Lecture 10:
Limit Theorems-cont'd
and
Introduction to Random Processes

ELE 525: Random Processes in Information Systems

Hisashi Kobayashi

Department of Electrical Engineering
Princeton University
October 14, 2013

Textbook: Hisashi Kobayashi, Brian L. Mark and William Turin, ***Probability, Random Processes and Statistical Analysis*** (Cambridge University Press, 2012)

Now let us consider an infinite sequence of events denoted as $E_1, E_2, \ldots, E_k, \ldots$. We are interested in finding how many of the E_n's occur. Let A_n represent the event that at least one of $E_n, E_{n+1}, E_{n+2}, \ldots$ occurs:

$$A_n = \bigcup_{k=n}^{\infty} E_k. \tag{11.54}$$

Then $\{A_n\}$ is a decreasing sequence.

Let A represent the event "Infinitely many of events $E_1, E_2, \ldots$ occur." A occurs if and only if A_n occurs for every n, This is because

1. If an infinite number of the E_k's occur, then A_n occurs for each n, thus $\cap_{n=1}^{\infty} A_n$ occurs.
2. Conversely, if $\cap_{n=1}^{\infty} A_n$ occurs, then A_n occurs for each n. Thus, for each n at least one of the events $E_k; k \geq n$ occurs, hence an infinitely number of the E_k's occur.

Thus,

$$A = \bigcap_{n=1}^{\infty} A_n = \bigcap_{n=1}^{\infty} \left(\bigcup_{k=n}^{\infty} E_k \right), \tag{11.55}$$

For the event A thus defined, we have

$$P[A] = 0 \iff \text{With probability 1, only finitely many of the events } E_n\text{'s occur, and}$$
$$P[A] = 1 \iff \text{With probability 1, infinitely many of events } E_n\text{'s occur.}$$

Furthermore, $A_1 \supset A_2 \supset \cdots$. Hence, from Theorem 11.13 we have

$$P[A] = \lim_{n\to\infty} P[A_n], \tag{11.56}$$

where $P[A_n]$ is bounded from above due to Theorem 11.15:

$$P[A_n] \leq \sum_{k=n}^{\infty} P[E_k]. \tag{11.57}$$

With these preparations, we are now in a position to state one of the most important theorems in probability theory, usually referred to as Borel[6]-Cantelli[7] Lemmas:

Theorem 11.16 (Borel-Cantelli Lemmas). *Let $\{E_k\}$ be an infinite sequence of events, and let A be the event that infinitely many of the events E_k's occur, as defined by (11.55). Then*

- ***First lemma:*** *Regardless of the events E_k's being independent or not,*

$$\boxed{\text{If } \sum_{k=1}^{\infty} P[E_k] < \infty, \text{ then } P[A] = 0,} \tag{11.58}$$

that is, with probability 1 only finitely many of the events $E_1, E_2, \ldots,$ occur.

- ***Second lemma;*** *Suppose that $E_1, E_2, \ldots$ are independent events. Then,*

$$\boxed{\text{If } \sum_{k=1}^{\infty} P[E_k] = \infty, \text{ then } P[A] = 1,} \tag{11.59}$$

that is, infinitely many of the events $E_1, E_2, \ldots$ occur with probability 1.

Proof. If $\sum_k P[E_k]$ converges, then (11.57) shows

$$\lim_{n\to\infty} P[A_n] \le \lim_{n\to\infty} \sum_{k=n}^{\infty} P[E_k] = 0. \tag{11.60}$$

By applying (11.56) to this decreasing sequence $\{A_n\}$, we find

$$P[A] = P\left[\lim_{n\to\infty} A_n\right] = \lim_{n\to\infty} P[A_n] = 0.$$

This completes the proof of the First Lemma.

