# 10

# Stochastic perturbation of discrete time systems

We have seen two ways in which uncertainty (and thus probability) may appear in the study of strictly deterministic systems. The first was the consequence of following a random distribution of initial states, which, in turn, led to a development of the notion of the Frobenius–Perron operator and an examination of its properties as a means of studying the asymptotic properties of flows of densities. The second resulted from the random application of a transformation $S$ to a system and led naturally to our study of the linear Boltzmann equation.

In this chapter we consider yet another source of probabilistic distributions in deterministic systems. Specifically, we examine discrete time situations in which at each time the value $x_{n+1} = S(x_n)$ is reached with some error. An extremely interesting situation occurs when this error is small and the system is "primarily" governed by a deterministic transformation $S$. We consider two possible ways in which this error might be small: Either the error occurs rather rarely and is thus small on the average, or the error occurs constantly but is small in magnitude. In both cases, we consider the situation in which the error is independent of $S(x_n)$ and are, thus, led to first recall the notion of independent random variables in the next section and to explore some of their properties in Sections 10.2 and 10.3.

## 10.1 Independent random variables

Let $(\Omega, \mathcal{F}, \text{prob})$ be a probability space. The random variables $\xi_1, \xi_2, \ldots, \xi_k, \xi_i$: $\Omega \to R$, are called **independent** if, for every sequence of Borel sets $B_1$, $B_2, \ldots, B_k$, the events

$$\{\xi_1 \in B_1\}, \{\xi_2 \in B_2\}, \ldots, \{\xi_k \in B_k\}$$

are independent. Thus, for every sequence of Borel sets $B_1, \ldots, B_k$, we have

$$\text{prob}\{\xi_1 \in B_1, \ldots, \xi_k \in B_k\} = \{\text{prob } \xi_1 \in B_1\} \cdots \text{prob}\{\xi_k \in B_k\}.$$

Thus, having a $k$-dimensional random vector $(\xi_1, \ldots, \xi_k)$, we may consider two different kinds of densities: the density of each random component $\xi_i$ and the joint density function for the random vector $(\xi_1, \ldots, \xi_k)$. Let the density of $\xi_i$ be denoted by $f_i(x)$, and the joint density of $\xi = (\xi_1, \ldots, \xi_k)$ be $f(x_1, \ldots, x_k)$. Then by definition, we have

$$\int_{B_i} f_i(x)\,dx = \text{prob}\{\xi_i \in B_i\}, \qquad \text{for } B_i \subset R, \tag{10.1.1}$$

and

$$\int \cdots \int_B f(x_1, \ldots, x_k)\,dx_1 \cdots dx_k = \text{prob}\{(\xi_1, \ldots, \xi_k) \in B\}, \qquad \text{for } B \subset R^k,$$

where $B_i$ and $B$ are Borel subsets of $R$ and $R^k$, respectively. In this last integral take

$$B = B_1 \times \underbrace{R \times \cdots \times R}_{k-1 \text{ times}}$$

so that we have

$$\begin{aligned} \text{prob}\{(\xi_1, \ldots, \xi_k) \in B\} &= \text{prob}\{\xi_1 \in B_1\} \\ &= \int_{B_1} \left\{ \int \cdots \int_{R^{k-1}} f(x, x_2, \ldots, x_k)\,dx_2 \cdots dx_k \right\} dx. \end{aligned} \tag{10.1.2}$$

By comparing (10.1.1) with (10.1.2), we see immediately that

$$f_1(x) = \int \cdots \int_{R^{k-1}} f(x, x_2, \ldots, x_k)\,dx_2 \cdots dx_k. \tag{10.1.3}$$

Thus, having the joint density function $f$ for $(\xi_1, \ldots, \xi_k)$, we can always find the density for $\xi_1$ from equation (10.1.3). In an entirely analogous fashion, $f_2$ can be obtained by integrating $f(x_1, x, \ldots, x_k)$ over $x_1, x_3, \ldots, x_k$. The same procedure will yield each of the densities $f_i$.

However, the converse is certainly not true in general since, having the density $f_i$ of each random variable $\xi_i (i = 1, \ldots, k)$, it is not usually possible to find the joint density $f$ of the random vector $(\xi_1, \ldots, \xi_k)$. The one important special case in which this construction is possible occurs when $\xi_1, \ldots, \xi_k$ are independent random variables. Thus, we have the following theorem.

***Theorem 10.1.1.*** If the random variables $\xi_1, \ldots, \xi_k$ are independent and have densities $f_1, \ldots, f_k$, respectively, then the joint density function for the random vector $(\xi_1, \ldots, \xi_k)$ is given by

$$f(x_1, \ldots, x_k) = f_1(x_1) \cdots f_k(x_k), \tag{10.1.4}$$

where the right-hand side is a product.

*Proof:* Consider a Borel set $B \subset R^k$ of the form

$$B = B_1 \times \cdots \times B_k, \tag{10.1.5}$$

where $B_1, \ldots, B_k \subset R$ are Borel sets. Then

$$\text{prob}\{(\xi_1, \ldots, \xi_k) \in B\} = \text{prob}\{\xi_1 \in B_1, \ldots, \xi_k \in B_k\},$$

and, since the random variables $\xi_1, \ldots, \xi_k$ are independent,

$$\text{prob}\{(\xi_1, \ldots, \xi_k) \in B\} = \text{prob}\{\xi_1 \in B_1\} \cdots \text{prob}\{\xi_k \in B_k\}.$$

With this equation and (10.1.1), we obtain

$$\begin{aligned}\text{prob}\{(\xi_1, \ldots, \xi_k) \in B\} &= \int_{B_1} f_1(x_1)\,dx_1 \cdots \int_{B_k} f_k(x_k)\,dx_k \\ &= \int \cdots \int_B f_1(x_1) \cdots f_k(x_k)\,dx_1 \cdots dx_k.\end{aligned} \tag{10.1.6}$$

Since, by definition, sets of the form (10.1.5) are generators of the Borel subsets in $R^k$, it is clear that (10.1.6) must hold for arbitrary Borel sets $B \subset R^k$. By the definition of the joint density, this implies that $f_1(x_1) \cdots f_k(x_k)$ is the joint density for the random vector $(\xi_1, \ldots, \xi_k)$. ■

As a simple application of Theorem 10.1.1, we consider two independent random variables $\xi_1$ and $\xi_2$ with densities $f_1$ and $f_2$, respectively. We wish to obtain the density of $\xi_1 + \xi_2$. Observe that, by Theorem 10.1.1, the random vector $(\xi_1, \xi_2)$ has the joint density $f_1(x_1)f_2(x_2)$. Thus, for an arbitrary Borel set $B \subset R$, we have

$$\text{prob}\{\xi_1 + \xi_2 \in B\} = \iint\limits_{x_1 + x_2 \in B} f_1(x_1)f_2(x_2)\,dx_1\,dx_2,$$

or, setting $x = x_1 + x_2$ and $y = x_2$,

$$\begin{aligned}\text{prob}\{\xi_1 + \xi_2 \in B\} &= \iint\limits_{B \times R} f_1(x - y)f_2(y)\,dx\,dy \\ &= \int_B \left\{\int_{-\infty}^{\infty} f_1(x - y)f_2(y)\,dy\right\} dx.\end{aligned}$$

From the definition of a density, this last equation shows that

$$f(x) = \int_{-\infty}^{\infty} f_1(x - y)f_2(y)\,dy \tag{10.1.7}$$

is the density of $\xi_1 + \xi_2$.

