20. Extinction Probability for Queues and Martingales

(Refer to section 15.6 in text (**Branching processes)** for discussion on the extinction probability).

20.1 Extinction Probability for Queues:

- A customer arrives at an empty server and immediately goes for service initiating a busy period. During that service period, other customers may arrive and if so they wait for service. The server continues to be busy till the last waiting customer completes service which indicates the end of a busy period. An interesting question is whether the busy periods are bound to terminate at some point ? Are they ?

Do busy periods continue forever? Or do such queues come to an end sooner or later? If so, how ?

• **Slow Traffic ($\rho \le 1$)**
Steady state solutions exist and the probability of extinction equals 1. (Busy periods are bound to terminate with probability 1. Follows from sec 15.6, theorem 15-9.)

• **Heavy Traffic ($\rho > 1$)**
Steady state solutions do not exist, and such queues can be characterized by their probability of extinction.

• Steady state solutions exist if the traffic rate $\rho < 1.$ Thus

$$p_k = \lim_{n\to\infty} P\{X(nT)=k\} \text{ exists if } \rho < 1.$$

• What if too many customers rush in, and/or the service rate is slow ($\rho \ge 1$) ? How to characterize such queues ?

Extinction Probability (π_0) for Population Models

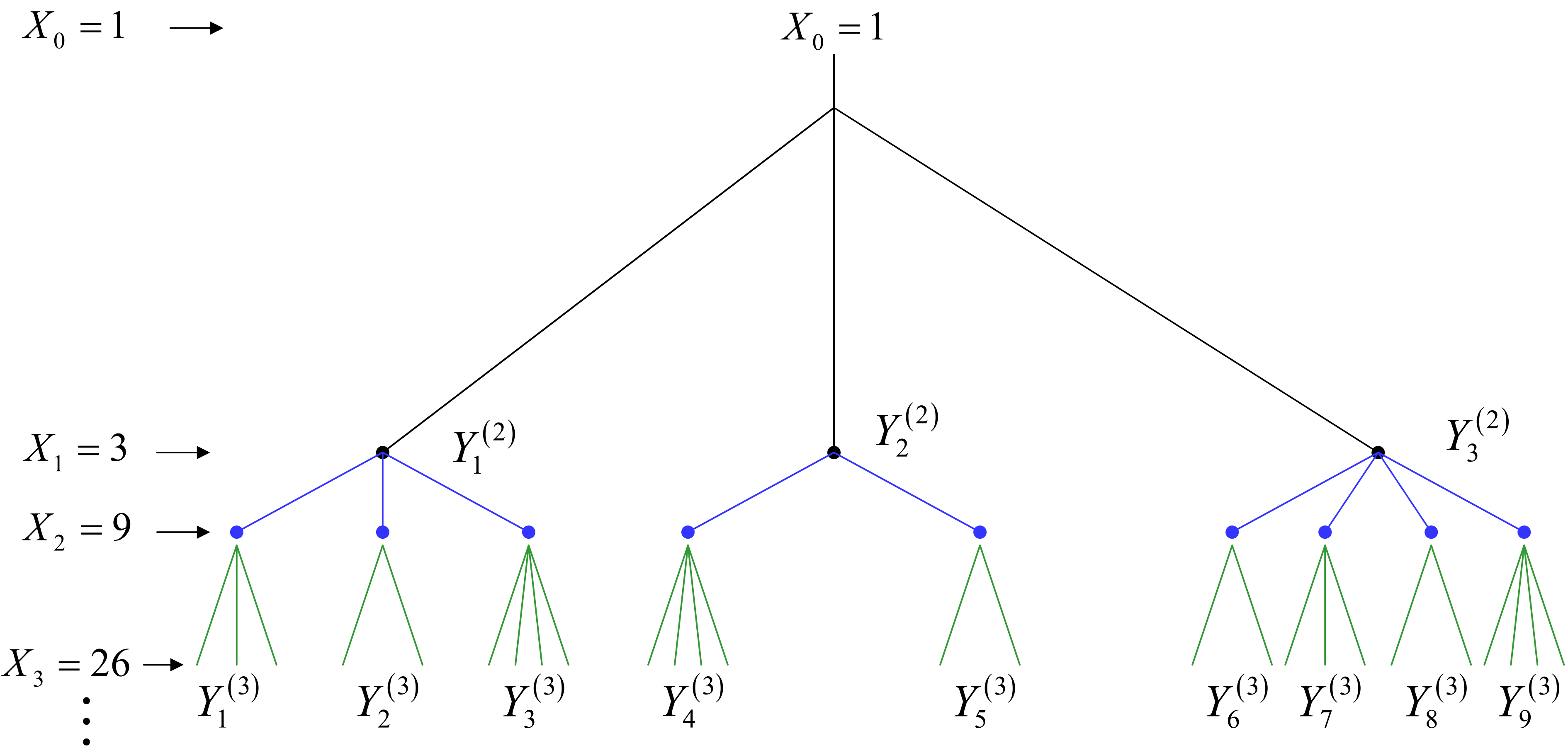


Fig 20.1

Queues and Population Models

- **Population models**

X_n : Size of the n[th] generation

$Y_i^{(n)}$: Number of offspring for the i^{th} member of the n[th] generation. From Eq.(15-287), Text

