3/7 Ch 5 [H]

Note Title 2015-03-05

Convergence of a sequence of numbers (Defn. 5.1)

$\{a_n : n = 1, 2, \ldots\}$ converges to b as $n \to \infty$, written

$a_n \to b$, as $n \to \infty$, or equivalently,

$\lim_{n \to \infty} a_n = b$ if

$\forall \varepsilon > 0$, $\exists n_0(\varepsilon)$, such that $\forall n > n_0(\varepsilon)$, we have

$$|a_n - b| < \varepsilon$$

Convergence almost surely (i.e., with probability 1) (Defn. 5.2)

Definition 5.2 The sequence of random variables $\{Y_n : n = 1, 2, \ldots\}$ ***converges almost surely*** to μ, written

$$Y_n \xrightarrow{a.s.} \mu, \text{ as } n \to \infty$$

or equivalently, the sequence ***converges with probability 1***, written

$$Y_n \longrightarrow \mu, \text{ as } n \to \infty \text{ w.p. } 1$$

if

$$\forall k > 0, \mathrm{P}\left\{\lim_{i \to \infty} |Y_n - \mu| > k\right\} = 0.$$

The $\mathbf{P}\{\ldots\}$ in the previous expression is over the set of sample paths. More precisely we might write

$$\forall k > 0, \mathbf{P}\left\{\omega : \lim_{n \to \infty} |Y_n(\omega) - \mu| > k\right\} = 0$$

I.e., sample paths that do not converge to their mean become very rare.