***Remark 10.1.1.*** From the definition of the density, it follows that, if $\xi$ has a density $f$, then $c\xi$ has a density $(1/|c|)f(x/c)$. To see this, write

$$\text{prob}\{c\xi \in A\} = \text{prob}\left\{\xi \in \frac{1}{c}A\right\} = \int_{(1/c)A} f(y)\,dy = \frac{1}{|c|}\int_A f\left(\frac{x}{c}\right)dx.$$

Thus, from (10.1.7), if $\xi_1$ and $\xi_2$ are independent and have densities $f_1$ and $f_2$, respectively, then $(c_1\xi_1 + c_2\xi_2)$ has the density

$$f(x) = \frac{1}{|c_1c_2|}\int_{-\infty}^{\infty} f_1\left(\frac{x-y}{c_1}\right)f_2\left(\frac{y}{c_2}\right)dy. \quad \square \tag{10.1.8}$$

## 10.2 Mathematical expectation and variance

In previous chapters we have, on numerous occasions, used the concept of mathematical expectation in rather specialized situations without specifically noting that it was, indeed, the mathematical expectation that was involved. We now wish to explicitly introduce this concept in its general sense.

Let $(\Omega, \mathcal{F}, \text{prob})$ be a probability space and let $\xi\colon \Omega \to R$ be a random variable. Then we have the following definition.

***Definition 10.2.1.*** If $\xi$ is integrable with respect to the measure "prob," then the **mathematical expectation** (or **mean value**) of $\xi$ is given by

$$E(\xi) = \int_\Omega \xi(\omega)\ \text{prob}(d\omega).$$

***Remark 10.2.1.*** By definition, $E(\xi)$ is the average value of $\xi$. A more illuminating interpretation of $E(\xi)$ is given by the law of large numbers [see equation (10.3.4)]. □

In the case when $\xi$ is a constant, $\xi = c$, then it is trivial to derive $E(c)$. Since $\text{prob}\{\Omega\} = 1$ for any constant $c$, we have

$$E(c) = c\int_\Omega \text{prob}(d\omega) = c. \tag{10.2.1}$$

Now we show how the mathematical expectation may be calculated via the use of a density function. Let $h\colon R^k \to R$ be a Borel measurable function, that is, $h^{-1}(\Delta)$ is a Borel subset of $R^k$ for each interval $\Delta$. Further, let $\xi = (\xi_1, \ldots, \xi_k)$ be a random vector with the joint density function $f(x_1, \ldots, x_k)$. Then we have the following theorem.

***Theorem 10.2.1.*** If $hf$ is integrable, that is,

$$\int \cdots \int_{R^k} |h(x_1 \cdots x_k)| f(x_1 \cdots x_k)\, dx_1 \cdots dx_k < \infty,$$

then the mathematical expectation of the random variable $h \circ \xi$ exists and is given by

$$E(h \circ \xi) = \int \cdots \int_{R^k} h(x_1, \ldots, x_k) f(x_1, \ldots, x_k)\, dx_1 \cdots dx_k. \qquad (10.2.2)$$

*Proof:* First assume that $h$ is a simple function, that is,

$$h(x) = \sum_{i=1}^{n} \lambda_i 1_{A_i}(x) \qquad x = (x_1, \ldots, x_k),$$

where the $A_i$ are mutually disjoint Borel subsets of $R^k$ such that $\cup_i A_i = R^k$. Then

$$h(\xi(\omega)) = \sum_{i=1}^{n} \lambda_i 1_{A_i}(\xi(\omega)) = \sum_{i=1}^{n} \lambda_i 1_{\xi^{-1}(A_i)}(\omega),$$

and, by the definition of the Lebesgue integral,

$$E(h \circ \xi) = \int_{\Omega} h(\xi(\omega))\, \mathrm{prob}(d\omega) = \sum_{i=1}^{n} \lambda_i\, \mathrm{prob}\{\xi^{-1}(A_i)\}.$$

Further, since $f$ is the density for $\xi$, we have

$$\mathrm{prob}\{\xi^{-1}(A_i)\} = \mathrm{prob}\{\xi \in A_i\} = \int_{A_i} f(x)\, dx, \qquad dx = dx_1 \cdots dx_k.$$

As a consequence,

$$E(h \circ \xi) = \sum_{i=1}^{n} \lambda_i \int_{A_i} f(x)\, dx = \int_{R^k} \sum_{i=1}^{n} \lambda_i 1_{A_i}(x) f(x)\, dx$$
$$= \int_{R^k} h(x) f(x)\, dx.$$

Thus, for the $h$ that are simple functions, equality (10.2.2) is proved. For an arbitrary $h$, $hf$ integrable, we can find a sequence $\{h_n\}$ of simple functions converging to $h$ and such that $|h_n| \leq |h|$. From equality (10.2.2), already proved for simple functions, we thus have

$$E(h_n \circ \xi) = \int_{R^k} h_n(x) f(x)\, dx.$$

By the Lebesgue dominated convergence theorem, since $|h_n f| \leq |h| f$, it follows that

$$\int_\Omega h(\xi(\omega))\ \mathrm{prob}(d\omega) = \int_{R^k} h(x)f(x)\,dx,$$

which completes the proof. ■

In the particular case that $k = 1$ and $h(x) = x$, we have from equation (10.2.2)

$$E(\xi) = \int_{-\infty}^{\infty} xf(x)\,dx. \tag{10.2.3}$$

Thus, if $f(x)$ is taken to be the mass density of a rod of infinite length, then $E(\xi)$ gives the center of mass of the rod.

From Definition 10.2.1, it follows that, for every sequence of random variables $\xi_1, \ldots, \xi_k$ and constants $\lambda_1, \ldots, \lambda_k$, we have

$$E\left(\sum_{i=1}^{k} \lambda_i \xi_i\right) = \sum_{i=1}^{k} \lambda_i E(\xi_i) \tag{10.2.4}$$

since the mathematical expectation is simply a Lebesgue integral on the probability space $(\Omega, \mathscr{F}, \mathrm{prob})$. Moreover, the mathematical expectation of $\sum \lambda_i \xi_i$ exists whenever all of the $E(\xi_i)$ exist.

We now turn to a consideration of the variance, starting with a definition.

***Definition 10.2.2.*** Let $\xi\colon \Omega \to R$ be a random variable such that $m = E(\xi)$ exists. Then the **variance** of $\xi$ is

$$D^2(\xi) = E((\xi - m)^2) \tag{10.2.5}$$

if the corresponding integral is finite.

Thus the variance of a random variable $\xi$ is just the average value of the square of the deviation of $\xi$ away from $m$. By the additivity of the mathematical expectation, equation (10.2.5) may also be written as

$$D^2(\xi) = E(\xi^2) - 2mE(\xi) + m^2 = E(\xi^2) - m^2. \tag{10.2.6}$$

If $\xi$ has a density $f(x)$, then by the use of equation (10.2.2), we can also write

$$D^2(\xi) = \int_{-\infty}^{\infty} (x - m)^2 f(x)\,dx$$

whenever the integral on the right-hand side exists. Finally, we note that for any constant $\lambda$,

$$D^2(\lambda\xi) = E(\lambda^2(\xi - m)^2) = \lambda^2 D^2(\xi).$$

Since in any application there is a certain inconvenience in the fact that $D^2(\xi)$ does not have the same dimension as $\xi$, it is sometimes more convenient to use the **standard deviation** of $\xi$, defined by

$$\sigma(\xi) = \sqrt{D^2(\xi)}\,.$$

For our purposes here, two of the most important properties of the mathematical expectation and variance of a random variable $\xi$ are contained in the next theorem.

***Theorem 10.2.2.*** Let $\xi_1, \ldots, \xi_k$ be independent random variables such that $E(\xi_i)$, $D^2(\xi_i)$, $i = 1, \ldots, k$ exist. Then

$$E(\xi_1 \cdots \xi_k) = E(\xi_1) \cdots E(\xi_k) \tag{10.2.7}$$

and

$$D^2(\xi_1 + \cdots + \xi_k) = D^2(\xi_1) + \cdots + D^2(\xi_k)\,. \tag{10.2.8}$$

*Proof:* The proof is easy even in the general case. However, to illustrate again the usefulness of (10.2.2), we will prove this theorem in the case when all the $\xi_i$ have densities. Thus, assume that $\xi_i$ has density $f_i$, $i = 1, \ldots, k$, and pick $h(x_1, \ldots, x_k) = x_1 \cdots x_k$. Since the $\xi_1, \ldots, \xi_k$ are independent random variables, by Theorem 10.1.1, the joint density function for the random vector $(\xi_1, \ldots, \xi_k)$ is

$$f_1(x_1) \cdots f_k(x_k)\,.$$

Hence, by equation (10.2.2),

$$\begin{aligned} E(\xi_1 \ldots \xi_k) &= \int \cdots \int_{R^k} x_1 \cdots x_k f_1(x_1) \cdots f_k(x_k)\, dx_1 \cdots dx_k \\ &= \int_{-\infty}^{\infty} x_1 f_1(x_1)\, dx_1 \cdots \int_{-\infty}^{\infty} x_k f_k(x_k)\, dx_k \\ &= E(\xi_1) \cdots E(\xi_k)\,, \end{aligned}$$

and (10.2.7) is therefore proved.