$$X_{n+1} = \sum_{k=1}^{X_n} Y_k^{(n)}$$

Let

$$a_k \triangleq P\{Y_i^{(n)} = k\}$$

- Offspring moment generating function:

$$P(z) = \sum_{k=0}^{\infty} a_k z^k \qquad (20\text{-}1)$$

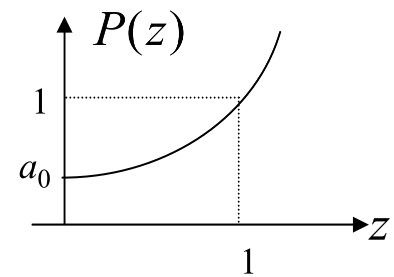


Fig 20.2

$$P_{n+1}(z) = \sum_{k=0}^{\infty} P\{X_{n+1} = k\}\, z^k = E\{z^{X_{n+1}}\}$$

$$= E(E\{z^{X_{n+1}} \mid X_n = j\}) = E\left(E\left\{ z^{\sum_{i=1}^{j} Y_i} \,\middle|\, X_n = j \right\} \right)$$

$$= E\{[P(z)]^j\} = \sum_j \{P(z)\}^j P\{X_n = j\} = P_n(P(z)) \qquad (20\text{-}2)$$

$$P_{n+1}(z) = P_n(P(z)) = P(P_n(z)) \qquad (20\text{-}3)$$

$$P\{X_n = k\} = ?$$

Extinction probability $= \lim_{n \to \infty} P\{X_n = 0\} = \pi_o = ?$

Extinction probability π_0 satisfies the equation $P(z) = z$ which can be solved iteratively as follows:

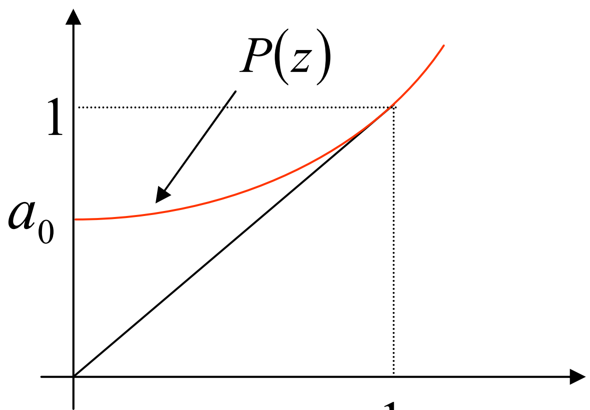
$$z_0 = P(0) \overset{\Delta}{=} a_0 \qquad (20\text{-}4)$$

and

$$z_k = P(z_{k-1}), \qquad k = 1, 2, \cdots \qquad (20\text{-}5)$$

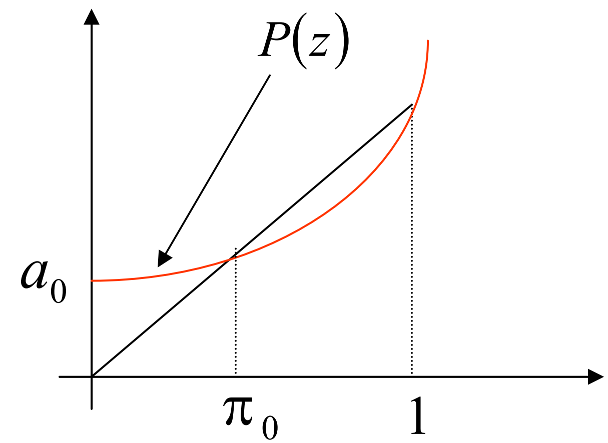
- Review Theorem 15-9 (Text)

Let $$\rho = E(Y_i) = P'(1) = \sum_{k=0}^{\infty} k\ P\{Y_i = k\} = \sum_{k=0}^{\infty} k a_k > 0 \qquad (20\text{-}6)$$



(a) $\rho \le 1$

Fig 20.3



(b) $\rho > 1$

$$\left.\begin{array}{l} \rho \le 1 \Rightarrow \pi_0 = 1 \\ \rho > 1 \Rightarrow \pi_0 \text{ is the unique solution of } P(z) = z\ , \\ a_0 < \pi_0 < 1 \end{array}\right\} \qquad (20\text{-}7)$$

- Left to themselves, in the long run, populations either die out completely with probability π_0, or explode with probability $1-\pi_0$. (Both unpleasant conclusions).