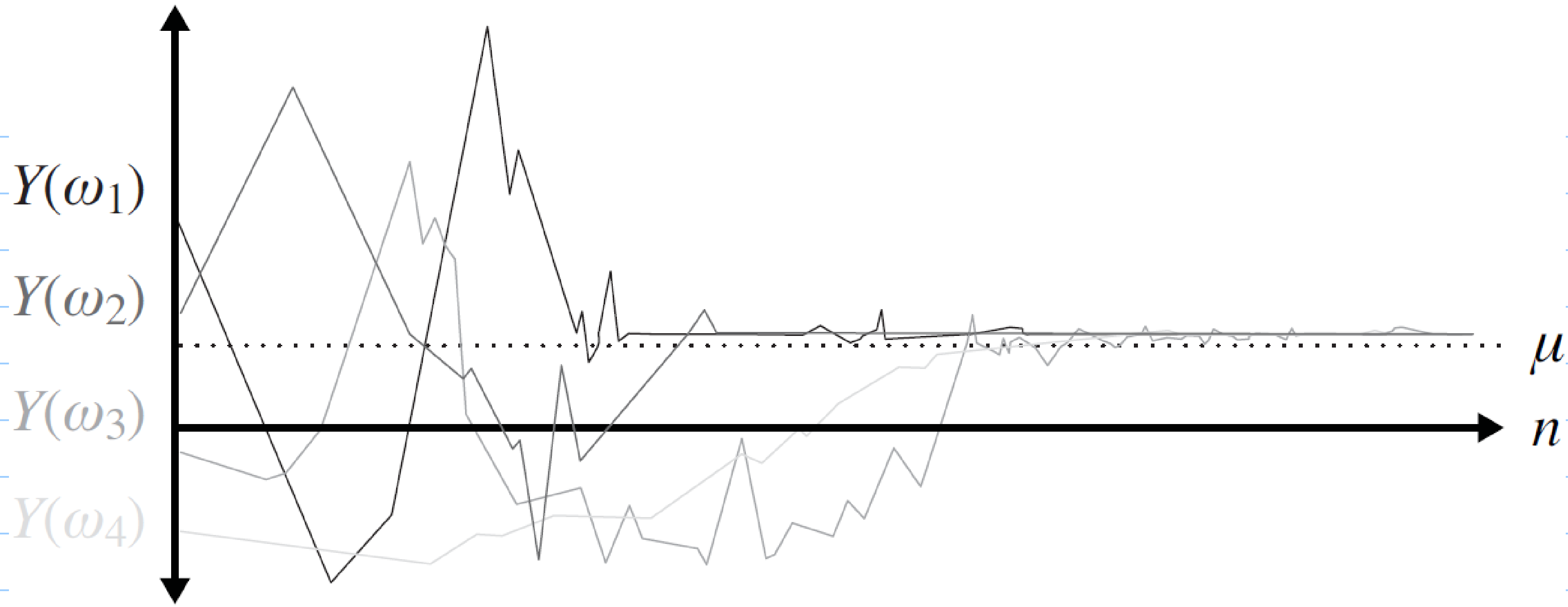


Figure 5.1 [H]. All four sample paths shown, after some point, behave well; the sequence of constants created by evaluating $\{Y_n \mid n=1,2,\ldots\}$ on that sample path converges to μ.

Convergence in probability

Definition 5.3 The sequence of random variables $\{Y_n : n = 1, 2, \ldots\}$ ***converges in probability*** to μ, written

$$Y_n \xrightarrow{P} \mu, \text{ as } n \to \infty$$

if

$$\forall k > 0, \lim_{n \to \infty} \mathbf{P}\{|Y_n - \mu| > k\} = 0.$$

The $\mathbf{P}\{\ldots\}$ in Definition 5.3 is over the set of possible sample paths, ω. More precisely we might write

$$\forall k > 0, \lim_{n \to \infty} \mathbf{P}\{\omega : |Y_n(\omega) - \mu| > k\} = 0 \quad (5.1)$$

It may be no sample path converges in the limit, but the values that are different from the mean on each sample path get rarer and rarer, so that their total mass goes to zero

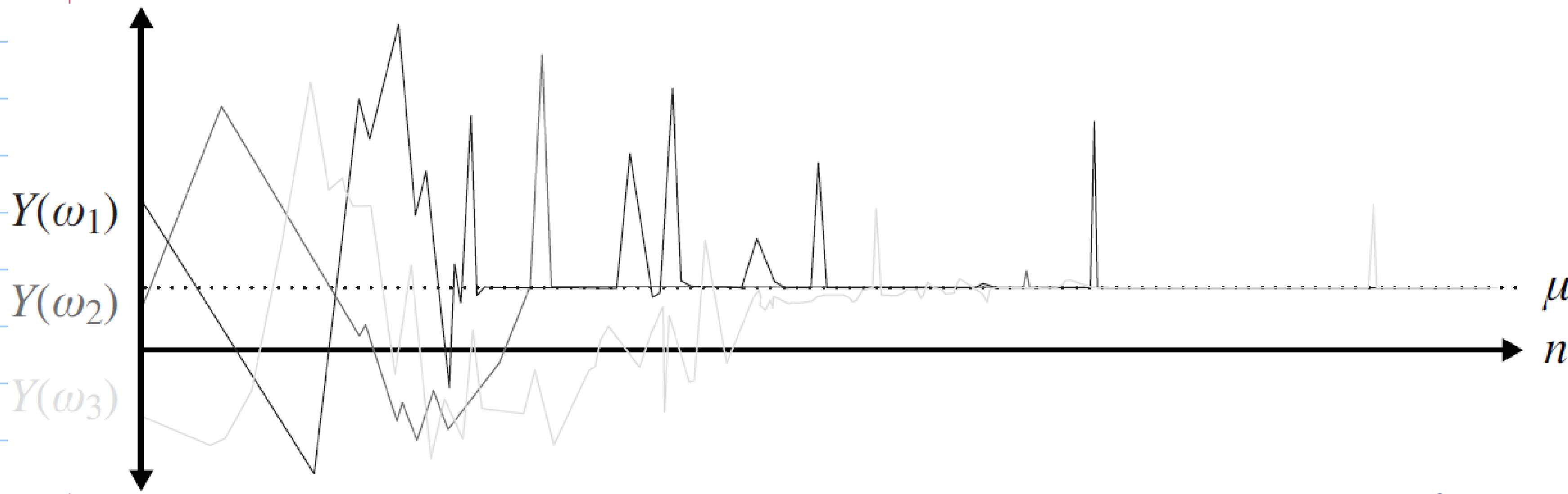


Figure 5.2 Illustration of convergence in probability without almost sure convergence.