Now set $E(\xi_i) = m_i$, so that

$$\begin{aligned} D^2(\xi_1 + \cdots + \xi_k) &= E((\xi_1 + \cdots + \xi_k - m_1 - \cdots - m_k)^2) \\ &= E\left( \sum_{i,j=1}^{k} (\xi_i - m_i)(\xi_j - m_j) \right). \end{aligned}$$

Since the $\xi_1, \ldots, \xi_k$ are independent, $(\xi_1 - m_1), \ldots, (\xi_k - m_k)$ are also independent. Therefore, by (10.2.4) and (10.2.7), we have

$$D^2(\xi_1 + \cdots + \xi_k) = \sum_{i=1}^{k} E((\xi_i - m_i)^2) + \sum_{i \neq j} E((\xi_i - m_i)(\xi_j - m_j))$$
$$= \sum_{i=1}^{k} D^2(\xi_i) + \sum_{i \neq j} (E(\xi_i) - m_i)(E(\xi_j) - m_j)\,.$$

Since $E(\xi_i) = m_i$, equation (10.2.8) results immediately. ■

***Remark 10.2.2.*** In Theorem 10.2.2, it is sufficient to assume that the $\xi_i$ are mutually independent, that is, $\xi_i$ is independent of $\xi_j$, $i \neq j$. □

To close this section on mathematical expectation and variance, we give two versions of the Chebyshev inequality, originally introduced in a special context in Section 5.7.

***Theorem 10.2.3.*** If $\xi$ is nonnegative and $E(\xi)$ exists, then

$$\text{prob}\{\xi \geqslant a\} \leqslant E(\xi)/a \qquad \text{for every } a > 0\,. \tag{10.2.9}$$

If $\xi$ is arbitrary but such that $m = E(\xi)$ and $D^2(\xi)$ exist, then

$$\text{prob}\{|\xi - m| \geqslant \varepsilon\} \leqslant D^2(\xi)/\varepsilon^2 \qquad \text{for every } \varepsilon > 0\,. \tag{10.2.10}$$

*Proof:* By the definition of mathematical expectation,

$$E(\xi) = \int_\Omega \xi(\omega)\,\text{prob}(d\omega) \geqslant \int_{\{\omega:\xi(\omega)\geqslant a\}} \xi(\omega)\,\text{prob}(d\omega)$$
$$\geqslant a\int_{\{\omega:\xi(\omega)\geqslant a\}} \text{prob}(d\omega) = a\ \text{prob}\{\xi \geqslant a\}\,,$$

which proves (10.2.9). [This is, of course, analogous to equation (5.7.9).]

Now replace $\xi$ by $(\xi - m)^2$ and $a$ by $\varepsilon^2$ in (10.2.9) to give

$$\text{prob}\{(\xi - m)^2 \geqslant \varepsilon^2\} \leqslant (1/\varepsilon^2)E((\xi - m)^2) = (1/\varepsilon^2)D^2(\xi)\,,$$

which is equivalent to (10.2.10) and completes the proof. ■

## 10.3 Stochastic convergence

There are several different ways in which the convergence of a sequence $\{\xi_n\}$ of random variables may be defined. For example, if $\xi_n \in L^p(\Omega, \mathcal{F}, \text{prob})$, then we may define both strong and weak convergences of $\{\xi_n\}$ to $\xi$ in $L^p(\Omega)$ space, as treated in Section 2.3.

In probability theory some of these types of convergence have special names. Thus, strong convergence of $\{\xi_n\}$ in $L^2(\Omega)$, defined by the relation

$$\lim_{n\to\infty} \|\xi_n - \xi\|_{L^2(\Omega)} = 0\,, \tag{10.3.1}$$

is denoted by

$$\text{l.i.m. } \xi_n = \xi$$

and called **convergence in mean.**

A second type of convergence useful in the treatment of probabilistic phenomena is given in the following definition.

***Definition 10.3.1.*** A sequence $\{\xi_n\}$ of random variables is said to be **stochastically convergent** to the random variable $\xi$ if, for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \text{prob}\{|\xi_n - \xi| \geqslant \varepsilon\} = 0\,. \tag{10.3.2}$$

The stochastic convergence of $\{\xi_n\}$ to $\xi$ is denoted by

$$\text{st-lim } \xi_n = \xi. \tag{10.3.3}$$

Note that in terms of $L^p$ norms, the mathematical expectation and variance of a random variable may be written as

$$E(|\xi|^p) = \int_\Omega |\xi|^p \,\text{prob}(d\omega) = \|\xi\|^p_{L^p(\Omega)}$$

and

$$D^2(\xi) = \int_\Omega |\xi - m|^2 \,\text{prob}(d\omega) = \|\xi - m\|^2_{L^2(\Omega)}\,.$$

This observation allows us to derive a connection between stochastic convergence and strong convergence from the Chebyshev inequality, as contained in the following proposition.

***Proposition 10.3.1.*** If a sequence $\{\xi_n\}$ of random variables, $\xi_n \in L^p(\Omega)$, is strongly convergent in $L^p(\Omega)$ to $\xi$, then $\{\xi_n\}$ is stochastically convergent to $\xi$. Thus, in particular, convergence in mean implies stochastic convergence.

*Proof:* We only consider $p < \infty$, since for $p = \infty$ the proposition is trivial. Applying the Chebyshev inequality (10.2.9) to $|\xi_n - \xi|^p$, we have

$$\text{prob}\{|\xi_n - \xi|^p \geqslant \varepsilon^p\} \leqslant (1/\varepsilon^p)E(|\xi_n - \xi|^p)$$

or, equivalently,

$$\text{prob}\{|\xi_n - \xi| \geqslant \varepsilon\} \leqslant (1/\varepsilon^p)\|\xi_n - \xi\|^p_{L^p(\Omega)}\,,$$

which completes the proof. ■

A third type of convergence useful for random variables is defined next.

***Definition 10.3.2.*** A sequence $\{\xi_n\}$ of random variables is said to converge **almost surely** to $\xi$ (or to converge to $\xi$ with probability 1) if

$$\lim_{n\to\infty} \xi_n(\omega) = \xi(\omega)$$

for almost all $\omega$. Equivalently, this condition may be written as

$$\text{prob}\{\lim_{n\to\infty} \xi_n(\omega) = \xi(\omega)\} = 1\,.$$

***Remark 10.3.1.*** For all of the types of convergence we have defined (strong and weak $L^p$ convergence, convergence in mean, stochastic convergence, and almost sure convergence), the limiting function is determined up to a set of measure zero. That is, if $\xi$ and $\bar{\xi}$ are both limits of the sequence $\{\xi_n\}$, then $\xi$ and $\bar{\xi}$ differ only on a set of measure zero. □

We now show the connection between almost sure and stochastic convergence with the following proposition.

***Proposition 10.3.2.*** If a sequence of random variables $\{\xi_n\}$ converges almost surely to $\xi$, then it also converges stochastically to $\xi$.

