Defining the p-dimensional vector random sequence $\mathbf{X}[n] = [X[n], \ldots, X[n-p+1]]^T$, and coefficient matrix

$$\mathbf{A} = \begin{bmatrix} a_1 & a_2 & \cdots & \cdots & a_p \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix},$$

we have

$$\mathbf{X}[n] = \mathbf{A}\mathbf{X}[n-1] + \mathbf{b}W[n].$$

Thus $\mathbf{X}[n]$ is a vector Markov random sequence with $\mathbf{b} = [b, 0, \ldots, 0]^T$. Such a vector transformation of a scalar equation is called a *state-variable representation* [6-8].

6.7 CONVERGENCE OF RANDOM SEQUENCES

Some nonstationary random sequences may converge to a limit as the sequence index goes to infinity. This asymptotic behavior is evidenced in probability theory by convergence of the fraction of successes in an infinite Bernoulli sequence, where the relevant theorems are called the laws of large numbers. Also, when we study the convergence of random processes in Chapter 8 we will sometimes make a sequence of finer and finer approximations to the output of a random system at a given time, say, t_0 , that is, $Y_n(t_0)$. The index n then defines a random sequence, which should converge in some sense to the true output. In this section we will look at several types of convergence for random sequences, that is, sequences of random variables.

We start by reviewing the concept of convergence for deterministic sequences. Let x_n be a sequence of complex (or real) numbers; then convergence is defined as follows.

Definition 6.7-1 A sequence of complex (or real) numbers x_n converges to the complex (or real) number x if given any $\varepsilon > 0$, there exists an integer n_0 such that whenever $n > n_0$, we have

$$|x_n - x| < \varepsilon. \quad \blacksquare$$

Note that in this definition the value n_0 may depend on the value ε; that is, when ε is made smaller then most likely n_0 will need to be made larger. Sometimes this dependence is formalized by writing $n_0(\varepsilon)$ in place of n_0 in this definition. This is often written as

$$\lim_{n\to\infty} x_n = x \qquad \text{or as} \qquad x_n \to x \text{ as } n \to \infty.$$

A practical problem with this definition is that one must have the limit x to test for convergence. For simple cases one can often guess what the limit is and then use the definition to verify that this limit indeed exists. Fortunately, for more complex situations there is an alternative in the *Cauchy criterion* for convergence, which we state as a theorem without proof.

Theorem 6.7-1 (Cauchy criterion [6-5].) A sequence of complex (or real) numbers x_n converges to a limit *if and only if* (iff)

$$|x_n - x_m| \to 0 \text{ as both } n \text{ and } m \to \infty.$$

The reason that this works for complex (or real) numbers is that the set of all complex (or real) numbers is *complete*, meaning that it contains all its limit points. For example, the set $\{0 < x < 1\} = (0,1)$ is not complete, but the set $\{0 \le x \le 1\} = [0,1]$ is complete because sequences x_n in these sets and tending to 0 or 1 have a *limit point* in the set $[0,1]$ but have no limit point in the set $(0,1)$. In fact, the set of all complex (or real) numbers is complete as well as n-dimensional linear vector spaces over both the real and complex number fields. Thus the Cauchy criterion for convergence applies in these cases also. For more on numerical convergence see [6-5].

Convergence for functions is defined using the concept of convergence of sequences of numbers. We say the sequence of functions $f_n(x)$ converges to the function $f(x)$ if the corresponding sequence of numbers converges for each x. It is stated more formally in the following definition.

Definition 6.7-2 The sequence of functions $f_n(x)$ converges (pointwise) to the function $f(x)$ if for each x_0 the sequence of complex numbers $f_n(x_0)$ converges to $f(x_0)$. ■

The Cauchy criterion for convergence applies to pointwise convergence of functions if the set of functions under consideration is complete. The set of continuous functions is not complete because a sequence of continuous functions may converge to a discontinuous function (cf. item (d) in Example 6.7-1). However, the set of bounded functions is complete [6-5].