Now we proceed to prove the Second Lemma. Take the complement of A_n of (11.54):

$$A_n^c = \bigcap_{k=n}^{\infty} E_k^c. \tag{11.61}$$

Then using the relation

$$A_n^c \subset \bigcap_{k=n}^{n+m} E_k^c, \quad \text{for every } m = 0, 1, 2, \ldots, \tag{11.62}$$

and the assumption of the Second Lemma that the E_k's are mutually independent and hence so are the E_k^c's (Problem 11.10), we find

$$P[A_n^c] \le P\left[\bigcap_{k=n}^{n+m} E_k^c\right] = P[E_n^c]\cdots P[E_{n+m}^c] = (1-P[E_n])\cdots(1-P[E_{n+m}])$$

$$\le \exp\left(-\sum_{k=n}^{n+m} P[E_k]\right), \quad \text{for every } m = 0, 1, 2, \ldots, \tag{11.63}$$

where we used the inequality $1 - x \le e^{-x}$, $x \ge 0$. If $\sum_{k=1}^{\infty} P[E_k] = \infty$, then $\sum_{k=n}^{n+m} P[E_k] \to \infty$ as $m \to \infty$. Hence, by taking the limit $m \to \infty$ in the last equation, we have

$$P[A_n^c] = 0 \ \text{ for every } n = 1, 2, \ldots.$$

Take the complement of A of (11.55):

$$A^c = \bigcup_{n=1}^{\infty} A_n^c.$$

Thus,

$$\boxed{P[A^c] \le \sum_{n=1}^{\infty} P[A_n^c] = 0.} \tag{11.64}$$

Therefore, we finally have $P[A] = 1 - P[A^c] = 1$. This completes the proof of the Second Lemma. Needless to say, when the E_n's are independent, the First Lemma still applies, that is,

$$\sum_{k=1}^{\infty} P[E_k] < \infty \implies P[A] = 0.$$

$\square$

11.3.2 Weak Law of Large Numbers (WLLN)

Let $\{X_k\} = (X_1, X_2, \ldots, X_k, \ldots)$ be a sequence of RVs with finite mean and variance:

$$E[X_k] = \mu_k, \text{ and } \operatorname{Var}[X_k] = \sigma_k^2, \quad k = 1, 2, \ldots.$$

$$S_n = \sum_{k=1}^{n} X_k, \quad n = 1, 2, \ldots.$$

$$m_n = E[S_n] = \sum_{k=1}^{n} \mu_k$$

$$s_n^2 = \operatorname{Var}[S_n] = E[(S_n - m_n)^2]$$

Define the nth **arithmetic average**[8] $\overline{X}_n$ by

$$\overline{X}_n = \frac{S_n}{n},$$

$$E[\overline{X}_n] = \frac{m_n}{n}$$

$$\operatorname{Var}[\overline{X}_n] = E\left[\left(\overline{X}_n - \frac{m_n}{n}\right)^2\right] = s_n^2/n^2 \qquad (11.69)$$

Suppose

$$\lim_{k\to\infty} \sigma_k^2 = 0.$$

By applying Chebyshev's inequality

$$P[|X_k - \mu_k| \geq \epsilon] \leq \frac{\sigma_k^2}{\epsilon^2}$$

$$\lim_{k\to\infty} P[|X_k - \mu_k| \geq \epsilon] = 0.$$

Similarly,

$$\boxed{\text{If } \lim_{n\to\infty} \frac{s_n^2}{n^2} = 0, \text{ then } \lim_{n\to\infty} P\left[\left|\overline{X}_n - \frac{m_n}{n}\right| \geq \epsilon\right] = 0,} \tag{11.73}$$

If the X_k's are *independent* and their variances are bounded, that is, if there exists a positive number M such that $\sigma_k^2 \leq M$ for all k, then (11.73) is satisfied because

$$\lim_{n\to\infty} \frac{s_n^2}{n^2} = \lim_{n\to\infty} \frac{1}{n^2} \sum_{k=1}^{n} \sigma_k^2 \leq \lim_{n\to\infty} \frac{M}{n} = 0. \tag{11.74}$$

