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

Pierre Lafaye de micheaux and Benoit Liquet

This paper describes the difficult concepts of convergence in probability, almost sure convergence, convergence in law and in r-th mean using a visual-minded and a graphical simulation-based approach. For this purpose, each probability of events is approximated by a frequency. An `R` package is available on CRAN which reproduces all the experiments done in this paper.

KEY WORDS: Convergence in probability, almost surely, in law, in r-th mean; R language; Visualization; Dynamic graphics; Monte Carlo Simulation.

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 in r-th mean (r) at the graduate level (see Sethuraman (1995)). Indeed, as pointed out by Bryce (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 (see 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 (for example Billingsley (1986), Chung (1974), 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 also to provide software useful for learning them. Each convergence notion provides an essential foundation for further

Pierre Lafaye de Micheaux is Associate Professor, Department of Statistics, LJK, Grenoble University, France (Email: `Pierre.Lafaye-de-Micheaux@upmf-grenoble.fr`). Benoit Liquet is Associate Professor, INSERM U897, ISPED, University Bordeaux 2, France (Email: `Benoit.Liquet@isped.u-bordeaux2.fr`). The authors thank R.Drouilhet, M.Lejeune and all the referees, editor and associate editor for their many helpful comments.

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'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. Second, a graphical simulation-based approach is used to illustrate this perspective and to investigate some modes of convergence in practical situations. Section 3 points out subtler distinctions between the various modes through examples. This is illustrated through exercises and solutions which emphasize our visualization approach in an online appendix. We propose an `R` (R Development Core Team, 2006) package named `ConvergenceConcepts`. The interactive part of the package provides an interesting pedagogic tool facilitating the visualization of the convergence concepts. The package also creates all the figures presented here, and can be used to investigate the convergence of any random variable. This approach aims to help students to develop intuition and logical thinking.

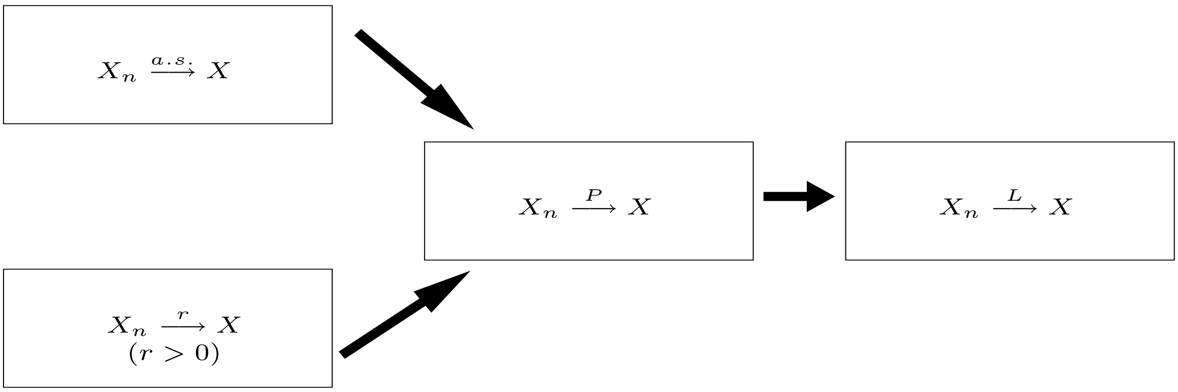

Figure 1: Universally valid implications of the four classical modes of convergence. See Ferguson (1996) for proofs.

2. MODES OF CONVERGENCE

Probability theory is the body of knowledge that enables us to reason formally 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 will mainly be interested 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 will use the following property

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

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

In the sequel, we will study the convergence (in some sense to be defined) of sequences of random variables X_n to X. We note $(x_n^j - x^j)_{n \in \mathbb{N}} = (x_1^j - x^j, x_2^j - x^j, \dots, x_n^j - x^j, \dots)$, the j-th sample path of $(X_n - X)_{n \in \mathbb{N}}$.

2.1 Convergence in Probability

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

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

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

Mind visualization approach:

We can mentally visualize the M sample paths of the stochastic process $(X_n - X)_{n=1,\dots,n_{max}}$. 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 an horizontal band $[-\epsilon, +\epsilon]$. This band can be chosen to be arbitrarily narrow. The sample paths should only be observed at each fixed position n, for example by mentally sliding along the n values axis a highlighting vertical bar. This is illustrated in Figure 2 which can be seen as a static example of our dynamic mental images. The evolution of $\hat{p}_n$ towards 0 informs us about the convergence (or not) in probability of X_n towards X.

In order to have a better understanding of 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.