*Proof:* Set

$$\eta_n(\omega) = \min(1, |\xi_n(\omega) - \xi(\omega)|)\,.$$

Clearly, $|\eta_n| \leqslant 1$. If $\{\xi_n\}$ converges almost surely to $\xi$, then $\{\eta_n\}$ converges to zero almost surely, and, by the Lebesgue dominated convergence theorem,

$$\lim_{n\to\infty} \|\eta_n\|_{L^1(\Omega)} = \lim_{n\to\infty} \int_\Omega \eta_n(\omega)\,\text{prob}(d\omega) = 0\,.$$

By Proposition 10.3.1 this implies that $\{\eta_n\}$ converges stochastically to zero. Since in the definition of stochastic convergence it suffices to consider only $\varepsilon < 1$, it then follows that

$$\text{prob}\{|\xi_n - \xi| \geqslant \varepsilon\} = \text{prob}\{\eta_n \geqslant \varepsilon\} \qquad \text{for } 0 < \varepsilon < 1\,.$$

Thus the stochastic convergence of $\{\eta_n\}$ to zero implies the stochastic convergence of $\{\xi_n\}$ to $\xi$ and the proof is complete. ■

As a simple illustration of the usefulness of the concept of stochastic convergence, we prove the simplest version of the law of large numbers given in the next theorem.

***Theorem 10.3.1 (Weak law of large numbers).*** Let $\{\xi_n\}$ be a sequence of independent random variables with

$$E(\xi_n) = m_n$$

and

$$M = \sup_n D^2(\xi_n) < \infty .$$

Then

$$\mathrm{prob}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}(\xi_i - m_i)\right| \geqslant \varepsilon\right\} < \frac{M}{n\varepsilon^2}$$

for every $\varepsilon > 0$. In particular, if $m_1 = m_2 = \cdots = m_n$, then

$$\text{st-lim}\, \frac{1}{n}\sum_{i=1}^{n}\xi_i = m . \tag{10.3.4}$$

*Proof:* Set

$$\eta_n = \frac{1}{n}\sum_{i=1}^{n}\xi_i .$$

Since the $\xi_i$ are independent random variables,

$$D^2(\eta_n) = D^2\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}D^2(\xi_i) \leqslant \frac{nM}{n^2} = \frac{M}{n}$$

and, clearly,

$$E(\eta_n) = \frac{1}{n}\sum_{i=1}^{n}m_i .$$

Thus, by the Chebyshev inequality (10.2.10),

$$\mathrm{prob}\{|\eta_n - E(\eta_n)| \geqslant \varepsilon\} \leqslant (1/\varepsilon^2)D^2(\eta_n) \leqslant M/n\varepsilon^2,$$

which completes the proof, as equation (10.3.4) is a trivial consequence. ∎

Equation (10.3.4) is a precise statement of our intuitive notion that the mathematical expectation or mean value of a random variable may be obtained by averaging the results of many independent experiments.

The term "weak law of large numbers" specifically refers to equation (10.3.4) because stochastic convergence is weaker than other types of convergence for which similar results can be proved. One of the most famous versions of these is the so-called **strong law of large numbers,** as contained in the Kolmogorov theorem.

***Theorem 10.3.2 (Kolmogorov).*** Let $\{\xi_n\}$ be a sequence of independent random variables with

$$E(\xi_n) = m_n \quad \text{and} \quad M = \sup_n D^2(\xi_n) < \infty .$$

Then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (\xi_i - m_i) = 0$$

with probability 1.

We will not give a proof of the Kolmogorov theorem [see Breiman (1968) for the proof] as it is not used in our studies of the flow of densities. We stated it only because of its close correspondence to the Birkhoff individual ergodic Theorem 4.2.4, which also deals with the pointwise convergence of averages. To illustrate this correspondence, consider the sequence

$$f(x), f(S(x)), \ldots, \tag{10.3.5}$$

which appears in the Birkhoff theorem, as a sequence of random variables on the probability space $(X, \mathcal{A}, \text{prob})$, where

$$\text{prob}(A) = \mu(A)/\mu(X).$$

These variables (10.3.5) are, in general, highly dependent since $S(x)$ is a function of $x$, $S^2(x) = S(S(x))$ is a function of $x$ and $S(x)$, and so on.

The reason that a probabilistic treatment of deterministic systems is often more difficult than problems in classical probability theory is directly related to the absence of independent random variables in the former. It is only in some special circumstances that independence may appear in deterministic systems under certain limiting cases, such as mixing and exactness.

## 10.4 Discrete time systems with randomly applied stochastic perturbations

In this section we consider the asymptotic behavior of a nonsingular transformation when a stochastic perturbation is randomly applied.

Let $(X, \mathcal{A}, \mu)$ be a measure space, and $S: X \to X$ a nonsingular transformation with associated Frobenius–Perron operator $P$. The following rules apply to the evolution of the point $x \in X$: At the $n$th instant of time we do not know the precise location of $x$ although we do know the density $f_n(x)$. At the next instant of time $(n + 1)$, the point moves with probability $(1 - \varepsilon)$ to the next location $S(x_n)$. However, there is a probability $\varepsilon$ that this new location will not be given by $S(x_n)$ but rather by a random variable, independent of $x_n$, with density $g(x)$. This process can be visualized as shown in Figure 10.4.1. To make it more precise, we follow the ideas of Chapter 8 in which we derived the linear Boltzmann equation. Thus, consider space $(X, \mathcal{A}, \mu_f)$, where $f$ is the density of the initial position of the point, and probability space $(\Omega, \mathcal{F}, \text{prob})$ related to the

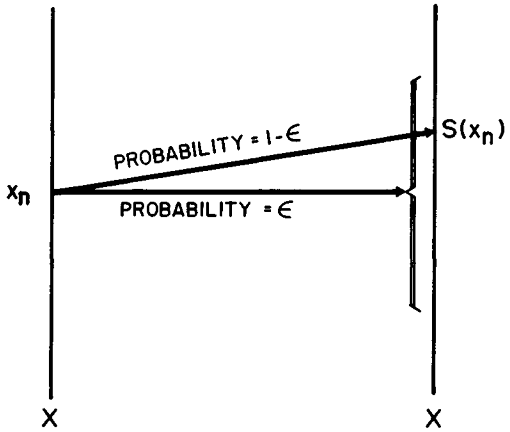


Figure 10.4.1. Schematic representation of the operation of a deterministic system with a randomly applied stochastic perturbation.

perturbations. With these two spaces we define the product space

$$\Omega \times X = \{(\omega, x) : \omega \in \Omega, x \in X\}$$

and the product measure

$$\text{Prob}(C \times A) = \text{prob}(C)\mu_f(A), \qquad \text{for } C \in \mathcal{F}, A \in \mathcal{A}$$

(see Theorem 2.2.2).

To describe the perturbations, consider a sequence of independent random variables

$$\xi_0, \eta_0, \xi_1, \eta_1, \ldots,$$

such that each $\eta_n: \Omega \to R$ takes only two values, 1 and 0, with the following probabilities:

$$\text{prob}(\eta_n = 1) = 1 - \varepsilon, \qquad \text{prob}(\eta_n = 0) = \varepsilon,$$

and each $\xi_n$ has the same density $g$. Then the equation

$$x_{n+1} = \eta_n S(x_n) + (1 - \eta_n)\xi_n, \qquad n = 0, 1, \ldots \tag{10.4.1}$$

gives the precise description of our intuitively introduced behavior of the sequence of $\{x_n\}$. Denote the density of $x_n$ by $f_n$. Our task now is to derive a relation between $f_n$ and $f_{n+1}$, assuming that the initial density $f_0 = f$ is given.