Queues :

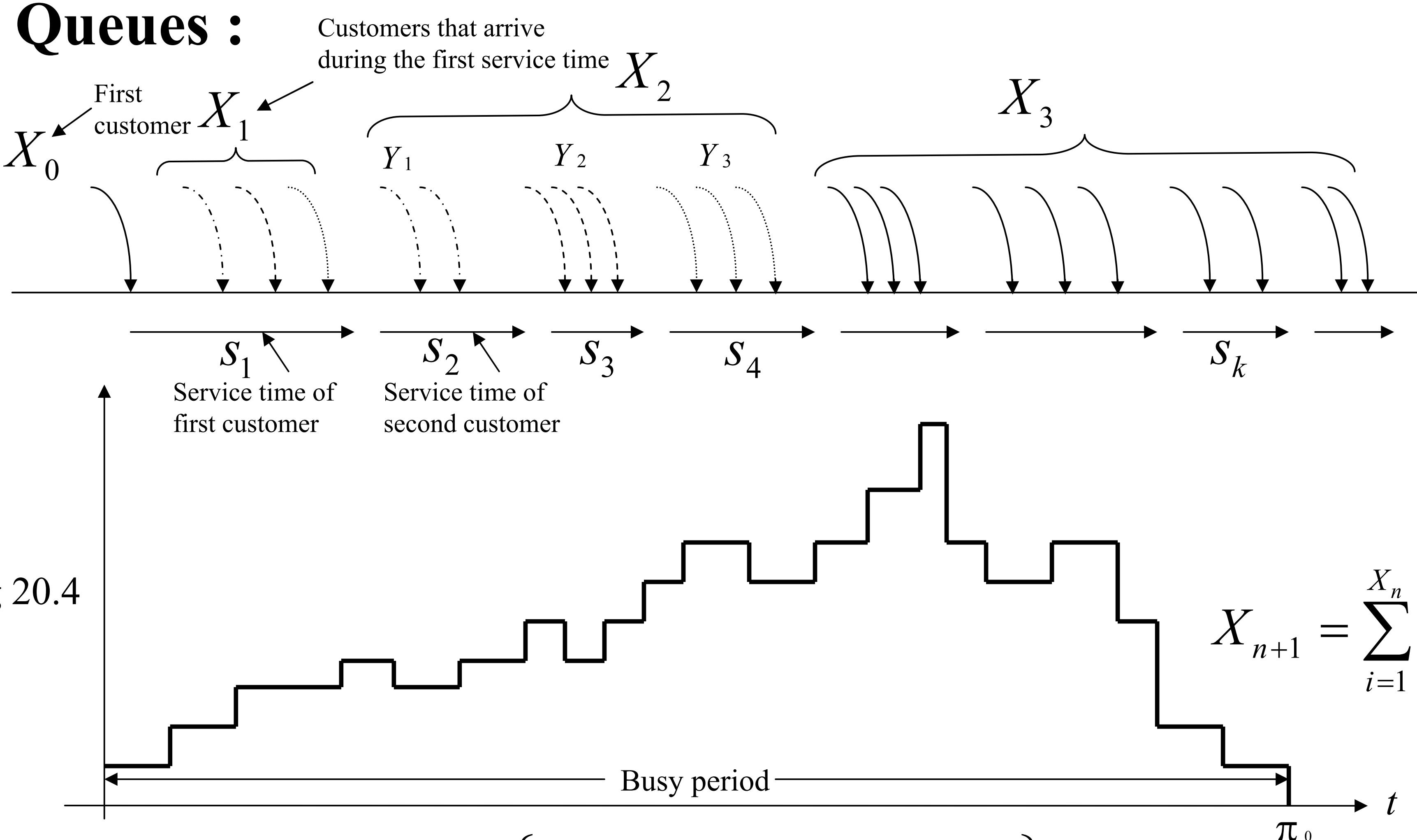


Fig 20.4

$$X_{n+1} = \sum_{i=1}^{X_n} Y_i$$

$$a_k = P(Y_i = k) = P\left\{ \begin{array}{c} k \text{ arrivals between two} \\ \text{successive departures} \end{array} \right\} \geq 0.$$

Note that the statistics $\{a_k\}$ depends both on the arrival as well as the service phenomena.

- $P(z)=\sum_{k=0}^{\infty} a_k z^k$: Inter-departure statistics generated by arrivals
- $\rho = P'(1)=\sum_{k=1}^{\infty} k a_k$: Traffic Intensity $\left.\begin{matrix}\le 1\\ >1\end{matrix}\right\} \Rightarrow$ $\begin{matrix}\text{Steady state}\\ \text{Heavy traffic}\end{matrix}$
- Termination of busy periods corresponds to extinction of queues. From the analogy with population models the extinction probability π_0 is the unique root of the equation $P(z)=z$

- Slow Traffic : $\rho \le 1 \;\Rightarrow\; \pi_0 = 1$

  Heavy Traffic : $\rho > 1 \;\Rightarrow\; 0 < \pi_0 < 1$

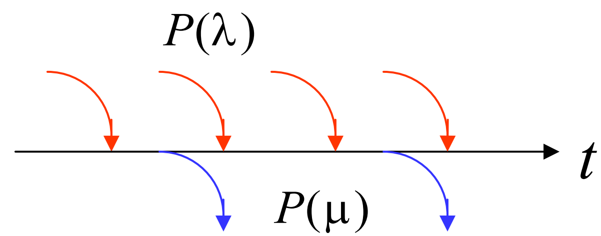
  i.e., unstable queues $(\rho > 1)$ either terminate their busy periods with probability $\pi_0 < 1$, or they will continue to be busy with probability 1- π_0. Interestingly, there is a finite probability of busy period termination even for unstable queues.