Example of convergence in probability from wikepedia

http://en.wikipedia.org/wiki/Convergence_of_random_variables

"Archer"

Suppose a person takes a bow and starts shooting arrows at a target. Let Xn be his score in n-th shot. Initially he will be very likely to score zeros, but as the time goes and his archery skill increases, he will become more and more likely to hit the bullseye and score 10 points. After the years of practice the probability that he hit anything but 10 will be getting increasingly smaller and smaller and will converge to 0. Thus, the sequence Xn converges in probability to X = 10.

Note that Xn does not converge almost surely however. No matter how professional the archer becomes, there will always be a small probability of making an error. Thus the sequence {Xn} will never turn stationary: there will always be non-perfect scores in it, even if they are becoming increasingly less frequent.

Examples of almost sure convergence (equivalently, convergence with probability 1) from Wikipedia

http://en.wikipedia.org/wiki/Convergence_of_random_variables

Consider an animal of some short-lived species. We record the amount of food that this animal consumes per day. This sequence of numbers will be unpredictable, but we may be quite certain that one day the number will become zero, and will stay zero forever after.

Consider a man who tosses seven coins every morning. Each afternoon, he donates one pound to a charity for each head that appeared. The first time the result is all tails, however, he will stop permanently.

Let X1, X2, … be the daily amounts the charity received from him.

We may be almost sure that one day this amount will be zero, and stay zero forever after that.

However, when we consider any finite number of days, there is a nonzero probability the terminating condition will not occur.

Almost sure convergence implies convergence in probability.

Weak Law of Large Numbers

Theorem 5.4 (Weak Law of Large Numbers) *Let $X_1, X_2, X_3, \ldots$ be i.i.d. random variables with mean $\mathbf{E}[X]$. Let*

$$S_n = \sum_{i=1}^{n} X_i \quad \text{and} \quad Y_n = \frac{S_n}{n}.$$

Then

$$Y_n \xrightarrow{P} \mathbf{E}[X], \text{ as } n \to \infty.$$

This is read as "Y_n converges in probability to $\mathbf{E}[X]$," which is shorthand for the following:

$$\forall k > 0, \lim_{n \to \infty} \mathbf{P}\{|Y_n - \mathbf{E}[X]| > k\} = 0.$$

Comments on Exercise 5.1. (Prove WLLNs)

<u>Markov's inequality.</u> Do example 4.15 [Trivedi]

If X is non-negative then

$$P\{X \geq t\} \leq \frac{E[X]}{t}, \quad \forall t > 0$$

Discrete case.

$$E[X] = \sum_x x P\{X = x\} = \sum_x x\, p_x = \sum_x x P(x) = \sum_{x<t} x P(x) + \sum_{x \geq t} x P(x) \geq$$

$$\geq \sum_{x \geq t} x P(x) \geq \sum_{x \geq t} t P(x) = t \sum_{x \geq t} P(x) = t P\{X \geq t\}, \text{ so}$$

$$P\{X \geq t\} \leq \frac{E[X]}{t}$$

Chebychev's inequality

Do Example 4.16 [Trivedi]

$$P\{|X - \underbrace{E[X]}_{\mu}| \geq t\} \leq \frac{\sigma_X^2}{t^2}$$

Complete proof of WLLNs.

Strong Law of Large Numbers

Theorem 5.5 (Strong Law of Large Numbers) *Let $X_1, X_2, X_3, \ldots$ be i.i.d random variables with mean $\mathbf{E}[X]$. Let*