If the X_k's are, in addition to being independent, identically distributed with common mean μ, then the condition in (11.73) obviously holds and the result takes the form

$$\lim_{n\to\infty} P[|\overline{X}_n - \mu| \geq \epsilon] = 0. \tag{11.75}$$

This last result involving i.i.d. RVs is often called the **weak law of large numbers (WLLN)**:

Theorem 11.17 (Weak law of large numbers). *Let $X_1, X_2, \ldots, X_k, \ldots$ be independent and identically RVs with finite mean $E[X_k] = \mu$ and finite variance. Let $\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$. Then we have*

$$\overline{X}_n \xrightarrow{\mathrm{P}} \mu,$$

which is equivalently stated as

$$\boxed{\lim_{n\to\infty} P[|\overline{X}_n - \mu| \geq \epsilon] = 0 \text{ for any } \epsilon > 0.} \tag{11.76}$$

□

As discussed above, the weak law of large numbers is easily generalizable to cases where the X_k's are not identically distributed or even not independent, although the i.i.d. assumption is commonly associated with the statement of the law.

11.3.3 Strong Laws of Large Numbers (SLLN)

Theorem 11.18 (Borel's Strong Law of Large Numbers). *Let $\{B_k\}$ be a sequence of Bernoulli trials with the probability of success p. Then the sequence $\{B_k\}$ obeys the strong law of large numbers, i.e.,*

$$\overline{B}_n \overset{\text{a.s.}}{\to} p, \tag{11.79}$$

or equivalently

$$\boxed{P\left[\lim_{n\to\infty} |\overline{B}_n - p| < \epsilon\right] = 1.} \tag{11.80}$$

Proof. The proof provided by Borel is based on the number theoretic interpretation. First, consider the case $p = 1/2$ (fair coin tossing). Any real number ω taken at random with uniform distribution in the interval $(0, 1)$, can be converted into an infinite sequence $\{B_k(\omega)\}$ by using the binary expansion:

$$\omega = \sum_{k=1}^{\infty} B_k(\omega) 2^{-k},$$

Note that compared with the weaker version, i.e., Bernoulli's Theorem (11.78), the "$\lim_{n\to\infty}$" moves inside the expression $P[\quad]$. The SLLN makes a statement regarding individual **sample sequences** or **sample paths** $B_k(\omega)$ that correspond to each sample point $\omega \in \Omega$ of this Bernoulli experiment. That is, for large n, the $\overline{B}_n(\omega)$ computed from any (except for those belonging to a set of probability measure zero) sample path $\{B_k(\omega)\}$ approaches arbitrarily close to p. Thus, the SLLN suggests that we can estimate the probability p with sufficient accuracy by conducting a single stream of Bernoulli experiments of sufficient length n. In contrast, the WLLN (i.e., Bernoulli's theorem) makes a statement regarding the **entire ensemble** of such sample paths. That is, when we consider all possible sample paths $\{B_k(\omega)\}$, $\omega \in \Omega$, then probabilistically speaking, the RV $\overline{B}_n(\omega)$ becomes arbitrarily close to the constant p, as n is made sufficiently large.

Theorem 11.19 (Kolmogorov's sufficient criterion for the SLLN when X_k's are independent).

Let $\{X_k\}$ be a sequence of independent RVs such that $E[X_k] = \mu_k$ and $\mathrm{Var}[X_k] = \sigma_k^2$. Define

$$\overline{X}_n = \frac{\sum_{k=1}^{n} X_k}{n} \quad \text{and} \quad \overline{\mu}_n = \frac{\sum_{k=1}^{n} \mu_k}{n}.$$

Then

$$\boxed{\text{If } \sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2} < \infty, \quad \text{then } \overline{X}_n - \overline{\mu}_n \overset{\text{a.s.}}{\to} 0,} \tag{11.81}$$

that is, the sequence $X_1, X_2, \ldots$ obeys the SLLN.