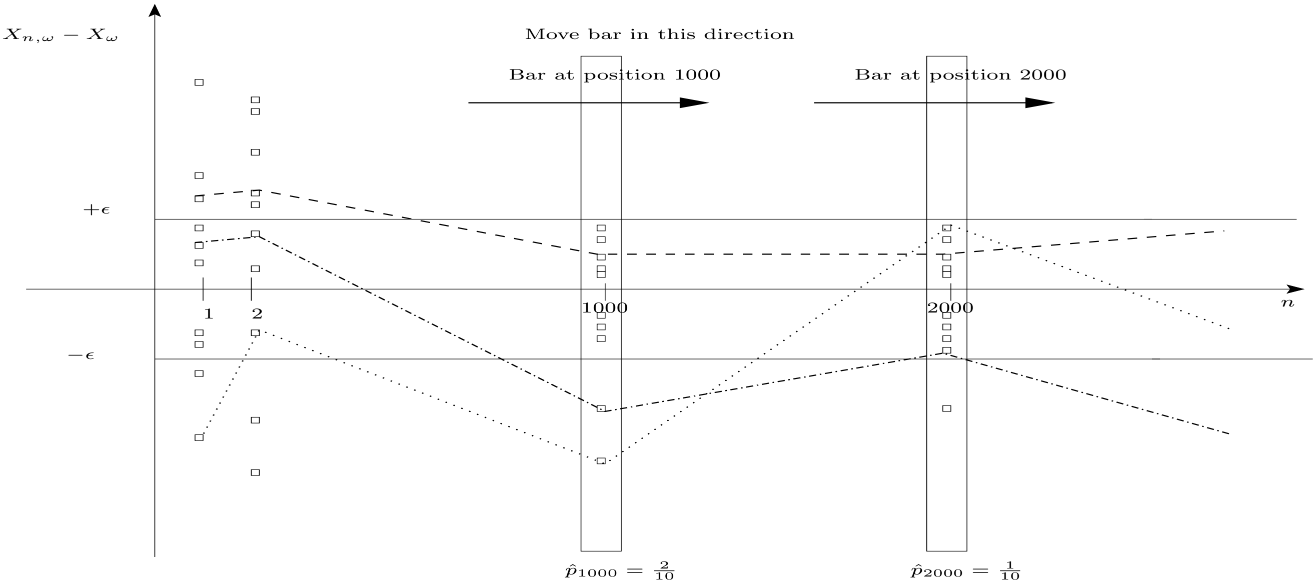

Figure 2: Seeing convergence in probability with $M = 10$ fictitious realizations. For $n = 1000$, $\hat{p}_n = 2/10$ since we can see two sample paths lying outside the band $[-\epsilon, +\epsilon]$ in the bar at position 1000. For $n = 2000$, $\hat{p}_n = 1/10$ since we can see one sample path lying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

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 independent and identically distributed (*i.i.d.*) $N(0, 1)$. 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 which are lying outside the horizontal band as indicated by 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 variable X_n to another random variable X, you 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 r-th mean (see Exercise 6).

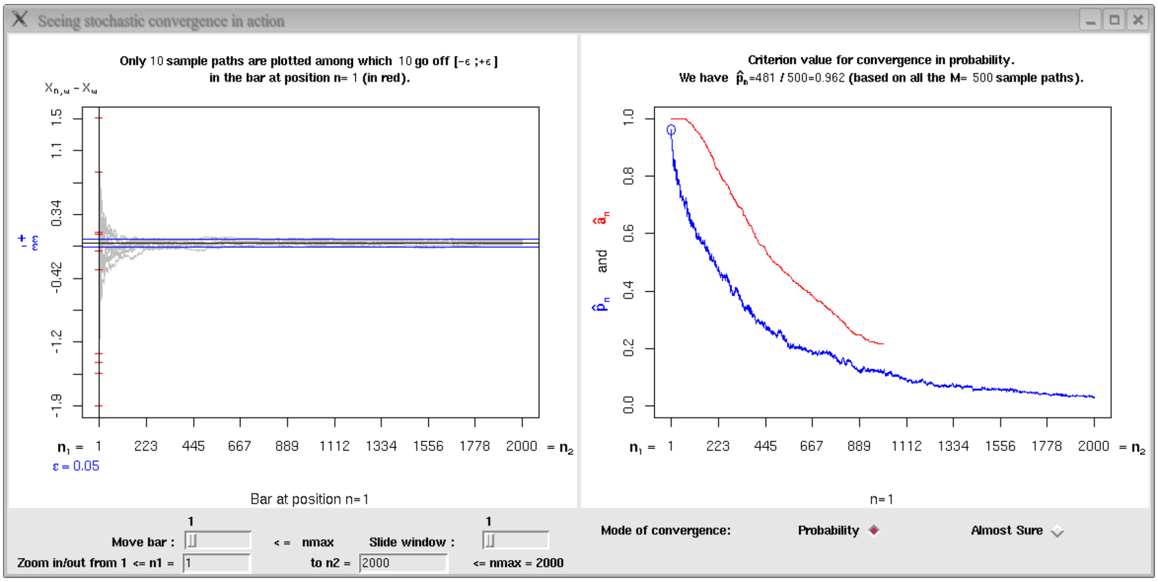

Figure 3: Ten sample paths of $\overline{Y}_n = X_n - X$ amid the 2000 (left); $\hat{p}_n$ and $\hat{a}_n$ moving towards 0 (right).

2.2 Almost sure convergence

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

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

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. So 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 will use the following lemma which contains an equivalent definition of almost sure convergence.

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

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

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 zero as n tends to infinity (for every $\epsilon > 0$). This shows that $a_n \geq p_n$ and consequently that almost sure convergence implies convergence in probability.

To understand 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 seen as a static example of our dynamic mental images. The evolution of $\hat{a}_n$ towards 0 informs us about the almost sure convergence (or not) of X_n towards X. Note that we always have (for the same X_i's) $\hat{a}_n \geq \hat{p}_n$ which is illustrated in Figure 3.