$$S_n = \sum_{i=1}^{n} X_i \quad \text{and} \quad Y_n = \frac{S_n}{n}.$$

Then

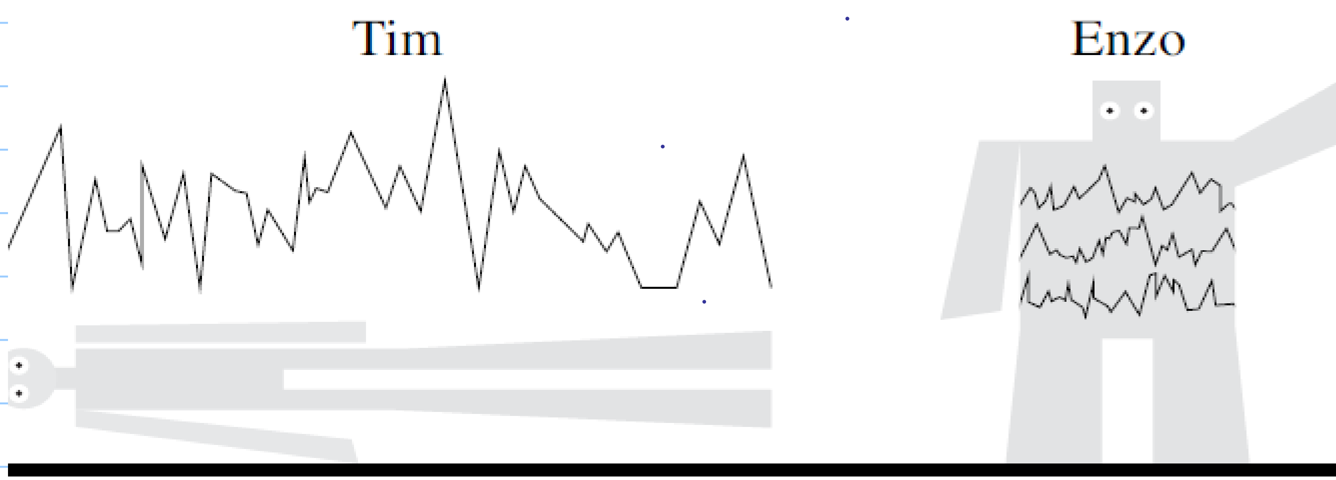
$$Y_n \xrightarrow{a.s.} \mathbf{E}[X], \text{ as } n \to \infty.$$

This is read as "Y_n converges almost surely to $\mathbf{E}[X]$" or "Y_n converges to $\mathbf{E}[X]$ with probability 1," which is shorthand for the following:

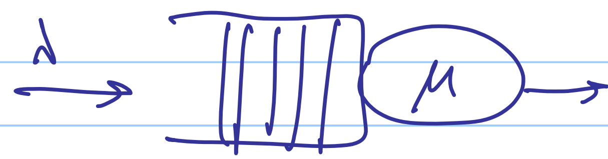
$$\forall k > 0, \mathbf{P}\left\{\lim_{n\to\infty} |Y_n - \mathbf{E}[X]| \geq k\right\} = 0.$$

Time Average vs. Ensemble Average

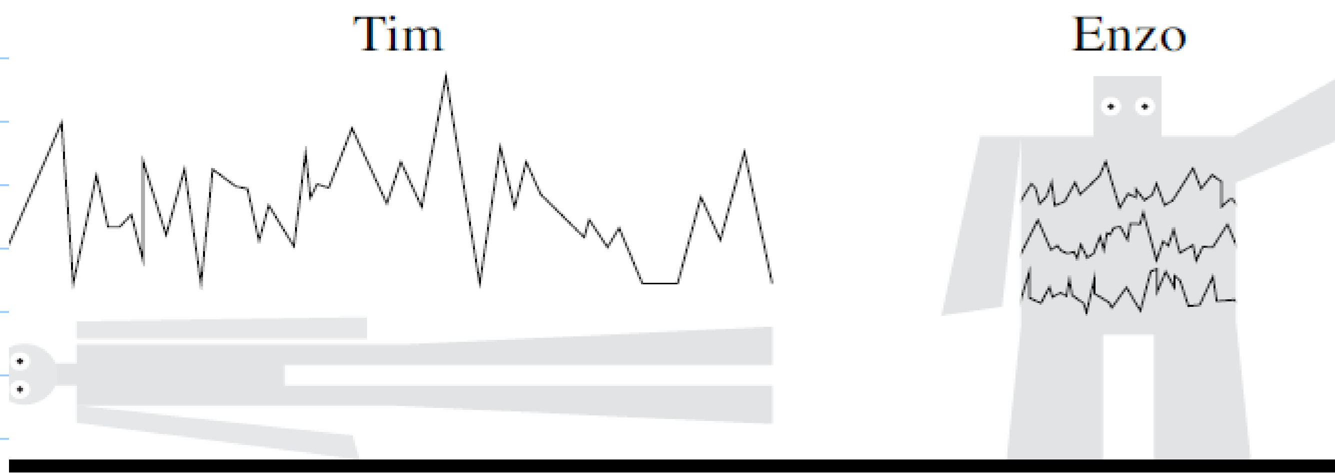
Tim & Enzo



Consider the following example: a FCFS queue in which each second a new job is added to the queue with probability p, and each second with probability q the job in service (if there is one) is completed, $q > p$.