The following are some examples of convergent sequences of numbers and functions. We leave the demonstration of these results as exercises for the reader.

Example 6.7-1

(Some convergent sequences.)

(a) $x_n = (1 - 1/n)a + (1/n)b \to a$ as $n \to \infty$.

(b) $x_n = \sin(\omega + e^{-n}) \to \sin\omega$ as $n \to \infty$.

(c) $f_n(x) = \sin[(\omega + 1/n)x] \to \sin(\omega x)$, as $n \to \infty$ for any (fixed) x.

(d) $f_n(x) = \left\{ \begin{array}{ll} e^{-n^2 x}, & \text{for } x > 0 \\ 1, & \text{for } x \le 0 \end{array} \right\} \to u(-x)$, as $n \to \infty$ for any (fixed) x.

The reader should note that in the convergence of the functions in (c) and (d), the variable x is held constant as the limit is being taken. The limit is then repeated for each such x value to find the limiting function.

Since a random variable is a function, a sequence of random variables (also called a random sequence) is a sequence of functions. Thus we can define the first and strongest type of convergence for random variables.

Definition 6.7-3 (Sure convergence.) The random sequence $X[n]$ *converges surely* to the random variable X if the sequence of functions $X[n, \zeta]$ converges to the function $X(\zeta)$ as $n \to \infty$ for all $\zeta \in \Omega$. ■

As a reminder, the functions $X(\zeta)$ are not arbitrary. They are random variables and thus satisfy the condition that the set $\{\zeta: X(\zeta) \leq x\} \subset \mathscr{F}$ for all x, that is, that this set be an event for all values of x. This is in fact necessary for the calculation of probability since the probability measure P is only defined for events. Such functions X are more generally called *measurable functions* and in a course on real analysis it is shown that the space of measurable functions is complete [6-1]. If we have a Cauchy sequence of measurable functions (random variables), then one can show that the limit function exists and is also measurable (a random variable). Thus the Cauchy convergence criterion also applies for random variables.

Most of the time we are not interested in precisely defining random variables for sets in Ω of probability zero because it is thought that these events will never occur. In this case, we can weaken the concept of sure convergence to the still very strong concept of almost-sure convergence.

Definition 6.7-4 (Almost-sure convergence.) The random sequence $X[n]$ converges *almost surely* to the random variable X if the sequence of functions $X[n, \zeta]$ converges for all $\zeta \in \Omega$ except possibly on a set of probability zero. ■

This is the strongest type of convergence normally used in probability theory. It is also called *probability-1* convergence. It is sometimes written

$$P\left\{\lim_{n\to\infty} X[n,\zeta] = X(\zeta)\right\} = 1,$$

meaning simply that there is a set A such that $P[A] = 1$ and $X[n]$ converges to X for all $\zeta \in A$. In particular $A \triangleq \{\zeta: \lim_{n\to\infty} X[n,\zeta] = X(\zeta)\}$. Here the set A^c is the probability-zero set mentioned in this definition. As shorthand notation we also use

$$X[n] \to X \text{ a.s.} \quad \text{and} \quad X[n] \to X \qquad \text{pr.1},$$

where the abbreviation "a.s." stands for *almost surely*, and "pr.1" stands for *probability one.*

An example of probability-1 convergence is the Strong Law of Large Numbers to be proved in the next section. Three examples of random sequences are next evaluated for possible convergence.

Example 6.7-2

(Convergence of random sequences.) For each of the following three random sequences, we assume that the probability space $(\Omega, \mathscr{F}, P)$ has sample space $\Omega = [0, 1]$. $\mathscr{F}$ is the family of Borel subsets of Ω and the probability measure P is Lebesgue measure, which on a real interval $(a, b]$, is just its length l, that is,

$$l(a,b] \triangleq b - a \quad \text{for } b \geq a.$$

(a) $X[n,\zeta] = n\zeta$.
(b) $X[n,\zeta] = \sin(n\zeta)$.
(c) $X[n,\zeta] = \exp[-n^2(\zeta - \frac{1}{n})]$.

