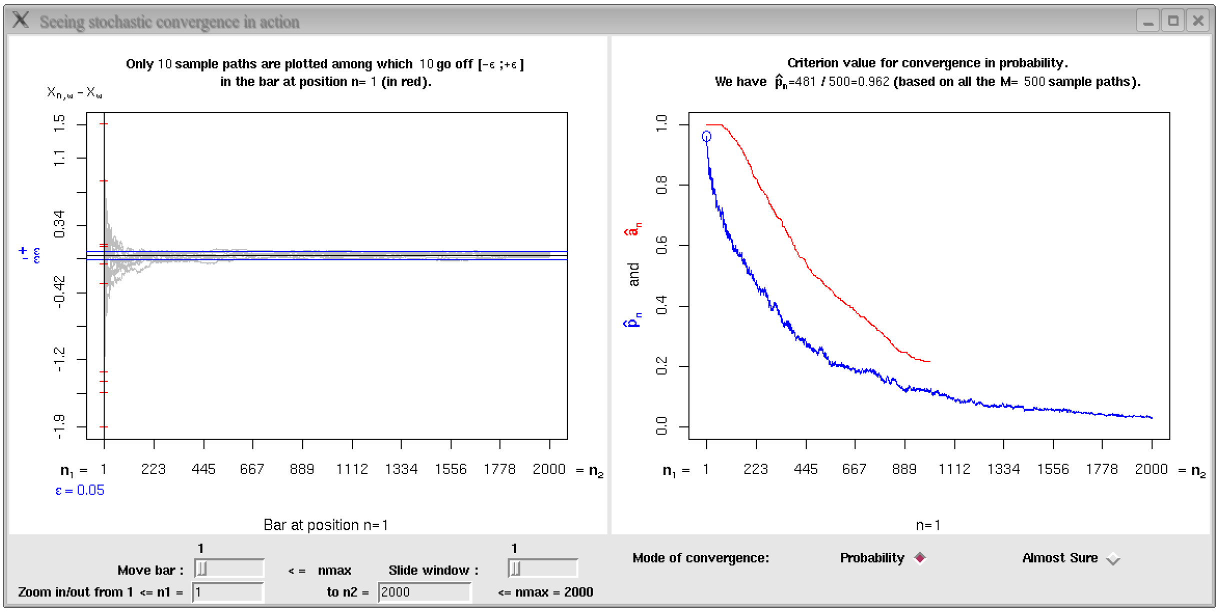

Figure 3. Ten sample paths of $\overline{Y}_n = X_n - X$ amid the 2,000 (left); $\hat{p}_n$ and $\hat{a}_n$ moving toward 0 (right).

able X_n to another random variable X, we can define the random variable $Y_n = X_n - X$ and study the convergence in probability of Y_n to the constant 0. This remark is also valid for almost sure convergence and convergence in rth mean (see Exercise 6).

2.2 Almost Sure Convergence

We write $X_n \xrightarrow{a.s.} X$ and say that the sequence $(X_n)_{n\in\mathbb{N}}$ converges almost surely to X if

$$P\left[\omega; \lim_{n\to\infty} X_{n,\omega} = X_\omega\right] = 1. \tag{2}$$

This means that $\lim_{n\to\infty} X_{n,\omega} = X_\omega$ for all paths $(X_{n,\omega})_{n\in\mathbb{N}}$, except for a set of null probability. Thus almost sure convergence is the familiar pointwise convergence of the sequence of numbers $X_{n,\omega}$ for every ω outside of a null event. To clarify the distinction between convergence in probability and almost sure convergence, we use the following lemma, which contains an equivalent definition of almost sure convergence.

Lemma 1 (Ferguson 1996, p. 5). $X_n \xrightarrow{a.s.} X$ if and only if

$$\forall \epsilon > 0, \qquad a_n = P[\omega; \exists k \geq n; |X_{k,\omega} - X_\omega| > \epsilon] \underset{n\to\infty}{\longrightarrow} 0.$$

Convergence in probability requires that the probability that X_n deviates from X by at least ϵ tends to 0 (for every $\epsilon > 0$). Convergence almost surely requires that the probability that there exists at least a $k \geq n$ such that X_k deviates from X by at least ϵ tends to 0 as n tends to infinity (for every $\epsilon > 0$). This demonstrates that $a_n \geq p_n$ and, consequently, that almost sure convergence implies convergence in probability.