Theorem 11.20 (Sufficient condition for the SLLN when X_k's are i.i.d.). *Let $\{X_k\}$ be a sequence of i.i.d. RVs with common mean $\mu = E[X_k]$. Then*

$$\boxed{\text{If } E[X_k^2] < \infty, \text{ then } \overline{X}_n \overset{\text{a.s.}}{\to} \mu, \text{ and } \overline{X}_n \xrightarrow{\text{m.s.}} \mu} \tag{11.82}$$

□

Subsequently in 1933, Kolmogorov showed the necessary and sufficient condition for the SLLN to hold in a sequence of i.i.d RVs. The following theorem is sometimes referred to as Kolmogorov's second theorem for the SLLN.

Theorem 11.21 (Strong Law of Large Numbers (SLLN) when X_k's are i.i.d.). *Let $\{X_k\}$ be a sequence of i.i.d. RVs. Then*

$$\boxed{\overline{X}_n \overset{\text{a.s.}}{\to} \mu, \textit{ if and only if } E[|X_k|] < \infty,} \tag{11.83}$$

where $\mu = E[X_1]$.

11.3.4 The Central Limit Theorem (CLT) Revisited

The SLLN states that the sample average $\overline{X}_n = S_n/n$ converges in a strong sense (i.e., for individual sample paths) to $\mu = E[X]$. But the law does not provide any information about the the distribution of $\overline{X}_n$ other than its mean. The central limit theorem (CLT) stated in Theorem 8.2 of Section 8.2.5 is concerned about this question. Specifically, Eq. (8.98) can now be paraphrased as

$$\frac{\sqrt{n}}{\sigma}(\overline{X}_n - \mu) \xrightarrow{\mathrm{D}} U. \tag{11.84}$$

where U is the standard normal variable. In this section we will discuss several variations of the central limit theorem (CLT).

Define the **normalized average**

$$Z_n = \frac{S_n - m_n}{s_n}, \tag{11.85}$$

Let us consider the simplest case where X_k's are not only **independent** but also **identically distributed** with common mean μ and variance σ^2. Then (11.85) reduces to

$$Z_n = \frac{S_n - n\mu}{\sqrt{n}\sigma} = \sum_{k=1}^{n} \frac{X_k - \mu}{\sqrt{n}\sigma} = \frac{\sqrt{n}}{\sigma}(\overline{X}_n - \mu). \tag{11.86}$$

Then we restate Theorem 8.2 of page 201 as follows:

Theorem 11.22 (Lindeberg-Lévy's CLT for i.i.d. RVs[10]).
Let $\{X_k\}$ be a sequence of i.i.d. RVs. Then,

$$\boxed{\text{if } E[X_1] = \mu < \infty, \;\; \mathrm{Var}[X_1] = \sigma^2 < \infty, \text{ then } Z_n \xrightarrow{\mathrm{D}} U,} \tag{11.87}$$

where Z_n is defined in (11.86) and U is the unit normal variable.

Proof. Note that we write $E[X_1]$ instead of $E[X_k]$ for all k, since that would be redundant, given the i.i.d. assumption. The proof was already given, when we discussed Theorem 8.2, where we used the characteristic function (CF) of the normalized average Z_n. The CF of Z_n tends to $e^{-u^2/2}$, the CF of the unit normal distribution. To justify the transition from the convergence of the CFs to convergence of the corresponding distributions, we need the continuity theorem (often referred to as Lévy Continuity Theorem), which states that a sequence $\{F_n(x)\}$ of probability distributions converges to a probability distribution $F(x)$, if and only if the sequence $\{\phi_n(x)\}$ of their CFs converges to a continuous limit $\phi(u)$. In this case $\phi(u)$ is the CF of $F(x)$, and the sequence $\{\phi_n(u)\}$ converges to $\phi(u)$ uniformly. See Feller [99] (p. 481 and pp. 487-491). $\square$

