Things to note :

- **The Kendall–Lee Notation for Queuing Systems**

1) The first characteristic specifies the nature of the arrival process. The following standard abbreviations are used:

M = Service times are iid and exponentially distributed.
D = Service times are iid and deterministic.
E_k = Service times are iid Erlangs with shape parameter k.
G = Service times are iid and follow some general distribution.

2) The second characteristic specifies the nature of the service times:

M = Interarrival times are independent, identically distributed (iid) random variables having an exponential distribution.
D = Interarrival times are iid and deterministic.
E_k = Interarrival times are iid Erlangs with shape parameter k.
GI = Interarrival times are iid and governed by some general distribution.

3) The third characteristic is the number of parallel servers.
4) The fourth characteristic describes the queue discipline:

FCFS = First come, first served
LCFS = Last come, first served
SIRO = Service in random order
GD = General queue discipline

5) The fifth characteristic specifies the maximum allowable number of customers in the system (including customers who are waiting and customers who are in service).
6) The sixth characteristic gives the size of the population from which customers are drawn. Unless the number of potential customers is of the same order of magnitude as the number of servers, the population size is considered to be infinite.

Laws of Motion for Birth–Death Processes:

Law 1 With probability $\lambda_j \Delta t + o(\Delta t)$, a birth occurs between time t and time $t + \Delta t$.[†] A birth increases the system state by 1, to $j + 1$. The variable λ_j is called the **birth rate** in state j. In most queuing systems, a birth is simply an arrival.

Law 2 With probability $\mu_j \Delta t + o(\Delta t)$, a death occurs between time t and time $t + \Delta t$. A death decreases the system state by 1, to $j - 1$. The variable μ_j is the **death rate** in state j. In most queuing systems, a death is a service completion. Note that $\mu_0 = 0$ must hold, or a negative state could occur.

Law 3 Births and deaths are independent of each other.

Laws 1–3 can be used to show that the probability that more than one event (birth or death) occurs between t and $t + \Delta t$ is $o(\Delta t)$. Note that any birth–death process is completely specified by knowledge of the birth rates λ_j and the death rates μ_j. Since a negative state cannot occur, any birth–death process must have $\mu_0 = 0$.

Symbols:

λ = average number of arrivals *entering* the system per unit time
L = average number of customers present in the queuing system
L_q = average number of customers waiting in line
L_s = average number of customers in service
W = average time a customer spends in the system
W_q = average time a customer spends in line
W_s = average time a customer spends in service

In these definitions, all averages are steady-state averages. For most queuing systems, Little's queuing formula may be summarized as in Theorem 3.

THEOREM 3

For *any* queuing system in which a steady-state distribution exists, the following relations hold:

$$L = \lambda W \quad (28)$$
$$L_q = \lambda W_q \quad (29)$$
$$L_s = \lambda W_s \quad (30)$$

Week 6

Section 20.6

The M/M/s/GD/∞/∞ Queuing System:

- Interarrival times are exponential.
- Service times are exponential.
- A single line of customers waiting to be served.
- Parallel servers.

Example：Banks and post office branches in which all customers wait in a single line for service can often be modeled as M/M/s/GD/∞/∞ queuing systems.

ρ = traffic intensity
λ = arrival rate
μ = service rate
s = number of servers

Formulas :

1) We call π_j the steady state, or equilibrium probability, of state j.

$$\pi_0 = \frac{1}{\sum_{i=0}^{(s-1)} \frac{(s\rho)^i}{i!} + \frac{(s\rho)^s}{s!(1-\rho)}}$$

$$\pi_j = \frac{(s\rho)^j \pi_0}{j!} \qquad (j = 1, 2, \ldots, s)$$

$$\pi_j = \frac{(s\rho)^j \pi_0}{s! s^{j-s}} \qquad (j = s, s+1, s+2, \ldots)$$

$$\lambda_j = \lambda \qquad (j = 0, 1, \ldots)$$
$$\mu_j = j\mu \qquad (j = 0, 1, \ldots, s)$$
$$\mu_j = s\mu \qquad (j = s+1, s+2, \ldots)$$