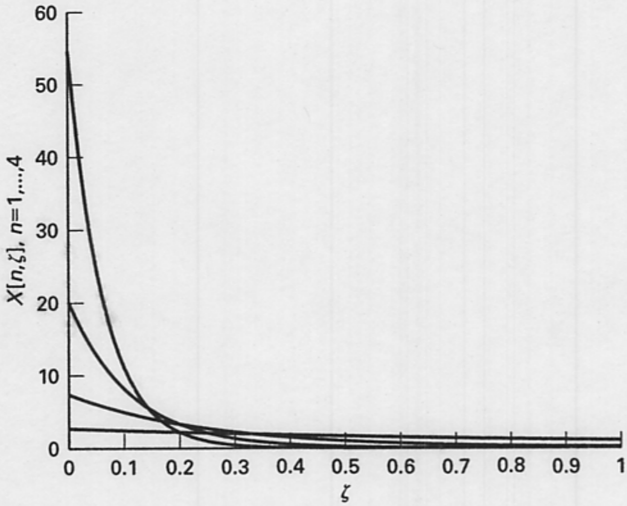

Figure 6.7-1 Plot of sequence (c) $X[n,\zeta]$ versus ζ for $\Omega = [0,1]$ for $n = 1, \ldots, 4$.

The sequence in (a) clearly diverges to $+\infty$ for any $\zeta \neq 0$. Thus this random sequence does not converge. The sequence in (b) does not diverge, but it oscillates between -1 and $+1$ except for the one point $\zeta = 0$. Thus this random sequence does not converge either.

Considering the random sequence in (c), the graph in Figure 6.7-1 shows that this sequence converges as follows:

$$\lim_{n\to\infty} X[n,\zeta] = \begin{cases} \infty \text{ for } \zeta = 0 \\ 0 \text{ for } \zeta > 0 \end{cases}.$$

Thus we can say that the random sequence converges to the (degenerate) random variable $X = 0$ with probability-1. We simply take $A = (0, 1]$ and note that $P[A] = 1$ and that $X[n,\zeta] \simeq 0$ for every ζ in A for sufficiently large n. We write $X[n] \to 0$ a.s. However $X[n]$ clearly does not converge surely to zero.

Thus far we have been discussing pointwise convergence of sequences of functions and random sequences. This is similar to considering a space of *bounded* functions $\mathscr{B}$ with the norm

$$||f||_\infty \triangleq \sup_x |f(x)|.^\dagger$$

When we write $f_n \to f$ in the function space $\mathscr{B}$, we mean that $||f_n - f||_\infty = \sup_x |f_n(x) - f(x)| \to 0$, giving us pointwise convergence. The space of continuous bounded functions is denoted L_∞ and is known to be complete ([6-1], p. 115).

†The supremum or sup operator is similar to the max operator. The supremum of a set of numbers is the smallest number greater than or equal to each number in the set, for example, $\sup\{0 < x < 1\} = 1$. Note the difficulty with max in this example since 1 is not included in the open interval $(0, 1)$, thus the max does not exist here!

Another type of function space of great practical interest uses the *energy norm* (cf. Eq. 4.4-6):

$$||f||_2 \triangleq \left(\int_{-\infty}^{+\infty} |f(x)|^2 dx \right)^{1/2}.$$

The space of integrable (measurable) functions with *finite energy norm* is denoted L^2. When we say a sequence of functions converges in L^2, that is, $||f_n - f||_2 \to 0$, we mean that

$$\left(\int_{-\infty}^{+\infty} |f_n(x) - f(x)|^2 dx \right)^{1/2} \to 0.$$

This space of integrable functions is also complete [6-1]. A corresponding concept for random sequences is given by mean-square convergence.