  π_0 : Measure of stability for unstable queues.

Example 20.1 : *M/M/* 1 queue

From (15-221), text,we have

$$a_k = \frac{\mu}{\lambda+\mu}\left(\frac{\lambda}{\lambda+\mu}\right)^k = \frac{1}{1+\rho}\left(\frac{\rho}{1+\rho}\right)^k, \qquad k = 0, 1, 2, \cdots \tag{20-8}$$



Number of arrivals between any two departures follows a geometric random variable.

$$P(z) = \frac{1}{[1+\rho(1-z)]}, \qquad \rho = \frac{\lambda}{\mu} \tag{20-9}$$

$$P(z) = z \Rightarrow \rho\, z^2 - (\rho+1)z + 1 = (z-1)\,(\rho\, z - 1) = 0$$

$$\pi_0 = \begin{cases} 1, & \rho \le 1 \\ \dfrac{1}{\rho}, & \rho > 1 \end{cases} \tag{20-10}$$

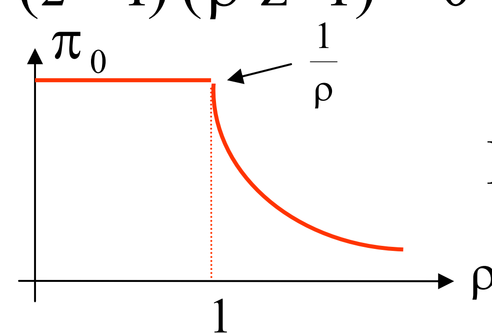


Fig 20.5

Example 20.2 : Bulk Arrivals $M^{[x]}/M/1$ queue

Compound Poisson Arrivals : Departures are exponential random variables as in a Poisson process with parameter μ. Similarly arrivals are Poisson with parameter λ. However each arrival can contain multiple jobs.

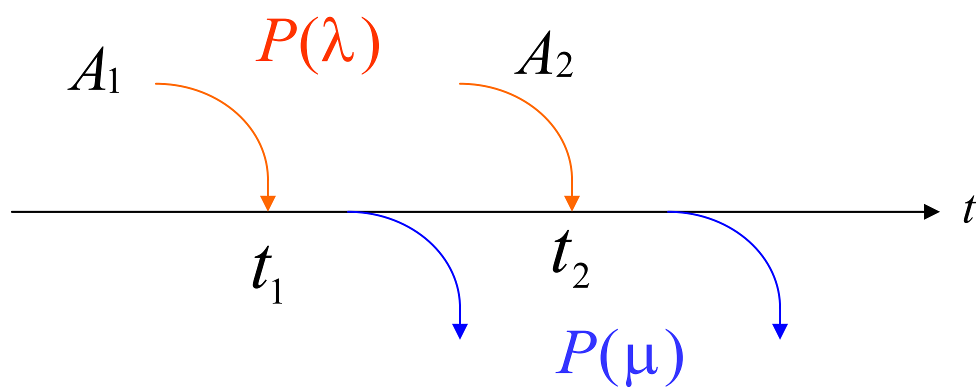


Fig 20.6

A_i : Number of items arriving at instant t_i

Let $P\{A_i = \mathrm{k}\} = c_k, \quad k = 0, 1, 2, \cdots$

and $C(z) = E\{z^{A_i}\} = \sum_{k=0}^{\infty} c_k z^k$ represents the bulk arrival statistics.

Inter-departure Statistics of Arrivals

$$P(z)=\sum_{k=0}^{\infty}P\{Y=\mathrm{k}\}\,z^k\;=E\{z^Y\}$$

$$=\sum_{n=0}^{\infty}\int_0^{\infty}E\{z^{(A_1+A_2+\cdots+A_n)}\;\big|n\text{ arrivals in }(0,t)\}\,P\{n\text{ arrivals in }(0,t)\}\,f_s(\mathrm{t})\,dt$$

$$=\sum_{n=0}^{\infty}\int_0^{\infty}[E\{z^{A_i}\}]^n\;e^{-\lambda t}\,\frac{(\lambda t)^n}{n!}\,\mu\;e^{-\lambda t}dt\;,\qquad\rho=\frac{\lambda}{\mu}$$