2) $\rho = \frac{\lambda}{s\mu}$. For $\rho < 1$,

The steady-state probability that all servers are busy is given by:

3) $$P(j \geq s) = \frac{(s\rho)^s \pi_0}{s!(1-\rho)}$$

- If ρ >= 1 no steady state exists. In other words, if the arrival rate is at least as large as the maximum possible service rate (λ >= sμ) the system “blows up.” **This is Important for this chapter,** since we use this formula -> $\rho = \frac{\lambda}{s\mu}$. For $\rho < 1$, to determine most of the variables.

  Since we need ρ to be smaller than 1, we can use this to determine an unknown variable.

$$L_q = \frac{P(j \geq s)\rho}{1 - \rho}$$

$$W_q = \frac{L_q}{\lambda} = \frac{P(j \geq s)}{s\mu - \lambda}$$

$$\begin{aligned} W &= \frac{L}{\lambda} \\ &= \frac{L_q}{\lambda} + \frac{1}{\mu} \\ &= W_q + \frac{1}{\mu} \\ &= \frac{P(j \geq s)}{s\mu - \lambda} + \frac{1}{\mu} \end{aligned}$$

- the probability that a given customer will not wait in queue = P(W > t). where t is time.

$$P(\mathbf{W} > t) = e^{-\mu t}\left\{1 + P(j \geq s)\,\frac{1 - \exp\left[-\mu t(s - 1 - s\rho)\right]}{s - 1 - s\rho}\right\}$$

- the probability that a customer will have to wait in line for more than t minutes.

$$P(\mathbf{W}_q > t) = P(j \geq s)\exp\left[-s\mu(1 - \rho)t\right]$$

QUESTIONS :

EXAMPLE 9 **Bank Tellers**

Consider a bank with two tellers. An average of 80 customers per hour arrive at the bank and wait in a single line for an idle teller. The average time it takes to serve a customer is 1.2 minutes. Assume that interarrival times and service times are exponential. Determine

1 The expected number of customers present in the bank

2 The expected length of time a customer spends in the bank

3 The fraction of time that a particular teller is idle

Solution 1 We have an $M/M/2/GD/\infty/\infty$ system with $\lambda = 80$ customers per hour and $\mu = 50$ customers per hour. Thus $\rho = \frac{80}{2(50)} = 0.80 < 1$, so a steady state does exist. (For $\lambda \geq 100$, no steady state would exist.) From Table 6, $P(j \geq 2) = .71$. Then (41) yields

$$L_q = \frac{.80(.71)}{1 - .80} = 2.84 \text{ customers}$$

and from (43), $L = 2.84 + \frac{80}{50} = 4.44$ customers.

2 Since $W = \frac{L}{\lambda}$, $W = \frac{4.44}{80} = 0.055$ hour $= 3.3$ minutes.

3 To determine the fraction of time that a particular server is idle, note that he or she is idle during the entire time that $j = 0$ and half the time (by symmetry) that $j = 1$. The probability that a server is idle is given by $\pi_0 + 0.5\pi_1$. Using the fact that $P(j \geq 2) = .71$, we obtain π_0 from (40):

$$\pi_0 = \frac{s!P(j \geq s)(1 - \rho)}{(s\rho)^2} = \frac{2!(.71)(1 - .80)}{(1.6)^2} = .11$$

Now (39.1) yields

$$\pi_1 = \frac{(1.6)^1\pi_0}{1!} = .176$$

Thus, the probability that a particular teller is idle is $\pi_0 + 0.5\pi_1 = .11 + 0.5(.176) = .198$. We could have determined π_0 directly from (39):

$$\pi_0 = \frac{1}{} = \frac{1}{1 + 1.6 + 6.4} = \frac{1}{9}$$

This is consistent with our computation of $\pi_0 = .11$.