To better explain this notion of almost sure convergence, we use the frequentist approach to approximate the probability a_n by $\hat{a}_n = \frac{1}{M} \times \#\{\exists k \in \{n, \ldots, n_{max}\}; |x_k^j - x^j| > \epsilon\}$.

Mind visualization approach. We can mentally visualize the pieces of sample paths inside the block $[n, n_{max}]$, where n_{max} should be chosen as large as possible. Then we can count the proportion $\hat{a}_n$ of the pieces of sample paths that are outside an horizontal band $[-\epsilon, +\epsilon]$. The aforementioned block is then mentally moved along the n values axis and $\hat{a}_n$ is updated accordingly, as illustrated in Figure 4, which can be considered a static example of our dynamic mental images. The evolution of $\hat{a}_n$ toward 0 informs us about the almost sure convergence (or not) of X_n toward X. Note that we always have (for the same X_i's) $\hat{a}_n \geq \hat{p}_n$, as illustrated in Figure 3.

Example 1 (Continuing). Figure 3 shows the almost sure convergence $X_n = \overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow{a.s.} X = 0$, where the random variables Y_i are iid $N(0, 1)$. We use $M = 500$ realizations and put $\epsilon = 0.05$. We compute $\hat{a}_n$ only for $n = 1, \ldots, Kn_{max} = 1000$ with $n_{max} = 2000$ and with $K = 0.5$ chosen in $(0, 1)$ to ensure sufficient future observations in the last blocks. This allows us to check whether some sample paths lie outside the band $[-\epsilon, +\epsilon]$ in the last block positions. We also compute $\hat{p}_n$ for $n = 1, \ldots, n_{max} = 2000$ to see convergence in probability, and add this to the same plot. We see that $\hat{p}_n$ and $\hat{a}_n$ go to 0.

2.3 Convergence in rth Mean

For a real number $r > 0$, we write $X_n \xrightarrow{r} X$ and say that the sequence $(X_n)_{n\in\mathbb{N}}$ converges to X in the rth mean if

$$e_{n,r} = E|X_n - X|^r \underset{n\to\infty}{\longrightarrow} 0. \tag{3}$$

Here we need to look at the convergence of one sequence of real numbers to 0. Suppose that we want to check the convergence in rth mean of some random variables X_n to X and that we cannot calculate $e_{n,r}$ explicitly. But if we have a generator