Let $N(v)$ be the number of jobs in the system at time v,



Tim generates one very long sequence of coin flips (a single process) that he uses to simulate the queue. He would log the number of jobs at each second, sum them up, then divide the sum by the total length (in seconds) of the simulation. He would call this "the average number of jobs."

Enzo generates many (say, 1000) simulations, each of length 1,000.

For <u>each simulation</u>, Enzo would sample the queue at time t, obtaining one value $N(t)$. He would then sum the $N(t)$ values thus obtained, and divide them by the number of $N(t)$ values (1,000).
Enzo would call this "the average number of jobs".

Who is right? Tim or Enzo?

Definition 5.6

$$\overline{N}^{\text{Time Avg}} = \lim_{t\to\infty} \frac{\int_0^t N(v)dv}{t}.$$

(Tim's)

Definition 5.7

$$\overline{N}^{\text{Ensemble}} = \lim_{t\to\infty} \mathbf{E}[N(t)] = \sum_{i=0}^{\infty} i p_i$$

(Enzo's)

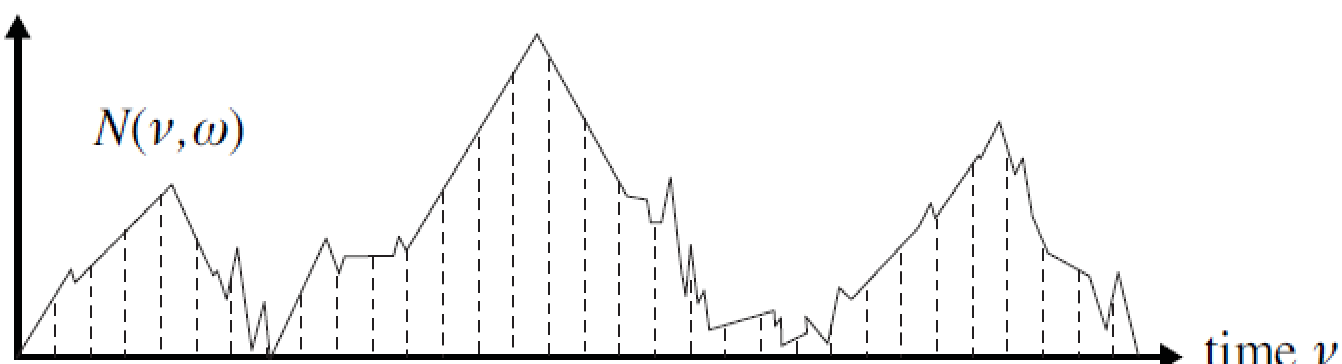
where

$$\begin{aligned} p_i &= \lim_{t\to\infty} \mathbf{P}\{N(t) = i\} \\ &= \text{mass of sample paths with value } i \text{ at time } t. \end{aligned}$$

Definition 5.8 (restatement of 5.6)

$$\overline{N}^{\text{Time Avg}}(\omega) = \lim_{t\to\infty} \frac{\int_0^t N(v,\omega)\,dv}{t},$$

where $N(v,\omega)$ represents the number of jobs in the system at time v under sample path ω.