EXAMPLE 10 Bank Staffing

The manager of a bank must determine how many tellers should work on Fridays. For every minute a customer stands in line, the manager believes that a delay cost of 5¢ is incurred. An average of 2 customers per minute arrive at the bank. On the average, it takes a teller 2 minutes to complete a customer's transaction. It cost the bank \$9 per hour to hire a teller. Interarrival times and service times are exponential. To minimize the sum of service costs and delay costs, how many tellers should the bank have working on Fridays?

Solution Since $\lambda = 2$ customers per minute and $\mu = 0.5$ customer per minute, $\frac{\lambda}{s\mu} < 1$ requires that $\frac{4}{s} < 1$ or $s \geq 5$. Thus, there must be at least 5 tellers, or the number of customers present will "blow up." We now compute, for $s = 5, 6, \ldots,$

$$\underline{\text{Expected service cost}} + \underline{\text{expected delay cost}}$$

Since each teller is paid $\frac{9}{60} = 15¢$ per minute,

$$\underline{\text{Expected service cost}} = 0.15s$$

As in Example 4,

$$\underline{\text{Expected delay cost}} = \left(\underline{\text{expected customers}}\right)\left(\underline{\text{expected delay cost}}\right)$$

But

$$\underline{\text{Expected delay cost}} = 0.05W_q$$

Since an average of 2 customers arrive per minute,

$$\underline{\text{Expected delay cost}} = 2(0.05W_q) = 0.10W_q$$

For $s = 5$, $\rho = \frac{2}{.5(5)} = .80$ and $P(j \geq 5) = .55$. From (42),

$$W_q = \frac{.55}{5(.5) - 2} = 1.1 \text{ minutes}$$

Thus, for $s = 5$,

$$\underline{\text{Expected delay cost}} = 0.10(1.1) = 11¢$$

and, for $s = 5$,

$$\underline{\text{Total expected cost}} = 0.15(5) + 0.11 = 86¢$$

Since $s = 6$ has a service cost per minute of $6(0.15) = 90¢$, 6 tellers cannot have a lower total cost than 5 tellers. Hence, having 5 tellers serve is optimal. Putting it another way, adding an additional teller can save the bank at most 11¢ per minute in delay costs. Since an additional teller cost 15¢ per minute, it cannot be optimal to hire more than 5 tellers.

START USING LINGO !!!