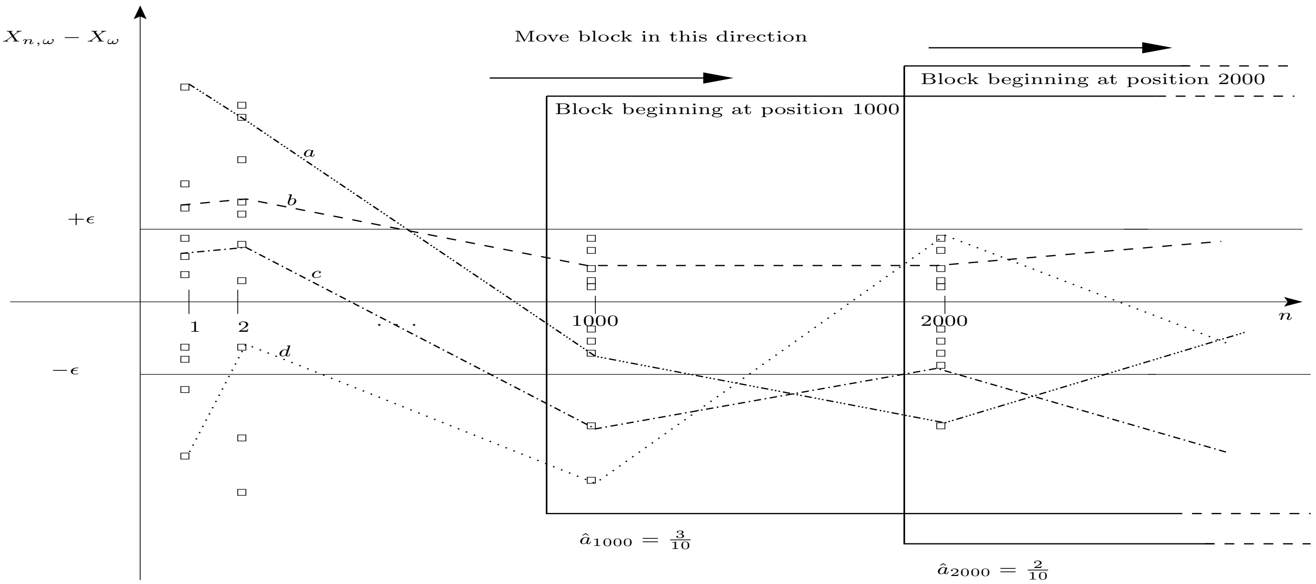

Figure 4: Seeing almost sure convergence with $M = 10$ fictitious realizations. For $n = 1000$, $\hat{a}_n = 3/10$ since we can see 3 sample paths (a, c, d) lying outside the band $[-\epsilon, +\epsilon]$ in the block beginning at position 1000. For $n = 2000$, $\hat{a}_n = 2/10$ since we can see 2 sample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

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 *i.i.d.* $N(0, 1)$. We use $M = 500$ realizations and we put $\epsilon = 0.05$. We compute $\hat{a}_n$ only for $n = 1, \ldots, K.n_{max} = 1000$ with $n_{max} = 2000$ and with $K = 0.5$ chosen in $(0, 1)$ in order to have enough future observations in the last blocks. This allows us to see if there are some sample paths that 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 it to the same plot. We can see that $\hat{p}_n$ and $\hat{a}_n$ go to 0.

2.3 Convergence in r-th mean

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

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

Here one has to look at the convergence of one sequence of real numbers to 0. Suppose that we would like to check the convergence in r-th mean of some random variables X_n to X and that we cannot calculate $e_{n,r}$ explicitly. However, if we have a generator of the $X_n - X$, 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, $n = n_{max}$ say, to see graphically if it approaches 0 or not.

See online Appendix Example 2 for an illustration.

2.4 Convergence in law (in distribution, weak convergence)

We shall 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, it is 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) which is involved. Note that we do not have to look at the realizations of the random variables, as the concept of convergence in law does not necessitate that X_n and X are close in any sense.

In practice, imagine that we would like 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 $F_n(t) = P[X_n \le t]$. However, if we have a generator of the X_n, we can use the frequentist approach to approximate the probability $F_n(t)$ by the empirical distribution function

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

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

See online Appendix Example 3 for an illustration.

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

We recalled in the introduction 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 the previously encountered modes of convergence, we think it is good pedagogic practice to provide examples where one weaker type of convergence is valid whereas a stronger type is not. We propose here one exercise with its solution. Five more exercises with their solutions are provided in the online Appendix. 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 not use our package as a black box to "prove" some convergence but rather to support their intuition, which should be based logically on the behaviour 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 $X_1 = 1$, $X_2 = 1_{[0,1/2)}(Z)$, $X_3 = 1_{[1/2,1)}(Z)$, $X_4 = 1_{[0,1/4)}(Z)$, $X_5 = 1_{[1/4,1/2)}(Z)$, ... Does $X_n \xrightarrow{a.s.} 0$? Does $X_n \xrightarrow{P} 0$? Does $X_n \xrightarrow{2} 0$?

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

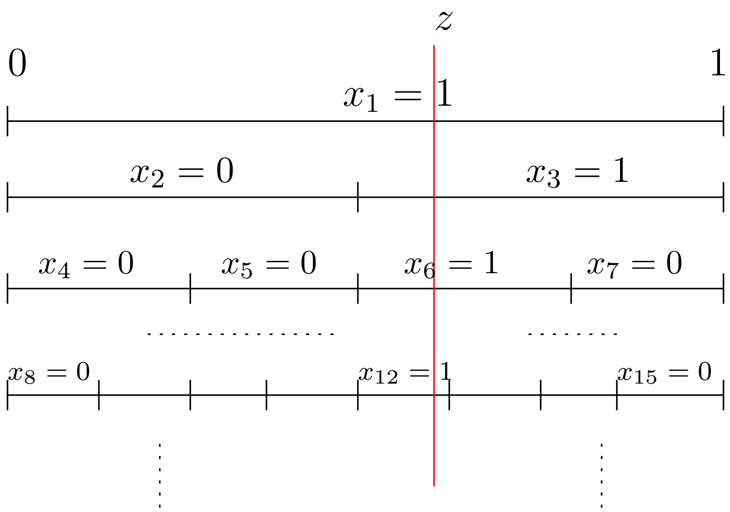

Figure 5: One fictitious sample path for X_n.

Lets 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 it will jump to 1 an infinite number of times after each fixed n value. So, with the help of Figure 4, one can immediately see that for all $n = 1, \ldots$, all the sample paths will jump to 1 somewhere (and even at many places) in the block beginning at position n. This shows that we cannot have almost sure convergence.

With regard to the question about convergence in probability, you should look back at Figure 2. If you have understood Figure 5, you can see that for each increasing fixed n value, the probability that the sample paths lies 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 whose length gets narrower. This lets us perceive that in this case we do have convergence in probability.