Note that

$$\text{Prob}\{x_{n+1} \in A\} = \text{Prob}\{x_{n+1} \in A \text{ and } \eta_n = 0\} + \text{Prob}\{x_{n+1} \in A \text{ and } \eta_n = 1\}. \tag{10.4.2}$$

Since, from (10.4.1), $x_{n+1}(\omega, x) = \xi_n(\omega)$ if $\eta_n(\omega) = 0$, and $x_{n+1}(\omega, x) = S(x_n)$ if $\eta_n(\omega) = 1$, we can rewrite equation (10.4.2) as

$$\begin{aligned}\operatorname{Prob}\{x_{n+1} \in A\} = \operatorname{Prob}\{&\xi_n \in A \text{ and } \eta_n = 0\}\\ &+ \operatorname{Prob}\{S(x_n) \in A \text{ and } \eta_n = 1\}.\end{aligned}$$

Since the events $\{\xi_n \in A\}$ and $\{\eta_n = 0\}$ are independent of each other and independent of the initial position $x$, we have

$$\begin{aligned}\operatorname{Prob}\{\xi_n \in A \text{ and } \eta_n = 0\} &= \operatorname{prob}\{\xi_n \in A \text{ and } \eta_n = 0\}\\ &= \operatorname{prob}\{\xi_n \in A\} \operatorname{prob}\{\eta_n = 0\}\\ &= \varepsilon \int_A g(x)\mu(dx).\end{aligned}$$

Further, since $x_n$ is dependent only on $\xi_1, \ldots, \xi_{n-1}$ and $\eta_1, \ldots, \eta_{n-1}$, we have

$$\operatorname{Prob}\{S(x_n) \in A \text{ and } \eta_n = 1\} = \operatorname{Prob}\{S(x_n) \in A\} \operatorname{prob}\{\eta_n = 1\}.$$

Finally, since $x_n$ has the density $f_n$ by assumption, this last formula implies

$$\operatorname{Prob}\{S(x_n) \in A \text{ and } \eta_n = 1\} = (1 - \varepsilon) \int_{S^{-1}(A)} f_n(x)\mu(dx).$$

Thus, combining the foregoing probabilities, we have

$$\operatorname{Prob}\{x_{n+1} \in A\} = (1 - \varepsilon) \int_{S^{-1}(A)} f_n(x)\mu(dx) + \varepsilon \int_A g(x)\mu(dx).$$

Using the definition of the Frobenius–Perron operator $P$ corresponding to $S$, this may be rewritten as

$$\operatorname{Prob}\{x_{n+1} \in A\} = \int_A [(1 - \varepsilon)Pf_n(x) + \varepsilon g(x)]\mu(dx)$$

for all Borel sets $\mathcal{A} \subset R$. Hence, if $x_n$ has density $f_n$, then this demonstrates that $x_{n+1}$ also has a density $f_{n+1}$ given by

$$f_{n+1} = (1 - \varepsilon)Pf_n + \varepsilon g. \tag{10.4.3}$$

We want to write the right-hand side of equation (10.4.3) in the form of a linear operator, and so we define $P_\varepsilon: L^1 \to L^1$ by

$$P_\varepsilon f = (1 - \varepsilon)Pf + \varepsilon g \int_X f(x)\mu(dx) \tag{10.4.4}$$

for all $f \in L^1$. Using the definition of $P_\varepsilon$, we may rewrite equation (10.4.3) in the form

$$f_{n+1} = P_\varepsilon f_n. \tag{10.4.5}$$

Our goal is to deduce as much as possible concerning the asymptotic behavior of $P_\varepsilon^n f_0$ for $f_0 \in D$.

The first result is contained in the following proposition.

***Proposition 10.4.1.*** Let the operator $P_\varepsilon: D \to D$ be defined by equation (10.4.4). Then $\{P_\varepsilon^n\}$ is asymptotically stable.

*Proof:* The proof is trivial. From the definition of $P_\varepsilon$ in (10.4.4), we have

$$P_\varepsilon^n f = P_\varepsilon(P_\varepsilon^{n-1} f) \geqslant \varepsilon g \int_X f(x)\mu(dx) = \varepsilon g$$

for all $f \in D$. Thus, $\varepsilon g$ is a nontrivial lower-bound function for $P_\varepsilon^n f$. Further since $P_\varepsilon$ is clearly a Markov operator, we have, by Theorem 5.6.2, that $\{P_\varepsilon^n\}$ is asymptotically stable. ■

***Remark 10.4.1.*** This simple result tells us that given any nonsingular transformation, the addition of even the smallest stochastic perturbation ensures that the system will be asymptotically stable regardless of the character of the deterministic system in the unperturbed case. □

However, much more can be determined about this stochastically perturbed deterministic system. We have the following result that explicitly gives the stationary density for $P_\varepsilon$.

***Proposition 10.4.2.*** Let the operator $P_\varepsilon: L^1 \to L^1$ be defined by equation (10.4.4). Then for $\varepsilon > 0$, the unique stationary density of $P_\varepsilon$ is given by

$$f_*^\varepsilon = \varepsilon \sum_{k=0}^{\infty} (1 - \varepsilon)^k P^k g. \tag{10.4.6}$$

*Proof:* Since $\|(1 - \varepsilon)^k P^k g\| \leqslant (1 - \varepsilon)^k \|g\|$ the series in (10.4.6) is absolutely convergent. Substitution of (10.4.6) into

$$P_\varepsilon f = (1 - \varepsilon)Pf + \varepsilon g \tag{10.4.7}$$

shows that $P_\varepsilon f_*^\varepsilon = f_*^\varepsilon$. ■

***Remark 10.4.2.*** It may happen that the limit of stationary densities $f_*^\varepsilon$ defined by equation (10.4.6) may not exist as $\varepsilon \to 0$. As a simple example consider $S: R^+ \to R^+$ given by $S(x) = \frac{1}{2}x$. In this case, the $k$-th iterate of the Frobenius–Perron operator is given by

$$P^k g(x) = 2^k g(2^k x)$$

and, thus,

$$f_*^\varepsilon(x) = \varepsilon \sum_{k=0}^{\infty} (1-\varepsilon)^k 2^k g(2^k x).$$

Now pick an arbitrarily small $h > 0$ and integrate $f_*^\varepsilon$ over $[0, h]$:

$$\int_0^h f_*^\varepsilon(x)\,dx = \varepsilon \sum_{k=0}^{\infty} 2^k (1-\varepsilon)^k \int_0^h g(2^k x)\,dx$$
$$= \varepsilon \sum_{k=0}^{\infty} (1-\varepsilon)^k \int_0^{h2^k} g(y)\,dy.$$

For $\delta > 0$ arbitrarily small, we can always find an $m$ such that, for all $k > m$,

$$\int_0^{h2^k} g(y)\,dy \geqslant \int_0^{\infty} g(y)\,dy - \delta = 1 - \delta$$

so

$$\int_0^h f_*^\varepsilon(x)\,dx \geqslant (1-\delta)\varepsilon \sum_{k=m}^{\infty} (1-\varepsilon)^k = (1-\delta)(1-\varepsilon)^m.$$

Thus, holding $\delta$ and $m$ fixed, assume $f_*^0(x) = \lim_{\varepsilon \to 0} f_*^\varepsilon(x)$ so that

$$\int_0^h f_*^0(x)\,dx \geqslant 1 - \delta \qquad \text{for every } h > 0.$$

Now it follows directly that $f_*^0 \in D$ cannot exist, for, if it did, then

$$\lim_{h \to 0} \int_0^h f_*^0(x)\,dx = 0,$$

which is a contradiction. □

***Theorem 10.4.1.*** Let $(X, \mathcal{A}, \mu)$ be a measure space, $S: X \to X$ a nonsingular transformation, $P$ the Frobenius–Perron operator corresponding to $S$, and $P_\varepsilon$ the operator defined by equation (10.4.4) with unique stationary density $f_*^\varepsilon$ given by (10.4.6). If the limit (strong or weak in $L^1$)

$$f_*^0(x) = \lim_{\varepsilon \to 0} f_*^\varepsilon(x)$$

exists, then $f_*^0$ is a stationary density for $P$.

*Proof:* Since $f_*^\varepsilon$ is a stationary density for $P_\varepsilon$, we have $P_\varepsilon f_*^\varepsilon = f_*^\varepsilon$ or, more explicitly,

$$(1-\varepsilon)Pf_*^\varepsilon + \varepsilon g = f_*^\varepsilon.$$

Under the assumption that $f_*^0$ exists, we immediately have $Pf_*^0 = f_*^0$, finishing the proof. ■

***Remark 10.4.3.*** In this context it is interesting to note that if $f_*^0$ exists it may depend on $g$. A simple example comes from $S: R \to R$ given by $S(x) = x$. Then $Pg = g$ for all $g \in D$ and

$$f_*^\varepsilon = g\varepsilon \sum_{k=0}^{\infty} (1 - \varepsilon)^k = g$$

so $f_*^0 = g$. □

Although this example shows that it is quite possible for $f_*^0$ to depend on $g$, the following theorem gives sufficient conditions for not only the existence of $f_*^0$ but also its value.