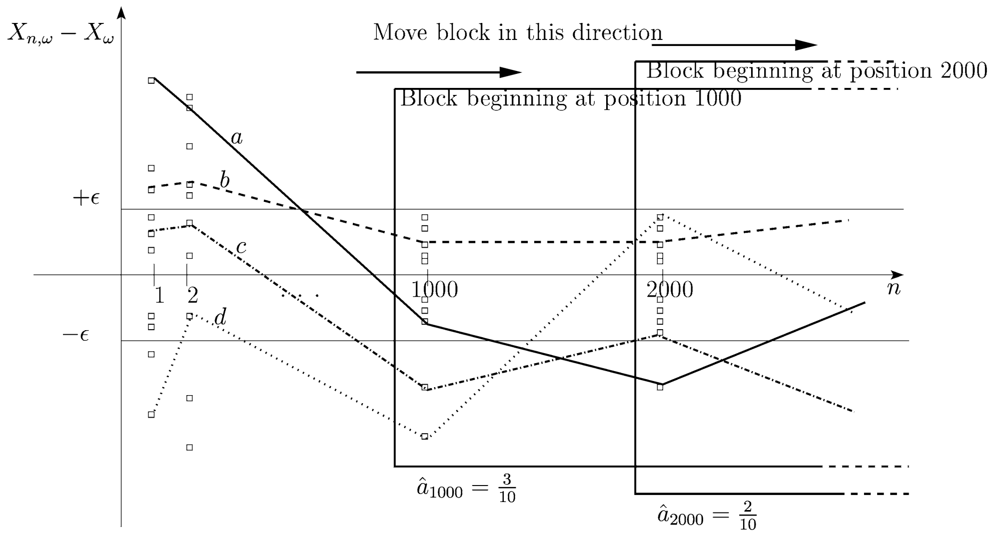

Figure 4. Seeing almost sure convergence with $M = 10$ fictitious realizations. For $n = 1000$, $\hat{a}_n = 3/10$ because we can see three sample paths (a, c, d) lying outside the band $[-\epsilon, +\epsilon]$ in the block beginning at position 1,000. For $n = 2000$, $\hat{a}_n = 2/10$, because we can see two sample paths (a and c) lying outside the band $[-\epsilon, +\epsilon]$ in the block beginning at position 2,000.

of the $X_n - X$, then we can use the following Monte Carlo approximation of $e_{n,r}$:

$$\hat{e}_{n,r} = \frac{1}{M} \sum_{j=1}^{M} |x_n^j - x^j|^r.$$

Then we can plot the $(\hat{e}_{n,r})_{n\in\mathbb{N}}$ sequence for $n = 1$ to a large value, say $n = n_{max}$, to see graphically whether or not it approaches 0.

See online Appendix Example 2 for an illustration (*http://www.biostatisticien.eu/ConvergenceConcepts*).

2.4 Convergence in Law (in Distribution, Weak Convergence)

We write $X_n \xrightarrow{L} X$ and say that the sequence $(X_n)_{n\in\mathbb{N}}$, with distribution functions $(F_n)_{n\in\mathbb{N}}$, converges to X in law if

$$l_n(t) = |F_n(t) - F(t)| \underset{n\to\infty}{\longrightarrow} 0 \tag{4}$$

at all t for which F (the distribution function of X) is continuous.

Here the notion of pointwise convergence of the real numbers $(F_n(t))_{n\in\mathbb{N}}$ to $F(t)$ (for every t at which F is continuous) is involved. Note that we need not look at the realizations of the random variables, because the concept of convergence in law does not require that X_n and X be close in any sense.

In practice, imagine that we want to check the convergence in law of some random variables X_n to a random variable X with known distribution function F, and that we do not have the distribution functions F_n of X_n (defined by $F_n(t) = P[X_n \leq t]$). But if we have a generator of the X_n, then we can use the frequentist approach to approximate the probability $F_n(t)$ by the empirical distribution function

$$\hat{F}_n(t) = \frac{\#\{x_n^j \leq t\}}{M}.$$

Then we can plot $\hat{F}_n(t)$ for different increasing values of n to evaluate whether it approaches $F(t)$. Alternatively, we could use a tridimensional plot of $\hat{l}_n(t) = |\hat{F}_n(t) - F(t)|$ as a function of n and t to evaluate whether it approaches the zero-horizontal plane.

See online Appendix Example 3 for an illustration (*http://www.biostatisticien.eu/ConvergenceConcepts*).

3. POINTING OUT THE DIFFERENCES BETWEEN THE VARIOUS MODES THROUGH EXAMPLES

In Section 1 we noted the only universally valid implications between the various modes of convergence. Under certain additional conditions, some important partial converses hold. Thus, to fully understand all of the modes of convergence described earlier, we believe that it is good pedagogic practice to provide examples in which one weaker type of convergence is valid but a stronger type is not valid. Here we propose an exercise with its solution. We provide five more exercises with solutions in the online Appendix (*http://www.biostatisticien.eu/ConvergenceConcepts*). Students should use our mind visualization approach to perceive the problem, then use our package to investigate it numerically and graphically before trying to demonstrate it rigorously. Students should use our package not as a black box to "prove" some convergence, but rather to support their intuition, which should be based logically on the behavior of the sequence of random variables under investigation.