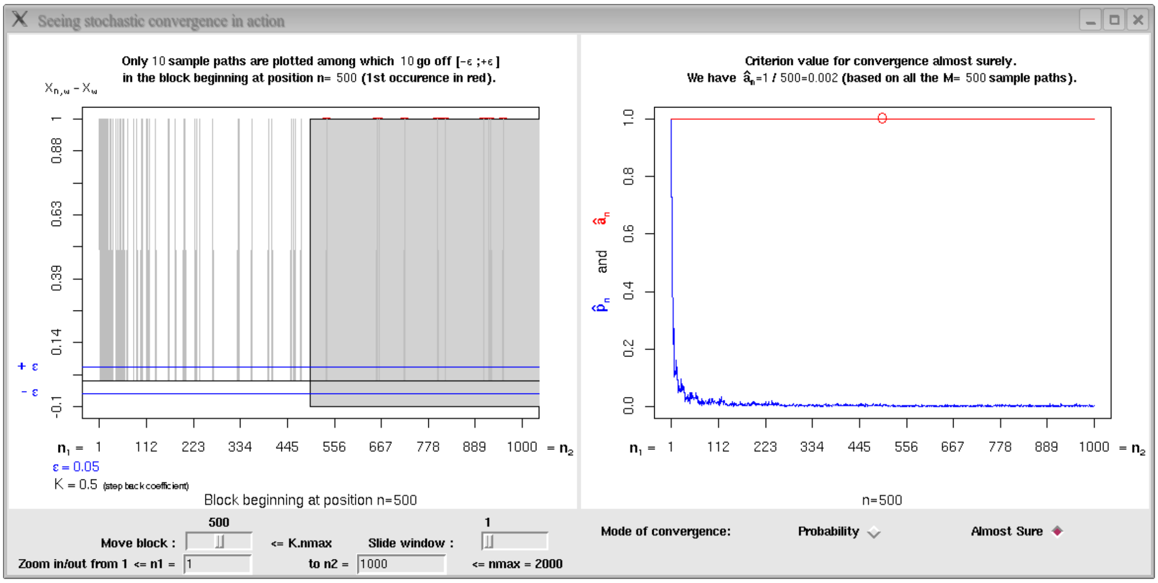

Figure 6: $\hat{p}_n$ going towards 0 and $\hat{a}_n$ equals 1.

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

Now X_n does not converge almost surely towards 0 since 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 $X_k = 1$. So $a_n = 1 \neq 0$.

However, X_n converges in probability to 0 since $p_n = P[X_n = 1] = \frac{1}{2^k}$ 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 $E[X_n^2] = \frac{1}{2^k}$ which shows that $X_n \xrightarrow{2} 0$.

REFERENCES

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

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

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

Chung, K. L. (1974), *A Course in probability theory*, (second edition), New York: Academic press.

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

Ferguson, T. (1996), *A course in large sample theory*, Chapman and Hall.

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

Lehmann, E. (2001), *Elements of Large-Sample Theory*, (second edition), Springer.

Marasinghe, M. G., Meeker, W. Q., Cook, D., Shin, T. (1996), "Using Graphics and Simulation to Teach Statistical Concepts," *The American Statistician*, 50, pp. 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 (2006), *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Serfling, R. J. (2002), *Approximation Theorems of Mathematical Statistics*, (second edition), Wiley-Interscience.

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

Online Appendix

To use our package named `ConvergenceConcepts`, download it from CRAN (`http://cran.r-project.org`) and install it along with the required dependencies. Then launch R and type in its console the following instructions:

```
require(ConvergenceConcepts)

investigate()
```

Now, you can investigate all the examples and exercises presented in this paper.

A Examples

A.1 Convergence in r-th mean

Example 2. We would like to investigate the convergence in r-th mean (for $r = 1, 2, 3$ say) of X_n towards $X = 0$, where the X_n are independent random variables such that $P[X_n = n^{0.4}] = 1/n$ and $P[X_n = 0] = 1 - 1/n$. One can show that $E|X_n|^r = n^{0.4r-1}$ and thus $X_n \stackrel{r}{\longrightarrow} 0$ for $r = 1, 2$ but not for $r = 3$. This can be observed on the following plot (see Figure 7) where we took $nmax = 2000$ and $M = 500$.

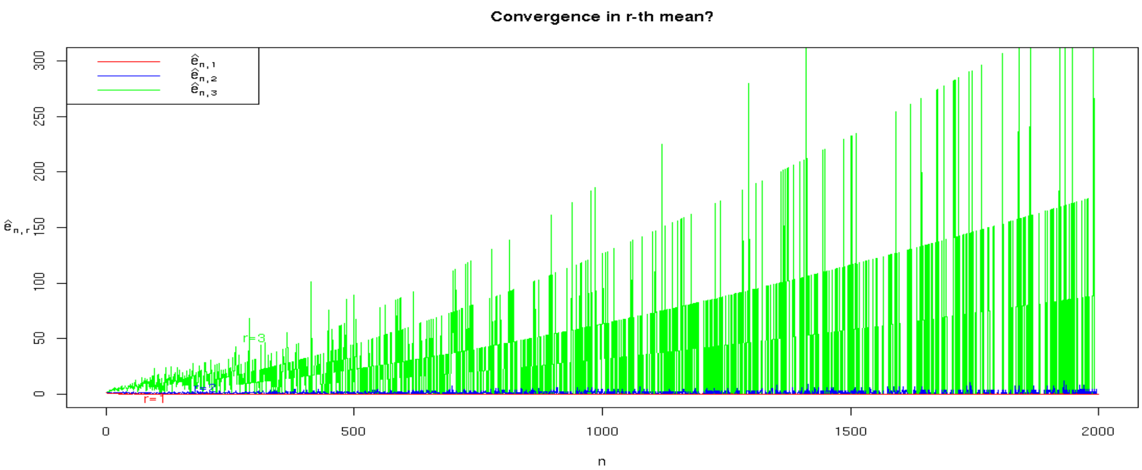

Figure 7: $\hat{e}_{n,1}$ (red) and $\hat{e}_{n,2}$ (blue) going towards 0 ; $\hat{e}_{n,3}$ (green) not going towards 0.