***Theorem 10.4.2.*** Let $(X, \mathcal{A}, \mu)$ be a finite measure space, $S: X \to X$ a measure-preserving ergodic transformation, and $P_\varepsilon$ the operator defined by equation (10.4.4) with the unique stationary density $f_*^\varepsilon$ given by (10.4.6). Then $f_*^0 = \lim_{\varepsilon\to 0} f_*^\varepsilon$ exists and is given by

$$f_*^0 = 1/\mu(X).$$

Although the proof of this theorem is straightforward, we will not give it in detail. It suffices to note that the proof is very similar to those of Sections 5.2 and 8.7, and the point of similarity resides in the fact that the series representations for $P^n f$ and $\hat{P}_t f$ in each of those sections have coefficients that sum to 1. Exactly the same situation occurs in the explicit representation (10.4.4) for $f_*^\varepsilon$ since

$$\varepsilon \sum_{k=0}^{\infty} (1 - \varepsilon)^k = 1.$$

## 10.5 Discrete time systems with constantly applied stochastic perturbations

In Section 10.4 we examined the asymptotic behavior of deterministic systems with a randomly applied stochastic perturbation. We now turn our attention to deterministic systems with constantly applied stochastic perturbations. Such dynamical systems have been considered by Kifer [1974] and Boyarsky [1984], and in a physical context by Feigenbaum and Hasslacher [1982].

Specifically consider the process defined by

$$x_{n+1} = S(x_n) + \xi_n, \tag{10.5.1}$$

where $S: R^d \to R^d$ is a measurable, though not necessarily nonsingular, transformation and $\xi_0, \xi_1, \ldots$ are independent random variables each having the same density $g$. We let the density of $x_n$ be denoted by $f_n$, and desire a relation connecting $f_{n+1}$ and $f_n$.

Assume $f_n \in D$. By (10.5.1), $x_{n+1}$ is the sum of two independent random variables: $S(x_n)$ and $\xi_n$. Note that $S(x_n)$ and $\xi_n$ are clearly independent since, in calculating $x_1, \ldots, x_n$, we only need $\xi_0, \ldots, \xi_{n-1}$. Let $h: R^d \to R$ be an arbitrary, bounded, measurable function. It is easy to find the mathematical expectation of $h(x_{n+1})$ since, by Theorem 10.2.1,

$$E(h(x_{n+1})) = \int_{R^d} h(x) f_{n+1}(x)\, dx. \tag{10.5.2}$$

Furthermore, because of (10.5.1) and the fact that the joint density of $(x_n, \xi_n)$ is just $f_n(y)g(z)$, we also have

$$\begin{aligned} E(h(x_{n+1})) &= E(h(S(x_n) + \xi_n)) \\ &= \int_{R^d}\int_{R^d} h(S(y) + z) f_n(y) g(z)\, dy\, dz. \end{aligned}$$

By a change of variables, this can be rewritten as

$$E(h(x_{n+1})) = \int_{R^d}\int_{R^d} h(x) f_n(y) g(x - S(y))\, dx\, dy. \tag{10.5.3}$$

Equating (10.5.2) and (10.5.3), and using the fact that $h$ was an arbitrary, bounded, measurable function, we immediately obtain

$$f_{n+1}(x) = \int_{R^d} f_n(y) g(x - S(y))\, dy. \tag{10.5.4}$$

***Remark 10.5.1.*** Our derivation of (10.5.4), though mathematically precise, is somewhat different from the usual method. Our reasons for this are threefold. First we were able to avoid the introduction of the concept of conditional probabilities. Second, the technique provides a clear proof that if $f_n(x)$ exists then $f_{n+1}(x)$ must also exist. To see this, take $h(x) = 1_A(x)$ in (10.5.3), so (10.5.3) becomes

$$\operatorname{prob}\{x_{n+1} \in A\} = \int_A \int_{R^d} f_n(y) g(x - S(y))\, dx\, dy$$

and, thus, by the definition of density, if $f_n$ exists then $f_{n+1}$ also exists and is given by (10.5.4). Finally, we have introduced this method of obtaining (10.5.4) because we use it later in deriving the Fokker–Planck equation that describes the evolution of densities for continuous time systems in the presence of a stochastic process. □

From our equation (10.5.4), we define an operator $\overline{P}: L^1 \to L^1$ by

$$\overline{P}f(x) = \int_{R^d} f(y)g(x - S(y))\,dy \tag{10.5.5}$$

for $f \in L^1$. In fact, $\overline{P}$ is a Markov operator that is quite easy to prove. Note first that if we set $K(x,y) = g(x - S(y))$, then, for $g \in D$, $K$ is a stochastic kernel (Section 5.7) and $\overline{P}$ is a Markov operator. Thus, in examining the behavior of the systems forming the subject of this section, we have available all of the tools developed in Section 5.7.

***Remark 10.5.2.*** In the special case in which $d = 1$ and $S = \lambda x$, equation (10.5.5) reduces to that considered in Example 5.7.2 with $a = 1$ and $b = -\lambda$. □

However, because of the characteristics of the function $g$ identified as a kernel, we can prove more than in Section 5.7. We start by stating and proving a result for the existence of a stationary density for $\overline{P}$.

***Theorem 10.5.1.*** Let the operator $\overline{P}: L^1 \to L^1$ be defined by (10.5.5) and let $g \in D$. If there exists a constant $M$ such that $g(x) \leq M < \infty$ for all $x \in R^d$ and a Liapunov function [see (5.7.8)] $V: R^d \to R$ such that, with $\alpha < 1$,

$$\int_{R^d} g(x - S(y))V(x)\,dx \leq \alpha V(y) + \beta, \qquad \text{for all } y \in R^d$$

then a stationary density for $\overline{P}$ exists.

*Proof:* Pick an arbitrary $f \in D$ such that

$$E_0(V\,|\,f) = \int_{R^d} V(x)f(x)\,dx < \infty,$$

and set

$$E_n(V\,|\,f) = \int_{R^d} V(x)\overline{P}^n f(x)\,dx.$$

Since $\overline{P}$ is a Markov operator defined by a kernel, we have, from the proof of Theorem 5.7.1, that

$$E_n(V\,|\,f) \leq [\beta/(1 - \alpha)] + \alpha^n E_0(V\,|\,f)$$

or, for sufficiently large $n$,

$$E_n(V\,|\,f) \leq [\beta/(1 - \alpha)] + 1\,.$$

Take the set $G_a$ to be

$$G_a = \{x \in R^d: V(x) < a\}$$

so that, by the Chebyshev inequality,

$$\int_{G_a} \overline{P}^n f(x)\,dx \geqslant \int_{R^d} \overline{P}^n f(x)\,dx - \frac{E_n(V \mid f)}{a}$$

and, thus,

$$\int_{R^d \setminus G_a} \overline{P}^n f(x)\,dx \leqslant \frac{E_n(V \mid f)}{a} \leqslant \frac{1}{a}\left[\frac{\beta}{1-\alpha} + 1\right].$$

We now use criterion 3 of Section 5.1 to show that $\{\overline{P}^n f\}$ is weakly precompact. Pick an $\varepsilon > 0$ and choose the constant $a$ such that

$$\frac{1}{a}\left[\frac{\beta}{1-\alpha} + 1\right] < \varepsilon.$$

Since $G_a$ is bounded this shows that condition (c) for weak precompactness (see Remarks 5.1.3 and 5.1.4) is satisfied.