- **The CLT for the case where the X_k are independent but not necessarily identically distributed.**

Define the third absolute central moment:

$$m_k^3 \triangleq E\left[|X_k - \mu_k|^3\right]. \tag{11.88}$$

Theorem 11.23 (Lyapunov's CLT for independent but non-identical RVs). *Let X_k's be independent RVs with $E[X_k] = \mu_k$ and $\mathrm{Var}[X_k] = \sigma_k^2$, and let s_n^2 be as defined in (11.67). Then,*

$$\boxed{\text{if } m_k < \infty \textit{ for all } k \text{ and } \lim_{n\to\infty} \frac{\sum_{k=1}^n m_k}{s_n} = 0, \text{ then } Z_n \xrightarrow{\mathrm{D}} U,} \tag{11.89}$$

where Z_n is defined in (11.86) and U is the unit normal variable.

Lyapunov generalized the sufficient condition (11.89) and showed that if

$$\lim_{n\to\infty} \frac{\sum_{k=1}^n E\left[(X_k - m_k)^{2+\delta}\right]}{s_n^{2+\delta}} = 0. \tag{11.90}$$

for some $\delta > 0$, then the Z_n converges in distribution to the unit normal variable U. The condition (11.89) corresponds to the case $\delta = 1$.

❖ The necessary and sufficient condition for the CLT

Consider the following quantities:

$$\sigma_k^{\star 2}(\epsilon) \triangleq \int_{|x-\mu_k| \le \epsilon s_n} (x-\mu_k)^2 dF_{X_k}(x), \quad k = 1, 2, \ldots, n, \tag{11.91}$$

$$s_n^{\star 2}(\epsilon) \triangleq \sum_{k=1}^{n} \sigma_k^{\star 2}(\epsilon). \tag{11.92}$$

Theorem 11.24 (Lindeberg-Feller's CLT for independent but non-identical RVs). *Let X_k's be independent RVs with $E[X_k] = \mu_k$, $\mathrm{Var}[X_k] = \sigma_k^2$, and $s_n^{\star 2}(\epsilon)$ defined in (11.92). Then,*

$$Z_n \xrightarrow{\mathrm{D}} U, \text{ if and only if } \lim_{n\to\infty} s_n^2 = \infty, \text{ and } \lim_{n\to\infty} \frac{s_n^{\star 2}(\epsilon)}{s_n^2} = 1 \text{ for every } \epsilon > 0, \tag{11.93}$$

where Z_n is defined in (11.86) and U is the unit normal variable.

The condition (11.93), called the **Lindeberg condition**, guarantees that the individual variances σ_k^2's are all small in comparison to their sum s_n^2. It can be shown that the **Lyapunov's condition** (11.90) satisfies Lindeberg's condition (11.93), because the sum in (11.93) is bounded by

$$\frac{1}{s_n^2}\sum_{k=1}^{n}\int_{|x-\mu_k|>\epsilon s_n}\frac{(x-\mu_k)^{2+\delta}}{\epsilon^\delta s_n^\delta}d_{X_k}(x) \le \frac{1}{\epsilon^\delta}\frac{\sum_{k=1}^{n}E\left[(X_k-m_k)^{2+\delta}\right]}{s_n^{2+\delta}}.$$

Before closing this section, we should reiterate that the preceding results on the CLT are concerned with convergence in distribution. In other words, the distribution of the normalized average Z_n converges to that of the unit normal variable U. This does not necessarily imply that the probability density function (PDF) of Z_n converges to that of U. If the X_k's are continuous RVs, then under some regularity conditions, the PDFs of Z_n will converge to that of U. But for a finite value of n, the normal distribution may well give a *poor approximation* to the tails of the PDF, as we discussed in Chapter 10. In fact the large deviations approximation discussed in that chapter was motivated by this poor approximation of tail end of the distribution by the normal distribution. If the X_k's are discrete RVs, the situation is more complicated. For example, the sum of independent Poisson random variables is also a Poisson variable for any n, although the *envelope* of the density function of Z_n may converge to that of U.