$$=\sum_{n=0}^{\infty}\mu\int_0^{\infty}e^{-(\lambda+\mu)t}\,\frac{[\lambda t\;C(z)]^n}{n!}=\frac{1}{1+\rho\{1-C(z)\}}\qquad(20\text{-}11)$$

$$P(z)=[1+\rho\{1-C(z)\}]^{-1},\;\text{Let}\quad c_k=(1-\alpha)\,\alpha^k,\qquad k=0,\;1,\;2,\;\cdots$$

$$C(z)=\frac{1-\alpha}{1-\alpha z},\qquad P(z)=\frac{1-\alpha\;z}{1+\alpha\rho-\alpha(1+\rho)z}\qquad(20\text{-}12)$$

$$\text{Traffic Rate}=P'(1)=\frac{\alpha\rho}{1-\alpha}>1$$

$$P(z)=z\;\;\Rightarrow\;\;(\mathrm{z}-1)[\alpha(1+\rho)\mathrm{z}-1]=0\quad\pi_0=1/(\alpha(1+\rho))$$

Bulk Arrivals (contd)

- Compound Poisson arrivals with geometric rate

$$\pi_0 = \frac{1}{\alpha(1+\rho)}, \qquad \rho > \frac{1-\alpha}{\alpha}. \qquad (20\text{-}13)$$

For $\alpha = \frac{2}{3}$, we obatain

$$\pi_0 = \frac{3}{2\,(1+\rho)}, \qquad \rho > \frac{1}{2} \qquad (20\text{-}14)$$

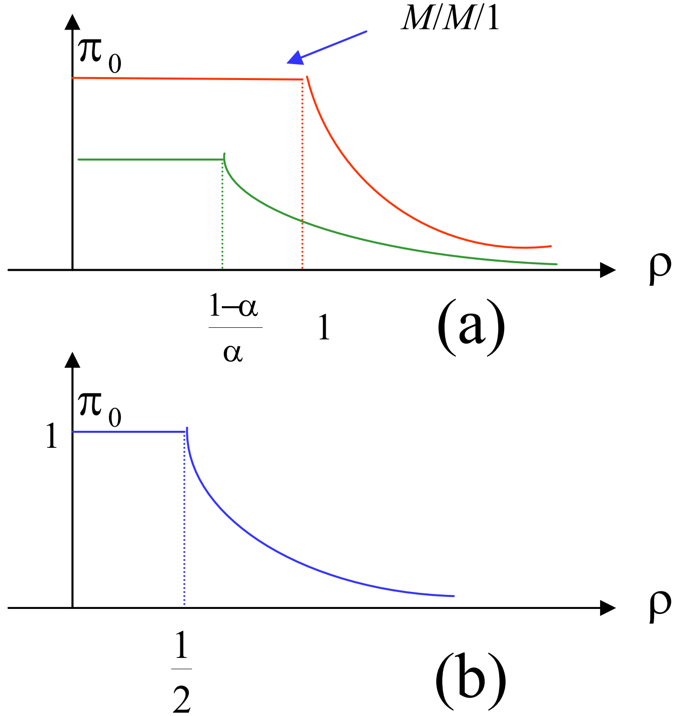


Fig 20.7

- **Doubly Poisson arrivals** $\Rightarrow \quad C(z) = e^{-\mu(1-z)}$ gives

$$P(z) = \frac{1}{1+\rho[1-C(z)]} = z \qquad (20\text{-}15)$$

Example 20.3 : *M/E*$_n$ / 1 queue (*n*-phase exponential service)

From (16-213)

$$P(z)=\left(1+\frac{\rho}{n}(1-z)\right)^{-n} \qquad (20\text{-}16)$$

$$P(z)=z \Rightarrow \left(x^{n}+x^{n-1}+\cdots+x-\frac{n}{\rho}\right)=0, \quad x=z^{\frac{1}{n}}$$

$$n=2 \Rightarrow \pi_0=\left(\frac{-1+\sqrt{1+8/\rho}}{2}\right)^{2}, \quad \rho>1$$

$$\pi_0 \approx (1+\rho/n)^{-n}, \quad n>>1 \qquad (20\text{-}17)$$

$\rho=2$	
$M/M/1 \to n=1$	$\pi_0=0.5$
$M/E_2/1 \to n=2$	$\pi_0=0.38$

Example 20.4 : *M/D/*1 queue

Letting $m\to\infty$ in (16.213),text, we obtain $P(z)=e^{-\rho(1-z)}$, so that

$$\pi_0 \approx e^{-\rho(1-e^{-\rho})}, \quad \rho>1. \qquad (20\text{-}18)$$

20.2 Martingales

Martingales refer to a specific class of stochastic processes that maintain a form of "stability" in an overall sense. Let $\{X_i, i \geq 0\}$ refer to a discrete time stochastic process. If n refers to the present instant, then in any realization the random variables $X_0, X_1, \cdots, X_n$ are known, and the future values $X_{n+1}, X_{n+2}, \cdots$ are unknown. The process is "stable" in the sense that conditioned on the available information (past and present), no change is expected on the average for the future values, and hence the conditional expectation of the immediate future value is the same as that of the present value. Thus, if

$$E\{X_{n+1} \mid X_n, X_{n-1}, \cdots, X_1, X_0\} = X_n \qquad (20\text{-}19)$$

for all n, then the sequence $\{X_n\}$ represents a **Martingale**.