A.2 Convergence in law

Example 3. Figure 8 shows the convergence in distribution of $X_n = \frac{1}{\sqrt{n}}\left[\frac{\sum_i^n Z_i - n}{\sqrt{2}}\right]$ towards $N(0, 1)$ where the Z_i are *i.i.d.* χ^2_1 random variables. On the left you can see an output of our `law.plot2d` function with the slider value fixed at $n = 70$. The distribution function of a standard Gaussian is plotted in black whereas the empirical distribution function of X_n based on $M = 5000$ realizations is plotted in red. We can move the slider and see that the red curve comes closer to the black one. Also, on the right you can see the tri-dimensional plot of $|\hat{F}_n(t) - F(t)|$ for $n = 1, \ldots, nmax = 200$ to see if gets closer to the zero horizontal plane. These plots suggest a convergence in distribution.

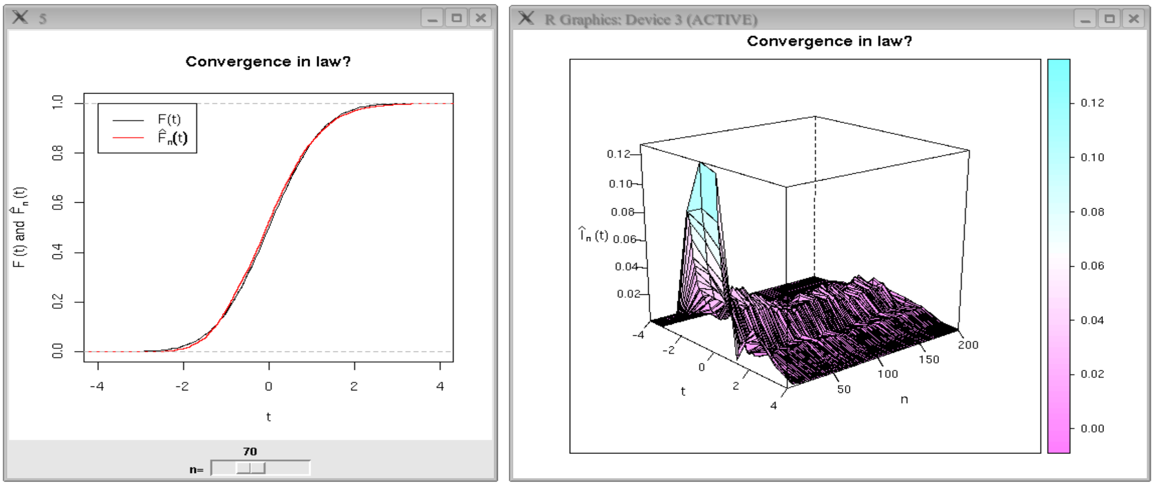

Figure 8: Convergence in distribution in action on a simulated example. Left: the distribution function of a standard Gaussian is plotted in black whereas the empirical distribution function of X_n ($n = 70$) based on $M = 5000$ realizations is plotted in red. Right: tri-dimensional plot of $|\hat{F}_n(t) - F(t)|$ as a function of n and t.

B Exercises

Exercise 2. Let $X_1, X_2, \ldots, X_n$ be *i.i.d.* $N(0, 1)$ random variables and $X = X_1$. Does $X_n \xrightarrow{L} X$? Does $X_n \xrightarrow{P} X$?

Exercise 3. Let $X_1, X_2, \ldots, X_n$ be independent random variables such that $P[X_n = \sqrt{n}] = \frac{1}{n}$ and $P[X_n = 0] = 1 - \frac{1}{n}$. Does $X_n \xrightarrow{2} 0$? Does $X_n \xrightarrow{P} 0$?

Exercise 4. Let Z be $U[0, 1]$ and let $X_n = 2^n 1_{[0,1/n)}(Z)$. Does $X_n \xrightarrow{r} 0$? Does $X_n \xrightarrow{a.s.} 0$?

Exercise 5. Let $Y_1, Y_2, \ldots, Y_n$ be independent random variables with mean 0 and variance 1. Define $X_1 = X_2 = 1$ and

$$X_n = \frac{\sum_{i=1}^{n} Y_i}{(2n \log \log n)^{1/2}}, n \geq 3.$$

Does $X_n \xrightarrow{2} 0$? Does $X_n \xrightarrow{a.s.} 0$?

Exercise 6. Let $Y_1, Y_2, \ldots, Y_n$ be independent random variables with uniform discrete distribution on $\{0, \ldots, 9\}$. Define

$$X_n = \sum_{i=1}^{n} \frac{Y_i}{10^i}.$$

It can be proved that $X_n \xrightarrow{a.s.} X = \sum_{i=1}^{\infty} \frac{Y_i}{10^i}$ which follows a $U[0, 1]$ distribution. Now, let $Z \sim U[0, 1]$ independent of X.

Does $X_n \xrightarrow{a.s.} Z$? Does $X_n \xrightarrow{L} Z$?

C Solutions to the exercises

Solution to Exercise 2.