12 Random Processes

12.1 Random Process

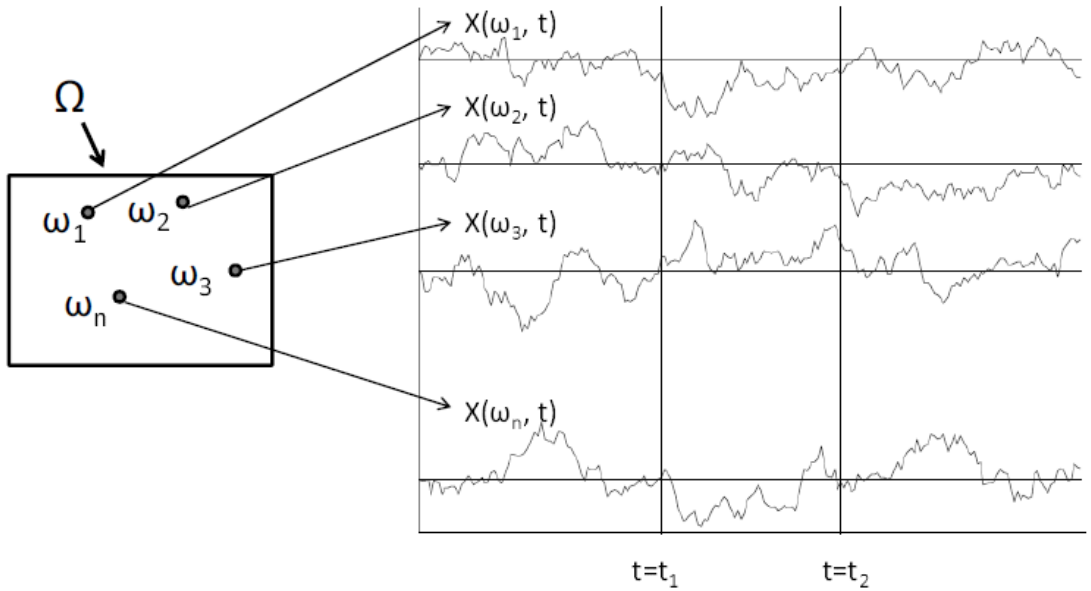


Figure 12.1 A random process $X(\omega, t)$ as a mapping from a sample point $\omega \in \Omega$ to a real-valued function.

A probability system, which is composed of a sample space, a set of real-valued time-indexed functions, and a probability measure, is called a **random process** or a **stochastic process**, and is usually denoted by a notation such as $X(t); t \in \mathcal{T}$, or simply as $X(t)$, if $\mathcal{T}$ is implicitly understood.

The individual time functions of the random process $X(t)$ are called **sample functions**

By definition, a random process implies the existence of an infinite number of random variables, one for each t in some range. Thus, we may speak of the probability density function $f_{X(t_1)}(\cdot)$ of the random variable $X(t_1)$ obtained by observing $X(t)$ at time t_1.

12.2 Classification of Random Processes

12.2.1 Discrete-Time vs. Continuous-Time Processes

A discrete-time random process or a random sequence, denoted as X_t or X_k .

A continuous-time random process , denoted as $X(t)$

The **term time** series is used often almost synonymously with discrete-time random process, in statistics, signal processing, econometrics and social sciences.