Now let $\delta = \varepsilon/M$. Pick a set $A \subset R^d$ such that $\mu(A) < \delta$, hence

$$\begin{aligned}\int_A \overline{P}^n f(x)\,dx &= \int_A dx \int_{R^d} g(x - S(y))\overline{P}^{n-1} f(y)\,dy \\ &\leqslant M \int_A dx \int_{R^d} P^{n-1} f(y)\,dy \\ &= M\mu(A) \leqslant \delta M = \varepsilon,\end{aligned}$$

which proves condition (b) for weak precompactness. Finally, since $\|P^n f\| = 1$ for all $n$, condition (a) is satisfied. It is clear that, if $\{\overline{P}^n f\}$ is weakly precompact and

$$A_n f = \frac{1}{n}\sum_{k=0}^{n-1} \overline{P}^k f,$$

then $\{A_n f\}$ is also weakly precompact. Thus, by Theorem 5.2.1, we conclude that there is a stationary density of $\overline{P}$ and the proof is finished. ■

We now turn to a theorem concerning the uniqueness of stationary densities for $\overline{P}$.

***Theorem 10.5.2.*** Let the operator $\overline{P}: L^1 \to L^1$ be defined by equation (10.5.5) and let $g \in D$. If $g(y) > 0$ for all $y \in R^d$ and if a stationary density $f_*$ for $\overline{P}$ exists, then $f_*$ is unique.

*Proof:* Assume there are two stationary densities for $\overline{P}$, namely, $f_1$ and $f_2$. Set $f = f_1 - f_2$, so we clearly have

$$\overline{P}f = f. \tag{10.5.6}$$

We may write $f = f^+ - f^-$ by definition, so that, if $f_1 \neq f_2$, then neither $f^+$ nor $f^-$ are zero. Since $\overline{P}f^+ = f^+$ (by Proposition 3.1.3), from (10.5.5) we have

$$\overline{P}f^+(x) = \int_{R^d} g(x - S(y))f^+(y)\,dy \tag{10.5.7}$$

and similarly for $\overline{P}f^-$. Since $f^+$ is not identically zero and $g$ is strictly positive, the integral in (10.5.7) is a nonzero function for every $x$ and, thus, $\overline{P}f^+(x) > 0$ for all $x$. Clearly, too, $\overline{P}f^-(x) > 0$ for all $x$ and, thus, the supports of $\overline{P}f^+$ and $\overline{P}f^-$ are not disjoint. By Proposition 3.1.2, then, we must have $\|\overline{P}f\| < \|f\|$, which contradicts equality (10.5.6). Thus, $f_1$ and $f_2$ must be identical almost everywhere if they exist. ■

***Remark 10.5.3.*** It certainly may happen that there is no solution to $\overline{P}f = f$ in $D$. As a simple example, consider $S(x) = x$ for all $x \in R$. Take $g$ to be the Gaussian density

$$g(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2),$$

so the operator $\overline{P}$ defined in (10.5.5) becomes

$$\overline{P}f(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp[-(x - y)^2/2]f(y)\,dy.$$

Note that $\overline{P}f(x)$ is simply the solution $u(t,x)$ of the heat equation (7.4.13) with $\sigma^2 = 1$ at time $t = 1$, assuming an initial condition $u(0,y) = f(y)$. Since this solution is given by a semigroup of operators [cf. equations (7.4.11) and (7.9.9)], it can be shown that

$$\begin{aligned}\overline{P}^n f(x) &= \frac{1}{\sqrt{2\pi n}} \int_{-\infty}^{\infty} \exp[-(x - y)^2/2n]f(y)\,dy \\ &\leqslant \frac{1}{\sqrt{2\pi n}} \int_{-\infty}^{\infty} f(y)\,dy = \frac{1}{\sqrt{2\pi n}}.\end{aligned}$$

Thus $\overline{P}^n f$ converges uniformly to zero as $n \to \infty$ for all $f \in D$, and there is no solution to $\overline{P}f = f$. □

If these conditions for the existence and uniqueness of stationary densities of $\overline{P}$ are strengthened somewhat, we can prove that $\{\overline{P}^n\}$ is asymptotically stable. In fact from our results of Theorem 5.7.1, we have the following corollary.

***Corollary 10.5.1.*** Let the operator $\overline{P}: L^1 \to L^1$ be defined by equation (10.5.5) and let $g \in D$. If there is a Liapunov function $V: R^d \to R$ such that

$$\int_{R^d} g(x - S(y))V(x)\,dx \leqslant \alpha V(y) + \beta \qquad \text{for all } y \in R^d, \tag{10.5.8}$$

for some nonnegative constants $\alpha, \beta, \alpha < 1$, and

$$\int_{R^d} \inf_{|y| \leqslant r} g(x - S(y))\,dx > 0 \tag{10.5.9}$$

for every $r > 0$, then $\{\overline{P}^n\}$ is asymptotically stable.

***Remark 10.5.4.*** Note that condition (10.5.9) is automatically satisfied if $g: R^d \to R$ is positive and continuous and $S: R^d \to R^d$ is continuous because

$$\inf_{|y| \leqslant r} g(x - S(y)) = \min_{|y| \leqslant r} g(x - S(y)) > 0$$

for every $x \in R^d$. □

***Example 10.5.1.*** Consider a point moving through $R^d$ whose trajectory is determined by

$$x_{n+1} = S(x_n) + \xi_n\,,$$

where $S: R^d \to R^d$ is continuous and satisfies

$$|S(x)| \leqslant \lambda |x|, \qquad \text{for } |x| \geqslant M, \tag{10.5.10}$$

where $\lambda < 1$ and $M > 0$ are given constants. Assume that $\xi_0, \xi_1, \ldots$ are independent random variables with the same density $g$, which is continuous and positive, and such that $E(\xi_n)$ exists. Then $\{\overline{P}^n\}$ defined by (10.5.5) is asymptotically stable.

To show this, it is enough to confirm that condition (10.5.8) is satisfied. Set $V(x) = |x|$, so

$$\begin{aligned}\int_{R^d} g(x - S(y))V(x)\,dx &= \int_{R^d} g(x - S(y))\,|x|\,dx \\ &= \int_{R^d} g(x)\,|x + S(y)|\,dx \\ &\leqslant \int_{R^d} g(x)\,(|x| + |S(y)|)\,dx. \\ &= |S(y)| + \int_{R^d} g(x)\,|x|\,dx.\end{aligned}$$

From (10.5.10) we also have

$$|S(y)| \leqslant \lambda|y| + \max_{|x|\leqslant M} |S(x)|$$

so that

$$\int_{R^d} g(x - S(y))V(x)\,dx \leqslant \lambda|y| + \max_{|x|\leqslant M} |S(x)| + \int_{R^d} g(x)\,|x|\,dx.$$

Thus, since $E(\xi_n)$ exists, equation (10.5.8) is satisfied with $\alpha = \lambda$ and

$$\beta = \int_{R^d} g(x)\,|x|\,dx + \max_{|x|\leqslant M} |S(x)|. \quad \square$$

It is important to note that throughout it has not been necessary to require that $S$ be a nonsingular transformation. Indeed, one of the goals of this section was to demonstrate that the addition of random perturbations to a singular transformation may lead to interesting results.

However, if $S$ is nonsingular, then the Frobenius–Perron operator $P$ corresponding to $S$ exists and allows us to rewrite (10.5.5) in an alternate form that will be of use in the following section. By definition,

$$\overline{P}f(x) = \int_{R^d} g(x - S(y))f(y)\,dy.$$

Assume $S$ is nonsingular, therefore the Frobenius–Perron and Koopman operators corresponding to $S$ exist. Let $h_x(y) = g(x - y)$, so we can write $\overline{P}f$ as

$$\overline{P}f(x) = \int_{R^d} h_x(S(y))f(y)\,dy = \langle f, Uh_x\rangle = \langle Pf, h_x\rangle,$$

or, more explicitly,

$$\overline{P}f(x) = \int_{R^d} g(x - y)Pf(y)\,dy. \tag{10.5.11}$$

By a change of variables, (10.5.11) may also be written as

$$\overline{P}f(x) = \int_{R^d} g(y)Pf(x - y)\,dy. \tag{10.5.12}$$

***Remark 10.5.5.*** Observe that for $d = 1$, equations (10.5.11) and (10.5.12) could also be obtained as an immediate consequence of equation (10.1.7) applied to equation (10.5.1) since $\xi_n$ and $S(x_n)$ are independent. $\square$

## 10.6 Small continuous stochastic perturbations of discrete time systems

This section examines the behavior of the system

$$x_{n+1} = S(x_n) + \varepsilon\xi_n, \qquad \varepsilon > 0, \tag{10.6.1}$$

where $S: R^d \to R^d$ is measurable and nonsingular. As in the preceding section, we assume the $\xi_n$ to be independent random variables each having the same density $g$.