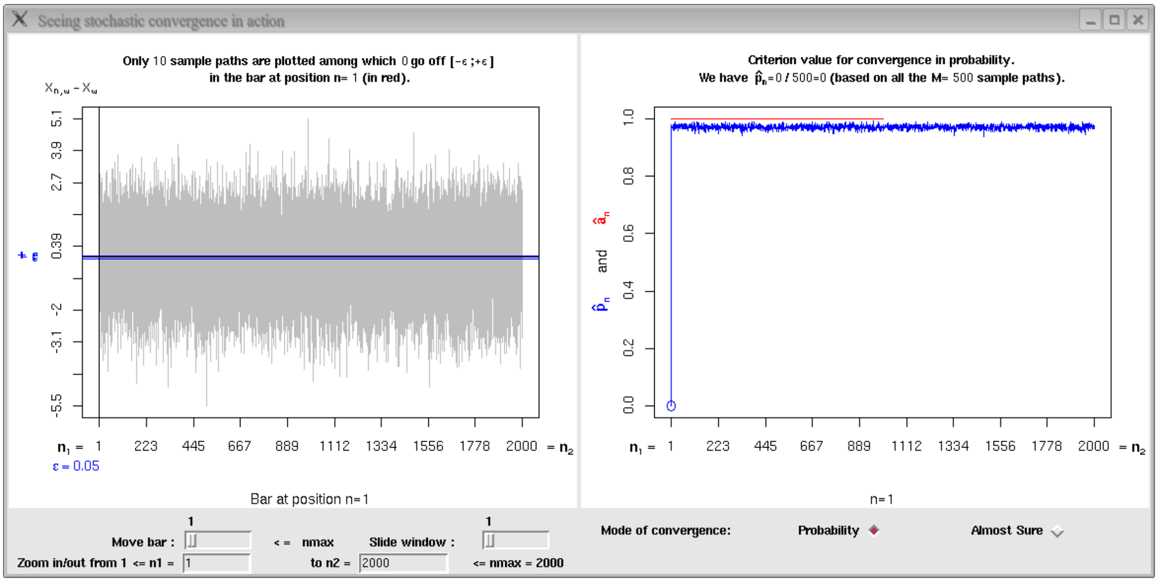

Figure 9: Ten sample paths of $X_n - X_1$ amid the 500 (left); $\hat{p}_n$ (resp. $\hat{a}_n$) going towards $p_n \neq 0$ (resp. $a_n = 1$) (right).

It is trivial that X_n converges in law to X_1 since for each n both X_n and X have the same distribution function. Now, since X_n and X are independent, $X_{n,\omega} - X_\omega$ has no particular reason to be close to 0 for any n or any ω. Thus we do not have $X_n \xrightarrow{P} X$. It can be seen on the plot of Figure 9 that $X_{n,\omega} - X_\omega$ tends to be far from 0 and that $\hat{p}_n$ and $\hat{a}_n$ are not going towards 0. Indeed, in this case, by noting that $X_n - X \sim N(0, 2)$, one can obtain explicitly

$$p_n = 2\left[1 - \Phi\left(\epsilon/\sqrt{2}\right)\right] \simeq 0.9718 \neq 0 \text{ (for } \epsilon = 0.05) \quad (5)$$

where $\Phi(\cdot)$ denotes the standard $N(0, 1)$ distribution function. Thus $X_n \overset{a.s.}{\nrightarrow} X$ and $X_n \overset{P}{\nrightarrow} X$.

Solution to Exercise 3.

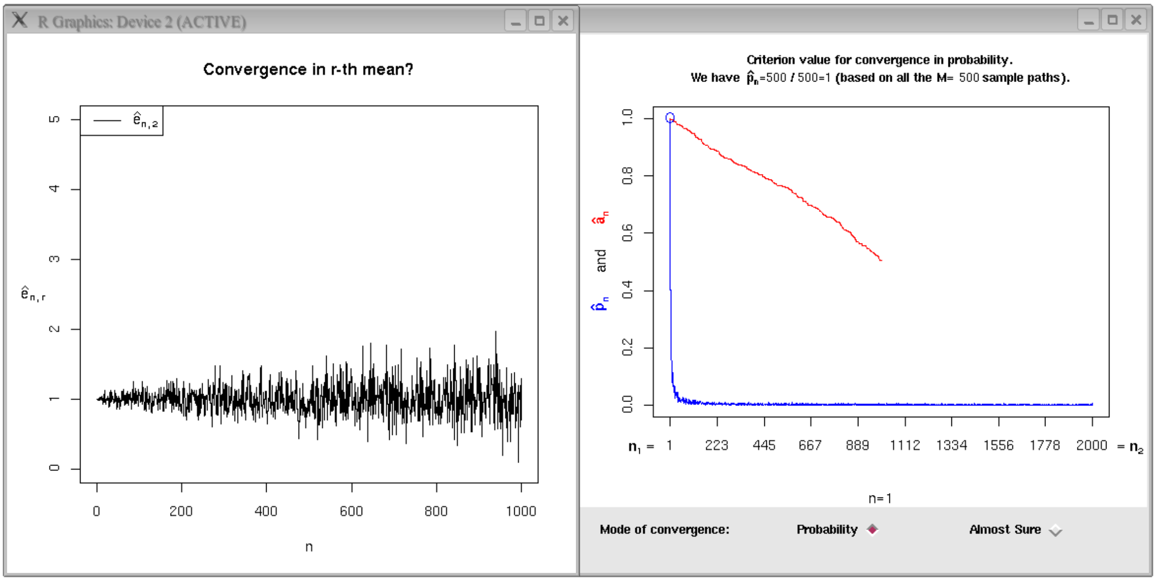

Figure 10: $\hat{e}_{n,2}$ not going towards 0 (left) and $\hat{p}_n$ going towards 0 (right).

We can mentally visualize each sample path to be essentially equal to 0, but to sometimes jump higher and higher, as n increases, with a decreasing probability. This gives us the intuition that X_n converges in probability to 0. On the other hand, for a fixed n, the mean of the $X_{n,\omega}^2$ values is taken away from 0 due to these few but very large values. But for increasing values of n, one can not say if the mean of the $X_{n,\omega}^2$ values will decrease or not. So we cannot tell more about the quadratic mean convergence to 0.

The intuition for convergence in probability is confirmed using our package ($\hat{p}_n$ is going to 0, see Figure 10). But we can expect that we do not have convergence in quadratic mean towards 0 because $\hat{e}_{n,2}$ is not going to 0 but oscillates around 1.