Exercise 1. Let Z be a uniform $U[0,1]$ random variable and define $X_n = 1_{[m.2^{(-k)};(m+1).2^{(-k)})}(Z)$, where $n = 2^k + m$ for $k \geq 1$ and with $0 \leq m < 2^k$. Thus $X_1 = 1$, $X_2 = 1_{[0,1/2)}(Z)$, $X_3 = 1_{[1/2,1)}(Z)$, $X_4 = 1_{[0,1/4)}(Z)$, and $X_5 = 1_{[1/4,1/2)}(Z), \ldots$. Does $X_n \xrightarrow{a.s.} 0$? Does $X_n \xrightarrow{P} 0$? Does $X_n \xrightarrow{2} 0$?

Solution to Exercise 1. Figure 5 explains the construction of X_n.

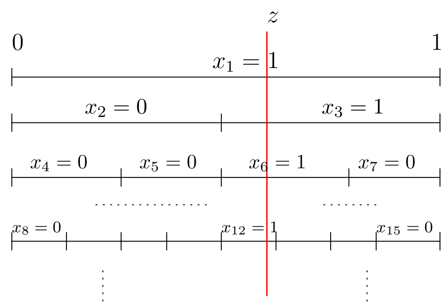

Figure 5. A fictitious sample path for X_n.

Let us apply our mental reasoning as explained in Section 2. Once a z value is randomly drawn, the entire associated sample path is fully determined. As n increases, each sample path "stays" for a longer time at 0 but eventually jumps to 1. In fact, a path will jump to 1 an infinite number of times after each fixed n value. Thus, with the help of Figure 4, we can immediately see that for all $n = 1, \ldots,$ all of the sample paths will jump to 1 somewhere (and even at many places) in the block beginning at position n. This demonstrates that we cannot have almost sure convergence. With regard to the question about convergence in probability, reconsider Figure 2. An understanding of Figure 5 allows us to see that for each increasing fixed n value, the probability that the sample paths lie outside a band $[-\epsilon, \epsilon]$ in the bar at position n corresponds to the proportion of $[0, 1]$-uniform z values falling into an interval with decreasing length. As such, we do have convergence in probability in this case.

Using our package, the user can interactively move the gray block on the left side of Figure 6 and thereby observe the pieces of sample paths that leave the horizontal band. For each sample path, red marks (just above the grey block) indicate the first time that this happens. Simultaneously, we can observe their proportion $\hat{a}_n$ (equaling 1 here) on the right side of Figure 6, represented by a sliding (red) circle. In the same way, we can graphically investigate convergence in probability by sliding the vertical bar (click first on radio button: Probability); we see that $\hat{p}_n$ is going toward 0. This confirms what we perceived by our mind visualization approach.

Now X_n does not converge almost surely toward 0, because we have $\forall\omega \lim_{n\to\infty} X_{n,\omega} \neq 0$. For all n, there always exists a $k \geq n$ such that $X_k = 1$; thus $a_n = 1 \neq 0$. But X_n converges in probability to 0, because $p_n = P[X_n = 1] = \frac{1}{2^k}$, which tends to 0 when $n = 2^k + m \to \infty$ with $0 \leq m < 2^k$. We also see that X_n^2 is a Bernouilli(p_n), so that $E[X_n^2] = \frac{1}{2^k}$, demonstrating that $X_n \xrightarrow{2} 0$.

SUPPLEMENTAL MATERIALS

See the online Supplement to investigate the examples and exercises presented in this paper. (.pdf)

[Received November 2006. Revised September 2008.]

REFERENCES

Billingsley, P. (1986), *Probability and Measure* (2nd ed.), New York: Wiley.

Bryce, G. R., Gould, R., Notz, W., and Peck, R. (2001), "Curriculum Guidelines for Bachelor of Science Degrees in Statistical Science," *The American Statistician*, 55, 7–13.

Chance, B., and Rossman, A. (2006), *Investigating Statistical Concepts, Applications, and Methods*, Belmont, CA: Duxbury.