Definition 6.7-5 (Mean-square convergence.) A random sequence $X[n]$ converges in the *mean-square sense* to the random variable X if $E\{|X[n] - X|^2\} \to 0$ as $n \to \infty$. ■

This type of convergence depends only on the second-order properties of the random variables and is thus often easier to calculate than a.s. convergence. A second benefit of the mean-square type of convergence is that it is closely related to the physical concept of power. If $X[n]$ converges to X in the mean-square sense, then we can expect that the variance of the error $\varepsilon[n] \triangleq X[n] - X$ will be small for large n. If we look back at Example 6.7-2c, we can see that this random sequence does not converge in the mean-square sense, so that the error variance or power as defined here would not ever be expected to be small. To see this, consider possible mean-square convergence to zero (since $X[n] \to 0$ a.s.),

$$\begin{aligned} E\{|X[n] - 0|^2\} &= E\{X[n]^2\} \\ &= \int_0^1 \exp(-2n^2\zeta) \exp 2n d\zeta \\ &= \exp(2n) \int_0^1 \exp(-2n^2\zeta) d\zeta \\ &= \exp(2n) \left[\frac{1 - \exp(-2n^2)}{2n^2} \right] \to \infty \quad \text{as } n \to \infty, \end{aligned}$$

hence $X[n]$ does not converge in the mean-square sense to 0.

Still another type of convergence that we will consider is called *convergence in probability*. It is weaker than probability-1 convergence and also weaker than mean-square convergence. This is the type of convergence displayed in the Weak Laws of Large Numbers to be discussed in the next section. It is defined as follows:

Definition 6.7-6 (Convergence in probability.) Given the random sequence $X[n]$ and the limiting random variable X, we say that $X[n]$ converges in probability to X if for every $\varepsilon > 0$,

$$\lim_{n \to \infty} P[|X[n] - X| > \varepsilon] = 0. \quad ■$$

We sometimes write $X[n] \to X(p)$, where (p) denotes the type of convergence. Also convergence in probability is sometimes called p-convergence.

One can use Chebyshev's inequality (Theorem 4.4-1), $P[|Y| > \varepsilon] \le E[|Y|^2]/\varepsilon^2$ for $\varepsilon > 0$, to show that mean-square convergence implies convergence in probability. For example, let $Y \triangleq X[n] - X$; then the preceding inequality becomes

$$P[|X[n] - X| > \varepsilon] \le E\left[|X[n] - X|^2\right]/\varepsilon^2.$$

Now mean-square convergence implies that the right-hand side goes to zero as $n \to \infty$, for any fixed $\varepsilon > 0$, which implies that the left-hand side must also go to zero which is the definition of convergence in probability. Thus we have proved the following result.

Theorem 6.7-2 Convergence of a random sequence in the mean-square sense implies convergence in probability. ■

The relation between convergence with probability-1 and convergence in probability is more subtle. The main difference between them can be seen by noting that the former talks about the probability of the limit while the latter talks about the limit of the probability. Further insight can be gained by noting that a.s. convergence is concerned with convergence of the entire sample sequences while p-convergence is concerned only with the convergence of the random variable at an individual n. That is to say, a.s. convergence is concerned with the joint events at an infinite number of times, while p-convergence is concerned with the simple event at time n, albeit large. One can prove the following theorem.

Theorem 6.7-3 Convergence with probability-1 implies convergence in probability.