Now, one can prove that X_n does not converge in a quadratic mean to 0 since $e_{n,2} = E|X_n|^2 = 1, \forall n$ and that X_n converges to 0 in probability since $p_n = \frac{1}{n} \to 0$.

Solution to Exercise 4.

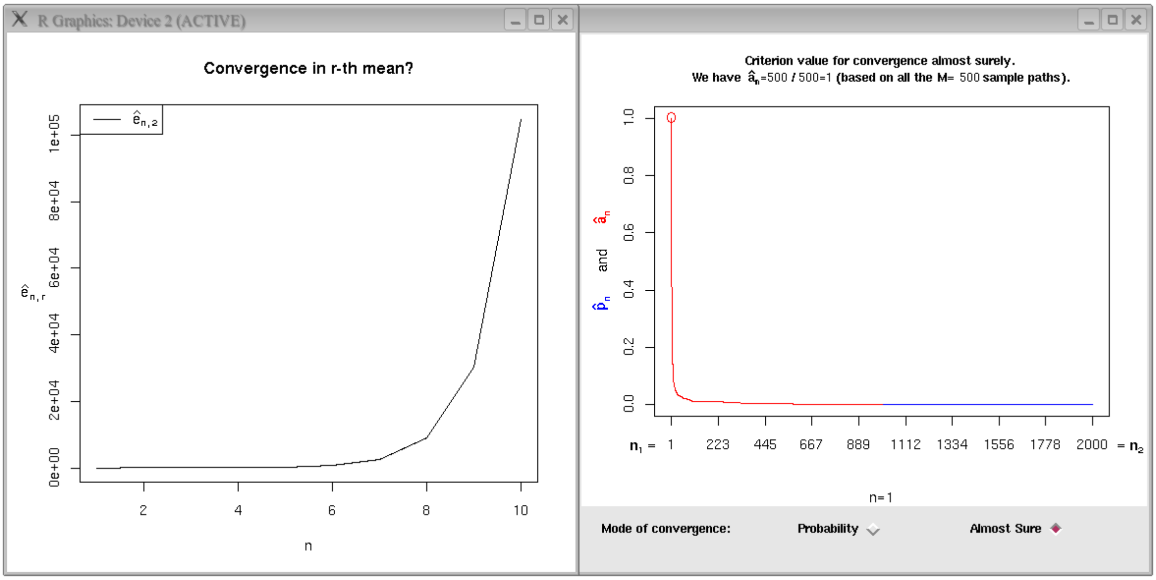

Figure 11: $\hat{e}_{n,2}$ not going towards 0 (left) and $\hat{a}_n$ going towards 0 (right). We plotted the left graph only for the very first n values since divergence is very fast here.

We can mentally visualize each sample path to be growing to large values then suddenly dropping to 0 and after that staying infinitely at this null value. These sample paths can also be visualized using our package with the possibility to use the "zoom in" facility. This gives us the intuition that X_n converges almost surely to 0 since $\forall\omega$, $\lim_{n\to\infty} X_{n,\omega} = 0$. On the other hand, for a fixed n, the mean of the $X^2_{n,\omega}$ values is taken away from 0 due to the small proportion of sample paths that take very large values. But for increasing values of n, one can not say if the mean of the $X^2_{n,\omega}$ values will decrease or not. So we cannot tell more using our intuition about the quadratic mean convergence to 0.

Convergence almost surely to 0 is illustrated using our package ($\hat{a}_n$ is going to 0, see Figure 11). But we can expect that we do not have convergence in quadratic mean towards 0 because $\hat{e}_{n,2}$ is not going to 0.

We can now prove that X_n does not converge to 0 in r-th mean since $E|X_n|^r = \frac{2^{rn}}{n} \to \infty$.

Solution to Exercise 5.

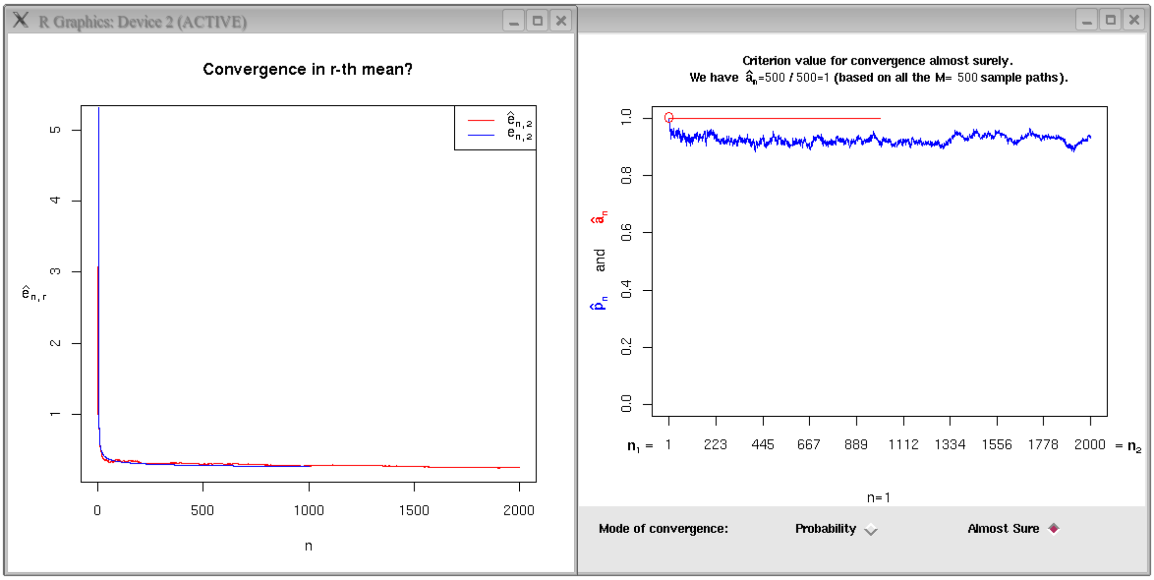

Figure 12: $\hat{e}_{n,2}$ going towards 0 (left) and $\hat{a}_n$ equals 1 (right).

