the probability of invariant events. A measure-preserving transformation T is ergodic if either $P(E) = 1$ or $P(E) = 0$ for all $E \in \mathcal{I}$, where $\mathcal{I}$ is the σ-field of invariant events under T. A stationary sequence $\{X_t(\omega)\}_{-\infty}^{+\infty}$ is said to be ergodic if $X_t(\omega) = X_1(T^{t-1}\omega)$ for every t where T is measure-preserving and ergodic. Some authors, such as Doob, use the term *metrically transitive* for ergodic. Events that are invariant under ergodic transformations either occur almost surely, or do not occur almost surely. In the case of **13.9**, Z must be a constant almost surely.

Intuitively, stationarity and ergodicity together are seen to be sufficient conditions for time averages and ensemble averages to converge to the same limit. Stationarity implies that, for example, $\mu = E(X_1(\omega))$ is the mean not just of X_1 but of any member of the sequence. The existence of events that are invariant under the shift transformation means that there are regions of the sample space which a particular realization of the sequence will never visit. If $P(TE\,\Delta\,E) = 0$, then the event E^c occurs with probability 0 in a realization where E occurs. However, if invariant events other than the trivial ones are ruled out, we ensure that a sequence will eventually visit all parts of the space, with probability 1. In this case time averaging and ensemble averaging are effectively equivalent operations.

The following corollary is the main reason for our interest in Theorem **13.10**.

13.12 Ergodic theorem Let $\{X_t(\omega)\}_1^\infty$ be a stationary, ergodic, integrable sequence. Then

$$\lim_{n\to\infty} S_n(\omega)/n = E(X_1),\ \text{a.s.} \tag{13.27}$$

Proof This is immediate from **13.10**, since by ergodicity, $E(X_1|\mathcal{I}) = E(X_1)$ a.s. ■

In an ergodic sequence, conditioning on events of probability zero or one is a trivial operation almost surely, in that the information contained in $\mathcal{I}$ is trivial. The ergodic theorem is an example of a law of large numbers, the first of several such theorems to be studied in later chapters. Unlike most of the subsequent examples this one is for stationary sequences. Its practical applications in econometrics are limited by the fact that the stationarity assumption is often inappropriate, but it is of much theoretical interest, because ergodicity is a very mild constraint on the dependence, as we now show.

A transformation that is measure-preserving eventually mixes up the outcomes in a non-invariant event A with those in A^c. The measure-preserving property rules out mapping sets into proper subsets of themselves, so we can be sure that $TA \cap A^c$ is nonempty. Repeated iterations of the transformation generate a sequence of sets $\{T^kA\}$ containing different mixtures of the elements of A and A^c. A positive dependence of B on A implies a negative dependence of B on A^c; that is, if $P(A \cap B) > P(A)P(B)$ then $P(A^c \cap B) = P(B) - P(A \cap B) < P(B) - P(A)P(B) = P(A^c)P(B)$. Intuition suggests that the average dependence of B on mixtures of A and A^c should tend to zero as the mixing-up proceeds. In fact, ergodicity can be characterized in just this kind of way, as the following theorem shows.