Chung, K. L. (1974), *A Course in Probability Theory* (2nd ed.), New York: Academic Press.

Dunn, P. K. (1999), "Three Tools for Interactively Visualizing Some Distribution Theory Concepts," *The American Statistician*, 53, 137–139.

Ferguson, T. (1996), *A Course in Large Sample Theory*, London: Chapman & Hall.

Fisher, R. A. (1956), *Statistical Methods and Scientific Inference*, Edinburgh: Oliver & Boyd.

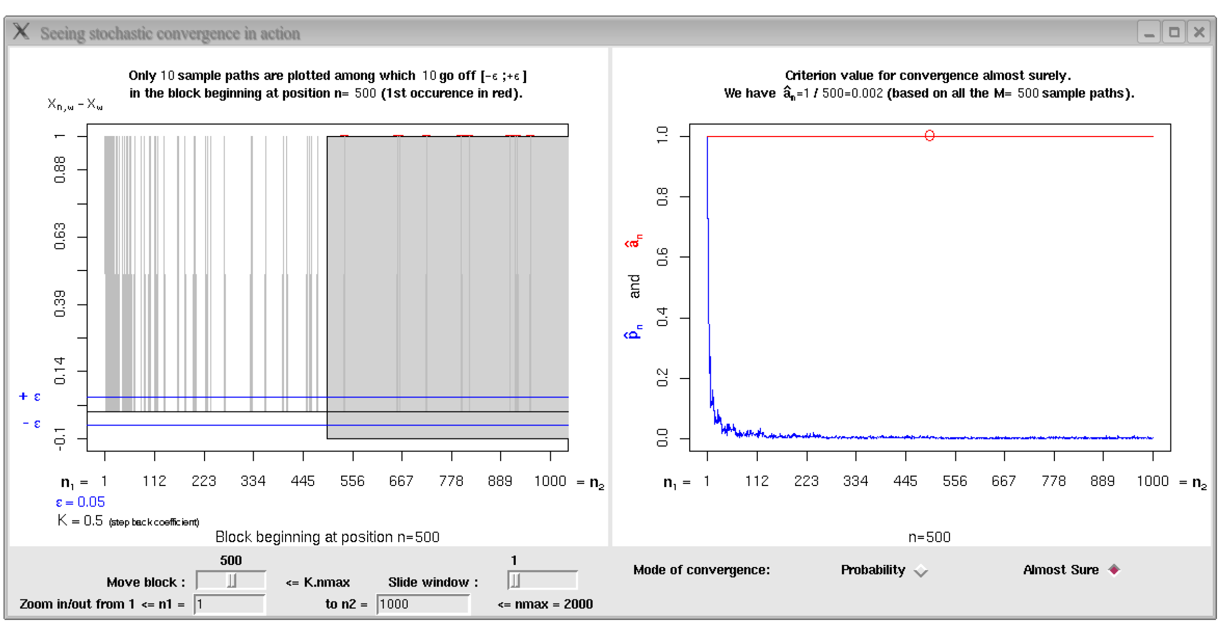

Figure 6. $\hat{p}_n$ going toward 0 and $\hat{a}_n$ equals 1.

Lehmann, E. (2001), *Elements of Large-Sample Theory* (2nd ed.), New York: Springer.

Marasinghe, M. G., Meeker, W. Q., Cook, D., and Shin, T. (1996), "Using Graphics and Simulation to Teach Statistical Concepts," *The American Statistician*, 50, 342–351.

Mills, J. D. (2002), "Using Computer Simulation Methods to Teach Statistics: A Review of the Literature," *Journal of Statistics Education*, 10 (1).

R Development Core Team (2008), *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing, ISBN 3–900051-07-0, URL *http://www.R-project.org*.

Serfling, R. J. (2002), *Approximation Theorems of Mathematical Statistics* (2nd ed.), New York: Wiley-Interscience.

Sethuraman, J. (1995), "Review: Large-Sample Methods in Statistics: An Introduction With Applications," *Journal of the American Statistical Association*, 90, 384.