Historically *martingales* refer to the "doubling the stake" strategy in gambling where the gambler doubles the bet on every loss till the almost sure win occurs eventually at which point the entire loss is recovered by the wager together with a modest profit. Problems 15-6 and 15-7, chapter 15, Text refer to examples of martingales. [Also refer to section 15-5, Text].

If $\{X_n\}$ refers to a Markov chain, then as we have seen, with

$$p_{ij} = P\{X_{n+1} = j \mid X_n = i\},$$

Eq. (20-19) reduces to the simpler expression [Eq. (15-224), Text]

$$\sum_j j\, p_{ij} = i. \tag{20-20}$$

For finite chains of size N, interestingly, Eq. (20-20) reads

$$P x_2 = x_2, \qquad x_2 = [1,\ 2,\ 3,\ \cdots,\ N]^T \tag{20-21}$$

implying that x_2 is a right-eigenvector of the $N \times N$ transition probability matrix $P = (p_{ij})$ associated with the eigenvalue 1. However, the "all one" vector $x_1 = [1,\ 1,\ 1,\ \cdots,\ 1]^T$ is always an eigenvector for any P corresponding to the unit eigenvalue [see Eq. (15-179), Text], and from Perron's theorem and the discussion there [Theorem 15-8, Text] it follows that, for finite Markov chains that are also martingales, P *cannot* be a *primitive* matrix, and the corresponding chains are in fact *not irreducible*. Hence every finite state martingale has at least two closed sets embedded in it. (The closed sets in the two martingales in Example 15-13, Text correspond to two absorbing states. Same is true for the Branching Processes discussed in the next example also. Refer to remarks following Eq. (20-7)).

Example 20.5: As another example, let $\{X_n\}$ represent the branching process discussed in section 15-6, Eq. (15-287), Text. Then Z_n given by

$$Z_n = \pi_0^{X_n}, \qquad X_n = \sum_{i=1}^{X_{n-1}} Y_i \tag{20-22}$$

is a martingale, where Y_i s are independent, identically distributed random variables, and π_0 refers to the extinction probability for that process [see Theorem 15.9, Text].

To see this, note that

$$E\{Z_{n+1} \mid Z_n, \cdots, Z_0\} = E\{\pi_0^{X_{n+1}} \mid X_n, \cdots, X_0\}$$

$$= E\{\pi_0^{\sum_{i=0}^{k} Y_i} \mid \underbrace{X_n = k}_{\substack{\text{since } \{X_n\} \text{ is} \\ \text{a Markov chain}}} \} = \prod_{i=1}^{X_n = k} \underbrace{[E\{\pi_0^{Y_i}\}]}_{\substack{\text{since } Y_i \text{ s are} \\ \text{independent of } X_n}} = \underbrace{[P(\pi_0)]^{X_n}}_{\text{use (15-2)}} = \pi_0^{X_n} = Z_n, \tag{20-23}$$

where we have used the Markov property of the chain, the common moment generating function $P(z)$ of $Y_i\, s$, and Theorem 15-9, Text.

Example 20.6 (DeMoivre's Martingale): The gambler's ruin problem (see Example 3-15, Text) also gives rise to various martingales. (see problem 15-7 for an example).

From there, if S_n refers to player A's cumulative capital at stage n, (note that $S_0 = \$\, a$), then as DeMoivre has observed

$$Y_n = \left(\frac{q}{p}\right)^{S_n} \tag{20-24}$$

generates a martingale. This follows since

$$S_{n+1} = S_n + Z_{n+1} \tag{20-25}$$

where the instantaneous gain or loss given by Z_{n+1} obeys

$$P\{Z_{n+1} = 1\} = p, \qquad P\{Z_{n+1} = -1\} = q, \tag{20-26}$$

and hence

$$E\{Y_{n+1} \mid Y_n, Y_{n-1}, \cdots, Y_0\} = E\{\left(\frac{q}{p}\right)^{S_{n+1}} \mid S_n, S_{n-1}, \cdots, S_0\}$$

$$= E\{\left(\frac{q}{p}\right)^{S_n + Z_{n+1}} \mid S_n\},$$

since $\{S_n\}$ generates a Markov chain.

Thus

$$E\{Y_{n+1} \mid Y_n, Y_{n-1}, \cdots, Y_0\} = \left(\tfrac{q}{p}\right)^{S_n} \left(\tfrac{q}{p} \cdot p + \left(\tfrac{q}{p}\right)^{-1} \cdot q\right) = \left(\tfrac{q}{p}\right)^{S_n} = Y_n \quad \text{(20-27)}$$

i.e., Y_n in (20-24) defines a martingale!