12.2.2 Discrete-State vs. Continuous-State Processes

State space: a set of possible values that a random process, discrete or continuous in time, may take on

Examples of **discrete states** :

A Bernoulli trial: S={ s, f } or S={ 0, 1}

A simple random walk: S = {0, ± 1, ±2, ±3, … } (step size)

The price of a stock: S= {0, 1, 2, …. } (unit price)

Examples of **continuous states**:

The temperature as a function of time: S =(-∞, ∞)

Gaussian process: S = (-∞, ∞)

Brownian motion or Wiener process, a limit of the random walk : S = (-∞, ∞)

Inter-arrival time of a Poisson process: S=[0, ∞)

Digital technology transforms a continuous-time, continuous-space process to a discrete-time, discrete-space process. E.g., CD and DVD.

12.2.3 Stationary vs. Non-Stationary Processes

A process $X(t)$ is called a **stationary process** if its distribution function

$$F_X(x;t) \triangleq P[X(t) \leq x], \quad t \in \mathcal{T} \tag{12.1}$$

is independent of time t. Otherwise $X(t)$ is called **nonstationary.**

Examples of **stationary process**:

An infinite series of Bernoulli trials

A Gaussian process

Examples of **nonstationary process**:

A random walk: The variance is proportional to the step size n.

Brownian motion: The variance is proportional to time t.

The price of a stock or the Dow Jones' index

Mathematically, a stationary process must have begun in the infinite past and will continue into the infinite future. We write $X(t)$; $-\infty < t < \infty$.

12.2.4 Independent versus dependent processes

Suppose we arbitrarily choose n time instants and consider the joint distribution function $F_{\boldsymbol{X}}(\boldsymbol{x}, \boldsymbol{t})$ of the set of random variables $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ where $X_i = X(t_i); i = 1, 2, \ldots, n$. If this distribution function factors into the product:

$$\begin{aligned} F_{\boldsymbol{X}}(\boldsymbol{x}; \boldsymbol{t}) &\triangleq F_{X_1 X_2 \cdots X_n}(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n) \\ &= F_{X_1}(x_1; t_1) F_{X_2}(x_2; t_2) \cdots F_{X_n}(x_n; t_n), \end{aligned} \tag{12.2}$$

for any finite n and for any choice of the instants $\boldsymbol{t}$, we say $X(t)$ is an **independent process**.

Examples of **independent** processes:

A Bernoulli trials

Step sizes of a random walk (i.e., the difference sequence of a random walk)

White noise (i.e., the power spectral is flat for all frequencies)

Interarrival times of a Poisson process

Many examples of random sequences X_k or X_n discussed in Chapter 11.

Examples of **dependent** process

A **random walk**

Brownian motion (integration of white noise)

A **Makov** process (discrete-time, continuous-time)

Packet traffic over LAN is known to have long-range dependency (LRD)

12.2.5 Markov chains and Markov processes

12.2.5.1 Discrete-time Markov chain (DTMC)

A discrete-time random process $\{X_k\}$ is called a **simple Markov chain**, if X_{k+1} is Independent of $X_1, X_2, \ldots, X_{k-1}$ in case X_k is known

i.e., if X_{k+1} depends on its past only through its most recent value X_k .

A Markov chain of **order *h*** is a sequence in which X_k depends on its past only through its ***h* previous values**, X_{k-1}, X_{k-2}, … , X_{k-h}.

$$p(x_k | x_{k-i};\ i \geq 1) = p(x_k | x_{k-1}, x_{k-2}, \ldots, x_{k-h}). \tag{12.3}$$