Proof (Adapted from Gnedenko [6-6].) Let $X[n] \to X$ a.s. and define the set A,

$$A \triangleq \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{m=1}^{\infty} \{\zeta : |X[n+m,\zeta] - X(\zeta)| < 1/k\}.$$

Then it must be that $P[A] = 1$. To see this we note that A is the set of ζ such that starting at some n and for all later n we have $|X[n,\zeta] - X(\zeta)| < 1/k$ and furthermore this must hold for all $k > 0$. Thus A is precisely the set of ζ on which $X[n,\zeta]$ is convergent. So $P[A]$ must be one. Eventually for n large enough and $1/k$ small enough we get $|X[n,\zeta] - X(\zeta)| < \varepsilon$, and the error stays this small for all larger n. Thus

$$P\left[\bigcup_{n=1}^{\infty} \bigcap_{m=1}^{\infty} \{|X[n+m] - X| < \varepsilon\}\right] = 1 \qquad \text{for all } \varepsilon > 0,$$

which implies by the continuity of probability,

$$\lim_{n\to\infty} P\left[\bigcap_{m=1}^{\infty} \{|X[n+m] - X| < \varepsilon\}\right] = 1 \qquad \text{for all } \varepsilon > 0,$$

which in turn implies the greatly weakened result

$$\lim_{n\to\infty} P[|X[n+m]-X|<\varepsilon]=1 \qquad \text{for all } \varepsilon>0, \tag{6.7-1}$$

which is equivalent to the definition of p-convergence. ■

Because of the gross weakening of the a.s. condition, that is, the enlargement of the set A in the foregoing proof, it can be seen that p-convergence does not imply a.s. convergence. We note in particular that Equation 6.7-1 may well be true even though no single sample sequence stays close to X for all $n+m>n$. This is in fact the key difference between these two types of convergence.

Example 6.7-3

Define a random pulse sequence $X[n]$ on $n\geq 0$ as follows: Set $X[0]=1$. Then for the *next two points* set exactly one of the $X[n]$'s to 1, the other to 0, equally likely. For the *next three points* set exactly one of the $X[n]$'s to 1 equally likely and set the others to 0. Continue this procedure for the *next four points*, setting exactly one of the $X[n]$'s to 1 equally likely, and so forth. A sample function would look like Figure 6.7-2.

Obviously this random sequence is slowly converging to zero in some sense as $n\to\infty$. In fact a simple calculation would show p-convergence and also m.s. convergence due to the growing distance between pulses as $n\to\infty$. In fact at $n\simeq\frac{1}{2}l^2$, the probability of a one (pulse) is only $1/l$. However, we do not have a.s. convergence, since *every* sample sequence has ones appearing arbitrarily far out on the n axis. Thus no sample sequences converge to zero.

One final type of convergence that we consider is not a convergence for random variables at all! Rather it is a type of convergence for distribution functions.

Definition 6.7-7 A random sequence $X[n]$ with probability distribution function $F_n(x)$ converges in distribution to the random variable X with probability distribution function $F(x)$ if

$$\lim_{n\to\infty} F_n(x)=F(x)$$

at all x for which F is continuous. ■

Note that in this definition we are not really saying anything about the random variables themselves, just their probability distribution functions (PDFs). Convergence in distribution

Figure 6.7-2 A sequence that is converging in probability but not with probability-1.

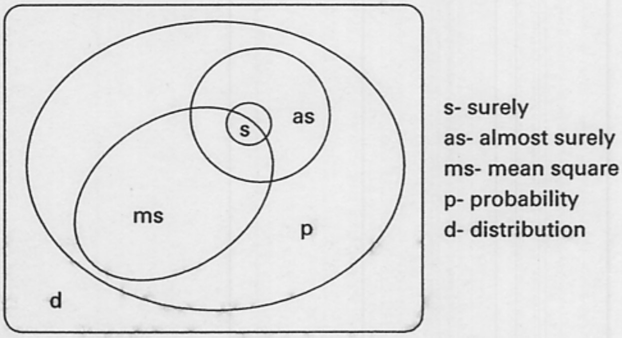

Figure 6.7-3 Venn diagram illustrating relationships of various possible convergence modes for random sequences.