Martingales have excellent convergence properties in the long run. To start with, from (20-19) for any *given* n, taking expectations on both sides we get

$$E\{X_{n+1}\} = E\{X_n\} = E\{X_0\}. \quad \text{(20-28)}$$

Observe that, as stated, (20-28) is true only when n is *known* or n is a *given* number.

As the following result shows, martingales do not fluctuate wildly. There is in fact only a small probability that a large deviation for a martingale from its initial value will occur.

Hoeffding's inequality: Let $\{X_n\}$ represent a martingale and $\sigma_1, \sigma_2, \cdots,$ be a sequence of real numbers such that the random variables

$$Y_i \triangleq \frac{X_i - X_{i-1}}{\sigma_i} \leq 1 \qquad \textit{with probability one.} \qquad (20\text{-}29)$$

Then

$$P\{|X_n - X_0| \geq x\} \leq 2e^{-(x^2/2\sum_{i=1}^{n}\sigma_i^2)} \qquad (20\text{-}30)$$

Proof: Eqs. (20-29)-(20-30) state that so long as the martingale increments remain bounded almost surely, then there is only a very small chance that a large deviation occurs between X_n and X_0. We shall prove (20-30) in three steps.

(i) For any convex function $f(x)$, and $0 < \alpha < 1,$ we have (Fig 20.8)

$$\alpha f(x_1) + (1-\alpha) f(x_2) \geq f(\alpha x_1 + (1-\alpha)x_2), \qquad (20\text{-}31)$$

which for $\alpha = \frac{1-a}{2},\ 1-\alpha = \frac{1+a}{2},$ $|a|<1,\ x_1=-1,\ x_2=1$ and $f(x)=e^{\varphi x},\ \varphi>0$ gives

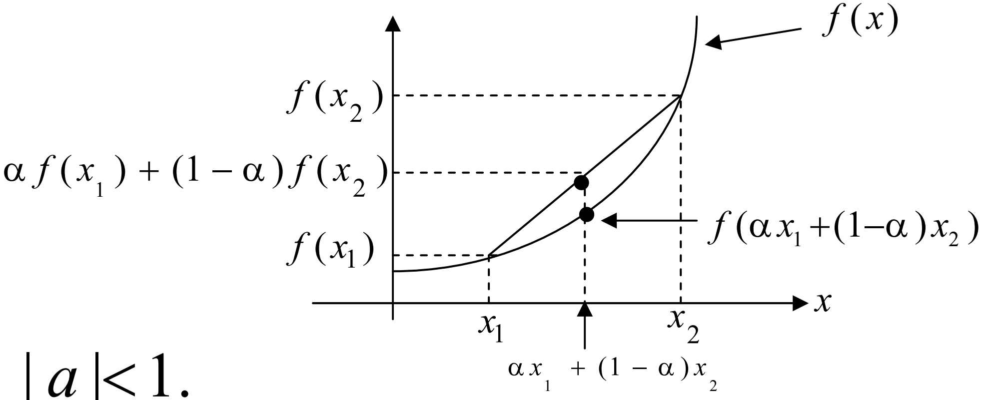


Fig 20.8

$$\frac{1}{2}(1-a)e^{-\varphi}+\frac{1}{2}(1+a)e^{\varphi}\ge e^{a\varphi}, \qquad |a|<1. \qquad (20\text{-}32)$$

Replacing a in (20-32) with any zero mean random variable Y that is bounded by unity almost everywhere, and taking expected values on both sides we get

$$E\{e^{\varphi Y}\}\le \tfrac{1}{2}(e^{\varphi}+e^{-\varphi})\le e^{\varphi^2/2} \qquad (20\text{-}33)$$

Note that the right side is independent of Y in (20-33). On the other hand, from (20-29)

$$E\{Y_i \mid X_i,\cdots,X_1,X_0\}=E(X_i \mid X_{i-1})-X_{i-1}=X_{i-1}-X_{i-1}=0 \qquad (20\text{-}34)$$

and since Y_i *s* are bounded by unity, from (20-32) we get (as in (20-33))

$$E\{e^{\varphi Y_i} \mid X_{i-1},\cdots,X_1,X_0\} \le e^{\varphi^2/2} \qquad (20\text{-}35)$$

(ii) To make use of (20-35), referring back to the Markov inequality in (5-89), Text, it can be rewritten as

$$P\{X \ge \alpha\} \le e^{-\theta\alpha} E\{e^{\theta X}\}, \quad \theta > 0 \qquad (20\text{-}36)$$

and with $X = X_n - X_0$ and $\alpha = x,$ we get

$$P\{X_n - X_0 \ge x\} \le e^{-\theta x} E\{e^{\theta(X_n - X_0)}\} \qquad (20\text{-}37)$$