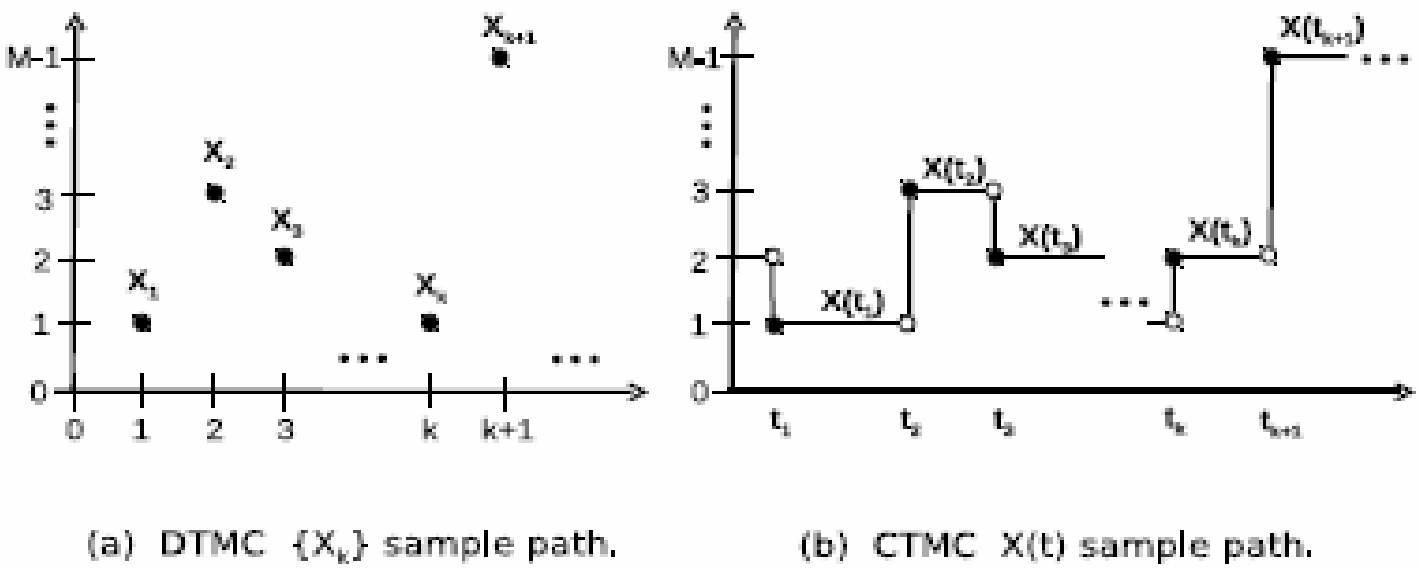
A Markov chain of order *h* defined over state space *S* can be transformed into a simple Markov chain by defining the state space $S^h = S \times S \cdots \times S$, the *h*-times Cartesian product of *S* with itself.

Markov chain models and related **hidden Markov models** (**HMM**s) are used in a variety of fields, including linguistic models for speech recognition, DNA and protein sequences, network traffic, etc. (cf. Chapters 1 and 20).

The simple Markov chain defined above is often referred to as a **discrete-time Markov chain (DTMC)**.

If there are M different states, we can label them, without loss of generality, by integers 0, 1, 2, M-1, i.e.,

$$\mathcal{S} = \{0, 1, 2, \ldots, M-1\}, \tag{12.4}$$



(a) DTMC $\{X_k\}$ sample path. (b) CTMC $X(t)$ sample path.

Figure 12.2 (a) A discrete-time Markov chain (DTMC); (b) a continuous-time Markov chain (CTMC).

12.5.2 Continuous-time Markov chain (CTMC)

For a given DTMC $\{X_k\}$ we can construct a **continuous-time Markov chain (CTMC)** $X(t)$

$$X(t) = i, \text{ for } t_k \le t < t_{k+1}, \text{ where } i = X(t_k), \text{ and } X(t_{k+1}) = j(\neq i). \tag{12.5}$$

and let the interval $\tau_k \triangleq t_{k+1} - t_k$ be **exponentially distributed** with mean λ_k^{-1}.

Figure 12.2 (b) of the previous slide shows a sample path of a CTMC.

The future behavior of $X(t)$; $t \ge t_n$ depends on its past $X(s)$; $-\infty < s < t_n$ only through its current state $X(t_n) = i \in S$, because of the **memoryless property** of the exponential distribution.

For a **Poisson arrival** process, let $X(t)$ be the cumulative number of arrivals (or births) up to time t. This **counting process** is a CTMC.