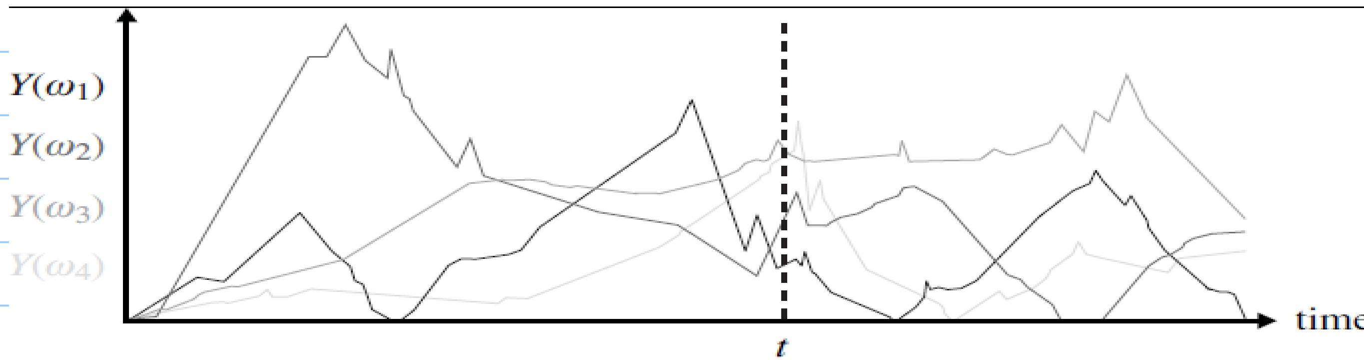
Time average (Fig. 5.5)

Example $N(0,\omega)=0$, $N(1,\omega)=1$, $N(2,\omega)=2$, $N(3,\omega)=3$, $N(\cancel{4},\omega)=2$ (no arrival, one departure), etc.

By time 4, the time average for this process (sample path, ω) is $(0+1+2+3+2)/5 = 8/5$.

We want t to be very large; we take the limit.

Note that we consider a single sequence. This seems suspicious!



Ensemble average (Fig. 5.6)

Here, we consider many sequences and look at the number of jobs in the system when the initial conditions do not matter any more (at "steady state").

Equivalence of averages & ergodicity

Theorem 5.9 *For an "ergodic" system (see Definition 5.10), the ensemble average exists and, with probability 1,*

$$\overline{N}^{\text{Time Avg}} = \overline{N}^{\text{Ensemble}}.$$

That is, for (almost) all sample paths, the time average along that sample path converges to the ensemble average.

Definition 5.10 An ***ergodic*** system is one that is positive recurrent, aperiodic, and irreducible.

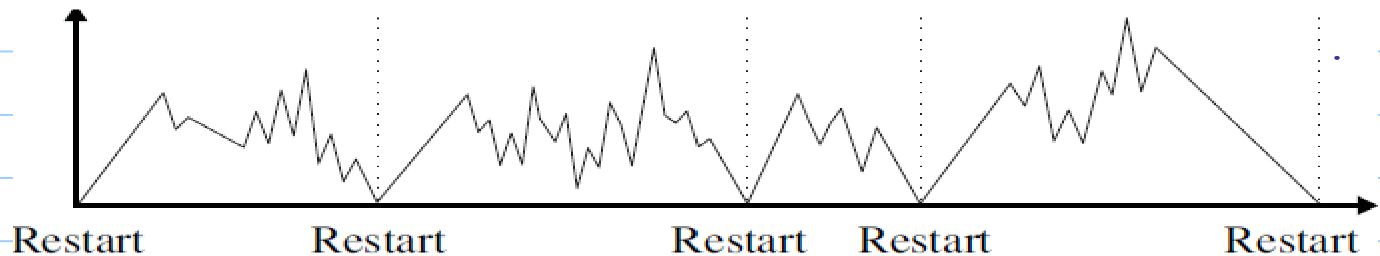


Fig. 5.7 Single Process Restarting Itself

- Simulation

If both methods give the same result, which one is more convenient?

- Average Time In System

$$\overline{T}^{\text{TimeAvg}} = \lim_{t\to\infty} \frac{\sum_{i=1}^{A(t)} T_i}{A(t)},$$ where T_i is the time in system of the ith arrival and $A(t)$ is the number of arrivals by time t. The Time Average is (assumed to be) associated with a single path.

$\overline{T}^{\text{Ensemble}} = \lim_{i \to \infty} E[T_i]$, where $E[T_i]$ is the time in system of the i-th job, and the average is taken over all sample paths.