But

use (20-29)

$$\begin{aligned} E\{e^{\theta(X_n - X_0)}\} &= E\{e^{\theta(X_n - X_{n-1}) + \theta(X_{n-1} - X_0)}\} \\ &= E[E\{e^{\theta(X_{n-1}-X_0)} e^{\theta\sigma_n Y_n} \mid X_{n-1},\cdots,X_1,X_0\}] \\ &= E[e^{\theta(X_{n-1}-X_0)} \underbrace{E\{e^{\theta\sigma_n Y_n} \mid X_{n-1},\cdots,X_1,X_0\}}_{\le e^{\theta^2\sigma_n^2/2} \text{ using (20-35)}}] \\ &\le E\{e^{\theta(X_{n-1}-X_0)}\} e^{\theta^2\sigma_n^2/2} \le e^{\theta^2 \sum_{i=1}^n \sigma_i^2/2}. \qquad (20\text{-}38) \end{aligned}$$

Substituting (20-38) into (20-37) we get

$$P\{X_n - X_0 \ge x\} \le e^{-(\theta x - \theta^2 \sum_{i=1}^n \sigma_i^2 / 2)} \tag{20-39}$$

(iii) Observe that the exponent on the right side of (20-39) is minimized for $\theta = x / \sum_{i=1}^n \sigma_i^2$ and hence it reduces to

$$P\{X_n - X_0 \ge x\} \le e^{-x^2 / 2\sum_{i=1}^n \sigma_i^2}, \qquad x > 0. \tag{20-40}$$

The same result holds when $X_n - X_0$ is replaced by $X_0 - X_n$, and adding the two bounds we get (20-30), the Hoeffding's inequality.

From (20-28), for any fixed n, the mean value $E\{X_n\}$ equals $E\{X_0\}$. Under what conditions is this result true if we replace n by a *random time* T? i.e., if T is a random variable, then when is

$$E\{X_T\} \stackrel{?}{=} E\{X_0\}. \qquad (20\text{-}41)$$

The answer turns out to be that T has to be a *stopping time*. What is a **stopping time**?

A stochastic process may be known to assume a particular value, but the time at which it happens is in general unpredictable or random. In other words, the nature of the outcome is fixed but the timing is random. When that outcome actually occurs, the time instant corresponds to a *stopping time*. Consider a gambler starting with \$a and let T refer to the time instant at which his capital becomes \$1. The random variable T represents a stopping time. When the capital becomes zero, it corresponds to the gambler's ruin and that instant represents another stopping time (Time to go home for the gambler!)

Recall that in a Poisson process the occurrences of the first, second, $\cdots$ arrivals correspond to stopping times $T_1, T_2, \cdots$. Stopping times refer to those random instants at which there is sufficient information to decide whether or not a specific condition is satisfied.

Stopping Time: The random variable T is a stopping time for the process $X(t)$, if for all $t \geq 0$, the event $\{T \leq t\}$ is a function of the values $\{X(\tau) \mid \tau > 0, \tau \leq t\}$ of the process up to t, i.e., it should be possible to decide whether T has occurred or not by the time t, knowing *only* the value of the process $X(t)$ up to that time t. Thus the Poisson arrival times T_1 and T_2 referred above are stopping times; however $T_2 - T_1$ is not a stopping time.

A key result in martingales states that so long as

T is a stopping time (under some additional mild restrictions)

$$E\{X_T\} = E\{X_0\}. \qquad (20\text{-}42)$$

Notice that (20-42) generalizes (20-28) to certain random time instants (stopping times) as well.

Eq. (20-42) is an extremely useful tool in analyzing martingales. We shall illustrate its usefulness by rederiving the gambler's ruin probability in Example 3-15, Eq. (3-47), Text.

From Example 20.6, Y_n in (20-24) refer to a martingale in the gambler's ruin problem. Let T refer to the random instant at which the game ends; i.e., the instant at which either player A loses all his wealth and P_a is the associated probability of ruin for player A, or player A gains all wealth \$$(a + b)$ with probability $(1 - P_a)$. In that case, T is a stopping time and hence from (20-42), we get

$$E\{Y_T\} = E\{Y_0\} = \left(\frac{q}{p}\right)^a \tag{20-43}$$

since player A starts with \$a in Example 3.15. But

$$E\{Y_T\} = \left(\frac{q}{p}\right)^0 P_a + \left(\frac{q}{p}\right)^{a+b}(1 - P_a)$$

$$= P_a + \left(\frac{q}{p}\right)^{a+b}(1 - P_a). \tag{20-44}$$

Equating (20-43)-(20-44) and simplifying we get

$$P_a = \frac{1 - \left(\frac{p}{q}\right)^b}{1 - \left(\frac{p}{q}\right)^{a+b}} \tag{20-45}$$

that agrees with (3-47), Text. Eq. (20-45) can be used to derive other useful probabilities and advantageous plays as well. [see Examples 3-16 and 3-17, Text].

Whatever the advantage, it is worth quoting the master Gerolamo Cardano (1501-1576) on this: "*The greatest advantage in gambling comes from not playing at all*."