Using LINGO for *M/M/s/GD/∞/∞* Computations

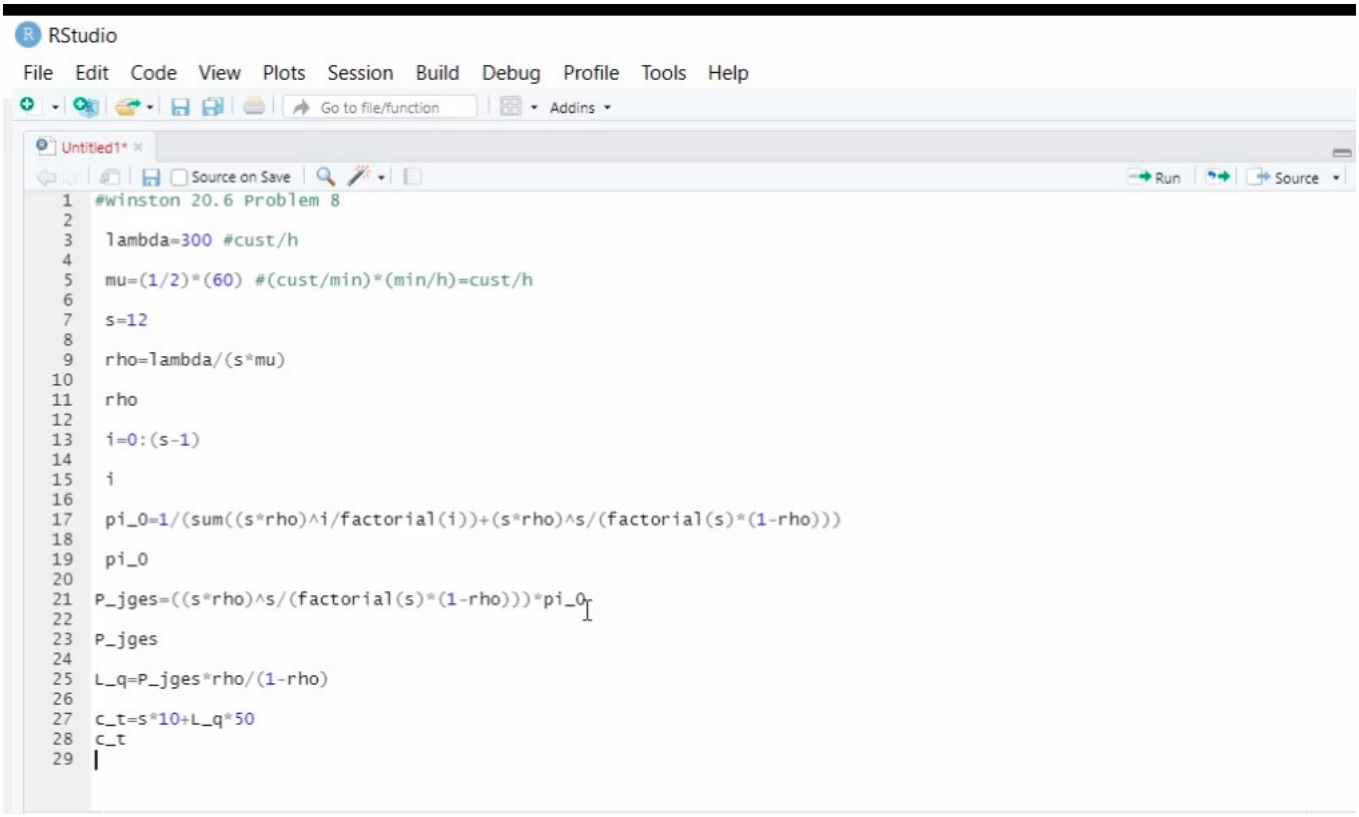
The LINGO function **@PEB( )** yields the probability that all servers are busy ($P(j \geq s)$) for an $M/M/s/GD/\infty/\infty$ system. The **@PEB** function has two arguments: the first is the value of λ/μ and the second is the number of servers. Thus, for Example 9, **@PEB** (80/50,2) = .711111 yields $P(j \geq 2)$.

The **@PEB** function can be used to solve queuing optimization problems with LINGO. For instance, to determine the cost-minimizing number of servers in Example 10, we would input the following problem into LINGO:

```
MODEL:
1) MIN=.10*@PEB(4,S)/(.5*S-2) + .15*S;
2) S>5;
END
```

In line 1 .10*@PEB(4,S)/(.5*S−2) is the expected cost per minute due to customers waiting in line, while .15*S is the per-minute service cost. Line 2 follows, because we need at least 5 servers for a steady state to exist. LINGO outputs $S = 5$ with an objective function value of .860823 (this is expected cost per minute).

8 An average of 300 customers per hour arrive at a huge branch of bank 2. It takes an average of 2 minutes to serve each customer. It costs $10 per hour to keep open a teller window, and the bank estimates that it will lose $50 in future profits for each hour that a customer waits in line. How many teller windows should bank 2 open?



A better way of doing it in terms of functions :

```
#Winston 20.6 Problem 8

lambda=300 #cust/h

mu=(1/2)*60 #(cust/min)*(min/h)=cust/h


rho=function(s)return(lambda/(s*mu))


pi_0=function(s){i=0:(s-1)
return(1/(sum((s*rho(s))^i/factorial(i))+(s*rho(s))^s/
(factorial(s)*(1-rho(s)))))}

P_jges=function(s)return((s*rho(s))^s/(factorial(s)*(1-
rho(s)))*pi_0(s))

L_q=function(s)return(P_jges(s)*rho(s)/(1-rho(s)))

ct=function(s)return(s*10+L_q(s)*50)

myct=sapply(s,ct)

myDF=data.frame(s,myct)

myDF

cat("The minimum total cost of $",min(myct),"per hour will be
attained by appointing",s[match(min(myct),myct)],"tellers.")
```

20.7 The M/G/∞/GD/∞/∞ and GI/G/∞/GD/∞/∞ Models:

- Systems in which a customer never has to wait for service to begin.(***infinite-server system***)
- Customer's entire stay in the system is service time.
  - (server available for each arrival)

System operates as follows:

1) Interarrival times are iid with common distribution A.
2) When a customer arrives, he or she immediately enters service.