Looking at the definition of X_n, we do not get a precise information on the sample paths. So, intuition cannot be of great help in this case. Thus, we use our package (with Y_i i.i.d. $N(0,1)$) to get some clue on quadratic convergence and almost sure convergence.

Figure 12 shows that $\hat{e}_{n,2}$ is going towards 0 and that $\hat{a}_n$ equals 1. This suggests a quadratic mean convergence, and not an almost sure convergence.

We can now prove that X_n converges in a quadratic mean to 0 since $E|X_n|^2 = \frac{1}{2\log\log n}$ for all n. We added a blue curve on the plot for the function $e_{n,2} = \frac{1}{2\log\log n}$ and we see that the blue and red curves are superposed.

To prove almost sure convergence, we have to use the law of the iterated logarithm (see Billingsley, 1995, p.154) that can be formulated as $P[X_n > 1-\epsilon,\ \text{infinitely often}] = 1$. This suffices to prove that X_n does not converge to 0 almost surely.

Solution to Exercise 6.

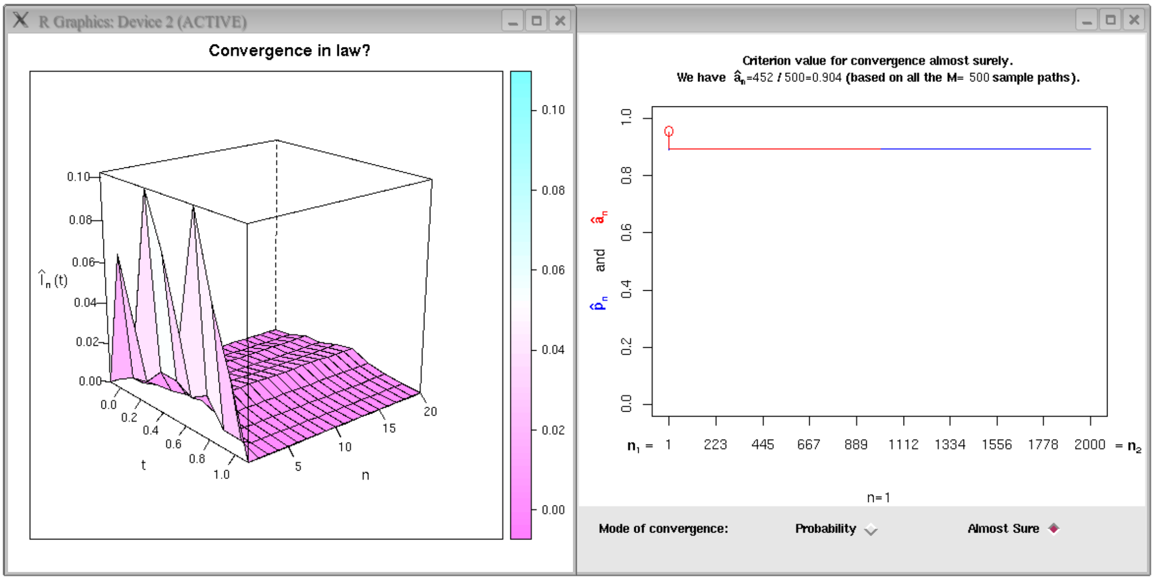

Figure 13: $\hat{l}_n(t)$ going towards 0 (left) ; $\hat{a}_n$ not going to 0 (right).

Since X_n and Z are independent, $X_{n,\omega} - Z_\omega$ has no particular reason to be close to 0 for any n or any ω. Thus we do not have $X_n \xrightarrow{a.s.} Z$. It can be seen on the plot of Figure 13 that $X_{n,\omega} - Z_\omega$ tends to be far from 0 and that $\hat{a}_n$ is not going towards 0. Using our package, we can also see that $\hat{l}_n(t)$ is going towards 0 forall t. This suggests a convergence in law of X_n towards Z. Indeed, as almost sure convergence implies convergence in law, we have $X_n \xrightarrow{L} X$ and since X and Z are both $U[0,1]$, $X_n \xrightarrow{L} Z$.

Now, lets us prove rigorously that $X_n \overset{a.s.}{\nrightarrow} Z$. We have $X_n - Z = X_n - X + X - Z \xrightarrow{a.s.} X - Z$ (by Slutsky theorem, see Ferguson (1996) p.42). Therefore $X_n - Z \xrightarrow{L} X - Z$ which implies that $\forall \epsilon > 0, p_n = P[|X_n - Z| > \epsilon] \underset{n\to\infty}{\longrightarrow} P[|X - Z| > \epsilon] = (1-\epsilon)^2 = 0.9025$ (for $\epsilon = 0.05$). Thus, $X_n \overset{P}{\nrightarrow} Z$ and so $X_n \overset{a.s.}{\nrightarrow} Z$. Note that the density function $p(.)$ of the difference of two $U[0,1]$ is given by $p(z) = (1+z)1_{\{-1\leq z\leq 0\}} + (1-z)1_{\{0\leq z\leq 1\}}$.

Note that if X and Z are two independent non constant random variables with the same law, we can have $X_n \xrightarrow{a.s.} X$ (i.e. $X_{n,\omega} \to X_\omega$ almost everywhere) but $X_n \overset{a.s.}{\nrightarrow} Z$ because we may not have $X_\omega = Z_\omega \ \forall \omega \in \Omega$. But, in the case where X and Z are constant random variables with the same law, we have obviously $X = Z$ and thus trivially $X_n \xrightarrow{a.s.} X$ implies that $X_n \xrightarrow{a.s.} Z$.