just means that as n gets large the PDFs are converging or becoming alike. For example, the sequence $X[n]$ and the variable X can be jointly independent even though $X[n]$ converges to X in distribution. This is radically different from the four earlier types of convergence, where as n gets large the random variables $X[n]$ and X are becoming very dependent because some type of "error" between them is going to zero. Convergence in distribution is the type of convergence that occurs in the Central Limit Theorem (Section 4.7). The relationships between these five types of convergence are shown diagrammatically in Figure 6.7-3, where we have used the fact that p-convergence implies convergence in distribution, which is shown below. Note that even sure convergence may not imply mean square convergence. This because the integral of the square of the limiting random variable, with respect to the probability measure, may diverge.

To see that p-convergence implies convergence in distribution, assume that the limiting random variable X is continuous so that it has a pdf. First we consider the conditional distribution function

$$F_{X[n]|X}(y|x) = P\{X[n] \le y|X = x\}.$$

From the definition of p-convergence, it should be clear that

$$F_{X[n]|X}(y|x) \to \begin{cases} 1, & y > x \\ 0, & y < x \end{cases}, \qquad \text{as } n \to \infty,$$

so that

$$F_{X[n]|X}(y|x) \to u(y - x), \quad \text{except possibly at the one point } y = x,$$

and hence

$$\begin{aligned} F_{X[n]}(y) = P\{X[n] \le y\} &= \int_{-\infty}^{+\infty} F_{X[n]|X}(y|x) f_X(x) dx \\ &\to \int_{-\infty}^{+\infty} u(y-x) f_X(x) dx \end{aligned}$$

$$= \int_{-\infty}^{y} f_X(x)dx$$
$$= F_X(y),$$

as was to be shown. In the case where the limiting random variable X is not continuous, we must exercise more care but the result is still true at all points x for which $F_X(x)$ is continuous. (See Problem 6.43.)

6.8 LAWS OF LARGE NUMBERS

The Laws of Large Numbers have to do with the convergence of a sequence of estimates of the mean of a random variable. As such they concern the convergence of a random sequence to a constant. The Weak Laws obtain convergence in probability, while the Strong Laws yield convergence with probability-1. A version of the Weak Law has already been demonstrated in Example 4.4-3. We restate it here for convenience.

Theorem 6.8-1 (Weak Law of Large Numbers.) Let $X[n]$ be an independent random sequence with mean μ_X and variance σ_X^2 defined for $n \geq 1$. Define another random sequence as

$$\hat{\mu}_X[n] \triangleq (1/n)\sum_{k=1}^{n} X[k] \qquad \text{for } n \geq 1.$$

Then $\hat{\mu}_X[n] \to \mu_x \quad (p) \quad \text{as } n \to \infty.$ ■

Remember, an independent random sequence is one whose terms are all jointly independent. Another version of the Weak Law allows the random sequence to be of nonuniform variance.

Theorem 6.8-2 (Weak Law—nonuniform variance.) Let $X[n]$ be an independent random sequence with constant mean μ_X and variance $\sigma_X^2[n]$ defined for $n \geq 1$. Then if

$$\sum_{n=1}^{\infty} \sigma_X^2[n]/n^2 < \infty,$$

$$\hat{\mu}_X[n] \to \mu_X \quad (p) \quad \text{as } n \to \infty. \quad ■$$

Both of these theorems are also true for convergence with probability-1, in which case they become Strong Laws. The theorems concerning convergence with probability-1 are best derived using the concept of a Martingale sequence. By introducing this concept we can also get another useful result called the Martingale convergence theorem, which is helpful in estimation and decision/detection theory.

Definition 6.8-1 †A random sequence $X[n]$ defined for $n \geq 0$ is called a *Martingale* if the conditional expectation

$$E\{X[n]|X[n-1], X[n-2], \ldots, X[0]\} = X[n-1] \qquad \text{for all } n \geq 1. \quad ■$$

†The material dealing with Martingale sequences can be omitted on a first reading.