Since the variables $\varepsilon\xi_n$ have the density $(1/\varepsilon)g(x/\varepsilon)$, see Remark 10.1.1, equation (10.5.12) takes the form

$$P_\varepsilon f(x) = \frac{1}{\varepsilon}\int_{R^d} g\left(\frac{y}{\varepsilon}\right) Pf(x - y)\,dy \tag{10.6.2}$$

and gives the recursive relation

$$f_{n+1} = P_\varepsilon f_n \tag{10.6.3}$$

that connects successive densities $f_n$ of $x_n$.

The operator $P_\varepsilon$ can also be written, via a change of variables, as

$$P_\varepsilon f(x) = \int_{R^d} g(y) Pf(x - \varepsilon y)\,dy. \tag{10.6.4}$$

Since

$$\int_{R^d} g(y) Pf(x)\,dy = Pf(x),$$

we should expect that in some sense $\lim_{\varepsilon\to 0} P_\varepsilon f(x) = Pf(x)$. To make this more precise, we state the following theorem.

***Theorem 10.6.1.*** For the system defined by equation (10.6.1)

$$\lim_{\varepsilon\to 0}\|P_\varepsilon f - Pf\| = 0, \qquad \text{for all } f \in L^1,$$

where $P$ is the Frobenius–Perron operator corresponding to $S$ and $P_\varepsilon$ is given by (10.6.4).

*Proof:* Since $P$ and $P_\varepsilon$ are linear we may restrict ourselves to $f \in D$. Write

$$Pf(x) = \int_{R^d} g(y) Pf(x)\,dy,$$

then

$$P_\varepsilon f(x) - Pf(x) = \int_{R^d} g(y)\,[Pf(x - \varepsilon y) - Pf(x)]\,dy.$$

Pick an arbitrarily small $\delta > 0$. Since $g$ and $Pf$ are both integrable functions on $R^d$, there must exist an $r > 0$ such that

$$\int_{|y|\geqslant r} g(y)\,dy \leqslant \frac{\delta}{4} \quad \text{and} \quad \int_{|x|\geqslant r/2} Pf(x)\,dx \leqslant \frac{\delta}{4}.$$

To calculate the norm of $P_\varepsilon f - Pf$,

$$\|P_\varepsilon f - Pf\| \leqslant \int_{R^d}\int_{R^d} g(y)\,|Pf(x - \varepsilon y) - Pf(x)|\,dx\,dy,$$

we split the integral into three parts,

$$\|P_\varepsilon f - Pf\| \leqslant I_1 + I_2 + I_3,$$

where

$$I_1 = \int_{|x|\leqslant r}\int_{|y|\leqslant r} g(y)\,|Pf(x - \varepsilon y) - Pf(x)|\,dx\,dy,$$

$$I_2 = \int_{R^d}\int_{|y|\geqslant r} g(y)\,|Pf(x - \varepsilon y) - Pf(x)|\,dx\,dy,$$

and

$$I_3 = \int_{|x|\geqslant r}\int_{|y|\leqslant r} g(y)\,|Pf(x - \varepsilon y) - Pf(x)|\,dx\,dy.$$

We consider each in turn.

With respect to $I_1$, note that, since the function $Pf$ is integrable, by Corollary 5.1.1, we may assume

$$\int_{|x|\leqslant r} |Pf(x - \varepsilon y) - Pf(x)|\,dx \leqslant \frac{\delta}{2}$$

for $|y| \leqslant r$ and $\varepsilon \leqslant \varepsilon_0$ with $\varepsilon_0$ sufficiently small. Hence

$$I_1 \leqslant \frac{\delta}{2}\int_{|y|\leqslant r} g(y)\,dy \leqslant \frac{\delta}{2}\int_{R^d} g(y)\,dy = \frac{\delta}{2}.$$

In examining $I_2$, we use the triangle inequality to write

$$I_2 \leqslant \int_{R^d}\int_{|y|\geqslant r} g(y)Pf(x - \varepsilon y)\,dx\,dy + \int_{R^d}\int_{|y|\geqslant r} g(y)Pf(x)\,dx\,dy.$$

Change the variables in the first integral to $v = y$ and $z = x - \varepsilon y$, then

$$\int_{R^d}\int_{|y|\geq r} g(y)Pf(x - \varepsilon y)\,dx\,dy = \int_{R^d}\int_{|v|\geq r} g(v)Pf(z)\,dz\,dv$$
$$= \int_{|v|\geq r} g(v)\,dv \leq \frac{\delta}{4}.$$

Further, we also have

$$\int_{R^d}\int_{|y|\geq r} g(y)Pf(x)\,dx\,dy \leq \frac{\delta}{4}$$

so that $I_2 \leq \delta/2$.

A similar sequence of operations with $I_3$ gives

$$I_3 \leq \int_{|x|\geq r}\int_{|y|\leq r} g(y)Pf(x - \varepsilon y)\,dx\,dy + \int_{|x|\geq r}\int_{|y|\leq r} g(y)Pf(x)\,dx\,dy$$
$$\leq \int_{|z+\varepsilon v|\geq r}\int_{|v|\leq r} g(v)Pf(z)\,dz\,dv + \int_{|x|\geq r}\int_{|y|\leq r} g(y)Pf(x)\,dx\,dy.$$

For $\varepsilon \leq \frac{1}{2}$ the inequalities $|z + \varepsilon v| \geq r$ and $|v| \leq r$ imply $|z| \geq \frac{1}{2}r$. So

$$I_3 \leq \int_{|z|\geq r/2}\int_{|v|\leq r} g(v)Pf(z)\,dz\,dv + \int_{|x|\geq r}\int_{|y|\leq r} g(y)Pf(x)\,dx\,dy$$
$$\leq \frac{\delta}{4} + \frac{\delta}{4} = \frac{\delta}{2}.$$

Thus

$$\|P_\varepsilon f - Pf\| \leq \tfrac{3}{2}\delta \qquad \text{for any } \varepsilon \leq \min(\tfrac{1}{2}, \varepsilon_0),$$

that is,

$$\lim_{\varepsilon\to 0}\|P_\varepsilon f - Pf\| = 0. \quad \blacksquare$$

As an immediate consequence of Theorem 10.6.1 we have the following corollary.

***Corollary 10.6.1.*** Suppose that $S$ and $g$ are given and that for every small $\varepsilon$, $0 < \varepsilon < \varepsilon_0$, the operator $P_\varepsilon$, defined by (10.6.4), has a stationary density $f_\varepsilon$. If the limit

$$f_* = \lim_{\varepsilon\to 0} f_\varepsilon$$

exists, then $f_*$ is a stationary density for the Frobenius–Perron operator corresponding to $S$.

*Proof:* Write

$$P_\varepsilon f_* = f_\varepsilon + P_\varepsilon(f_* - f_\varepsilon) .$$

Since $P_\varepsilon$ is contractive,

$$\|P_\varepsilon(f_* - f_\varepsilon)\| \leqslant \|f_* - f_\varepsilon\| .$$

Thus $f_\varepsilon + P_\varepsilon(f_* - f_\varepsilon) \to f_*$ as $\varepsilon \to 0$ and, as a consequence, $P_\varepsilon f_* \to f_*$. However, Theorem 10.6.1 also tells us that $P_\varepsilon f_* \to P f_*$, so $P f_* = f_*$. ■