- *Examples*: Industry, College program.

Formulas:

Let L be the expected number of customers in the system in the steady state

$$L = \frac{\lambda}{\mu}$$

And W be the expected time that a customer spends in the system.

$$W = \frac{1}{\mu}.$$

The steady-state probability that j customers are present :

$$\pi_j = \frac{(\frac{\lambda}{\mu})^j e^{-\lambda/\mu}}{j!}$$

EXAMPLE 11 Smalltown Ice Cream Shops

During each year, an average of 3 ice cream shops open up in Smalltown. The average time that an ice cream shop stays in business is 10 years. On January 1, 2525, what is the average number of ice cream shops that you would find in Smalltown? If the time between the opening of ice cream shops is exponential, what is the probability that on January 1, 2525, there will be 25 ice cream shops in Smalltown?

Solution We are given that $\lambda = 3$ shops per year and $\frac{1}{\mu} = 10$ years per shop. Assuming that the steady state has been reached, there will be an average of $L = \lambda(\frac{1}{\mu}) = 3(10) = 30$ shops in Smalltown. If interarrivals of ice cream shops are exponential, then

$$\pi_{25} = \frac{(30)^{25}e^{-30}}{25!} = .05$$

Of course, we could also compute the probability that there are 25 ice cream shops with the Excel formula

$$=\text{POISSON}(30,25,0)$$

This yields .045.

20.9 Finite Source Models: The Machine Repair Model

- arrival rates that are independent of the state of the system

Examples :

1) If customers do not want to wait in long lines, the arrival rate may be a decreasing function of the number of people present in the queuing system.

2) If arrivals to a system are drawn from a small population, the arrival rate may greatly depend on the state of the system. (**finite source models**).

Formulas:

The total rate of arrivals when the state is j,

$$\lambda_j = \underbrace{\lambda + \lambda + \cdots + \lambda}_{(K-j)\lambda\text{'s}} = (K-j)\lambda$$

and the death rate when the state is j,

$\mu_j = j\mu \qquad (j = 0, 1, \ldots, R)$,where R is the max number of servers (j <= R)

$\mu_j = R\mu \qquad (j = R+1, R+2, \ldots, K)$,where K is the max population and therefore we **use this one** when we ask the death rate for when **j is more than the servers**.

define $\rho = \frac{\lambda}{\mu}$,

steady-state probability distribution at state j :

$$\pi_j = \binom{K}{j} \rho^j \pi_0 \qquad (j = 0, 1, \ldots, R)$$

$$= \frac{\binom{K}{j} \rho^j j! \pi_0}{R! R^{j-R}} \qquad (j = R+1, R+2, \ldots, K)$$

Where $\binom{K}{j} = \frac{K!}{j!(K-j)!}$

To get π_0 we need to get all the other π_j's and use the formula :

$$\pi_0 + \pi_1 + \cdots + \pi_k = 1.$$

L = expected number of broken machines

$$L = \sum_{j=0}^{j=K} j\pi_j$$

L_q = expected number of machines waiting for service

$$L_q = \sum_{j=R}^{j=K} (j - R)\pi_j$$

the average number of arrivals per unit time is given by :

$$\bar{\lambda} = \sum_{j=0}^{j=K} \pi_j \lambda_j = \sum_{j=0}^{j=K} \lambda(K - j)\pi_j = \lambda(K - L)$$

W = average time a machine spends broken (down time)

$$W = \frac{L}{\bar{\lambda}}$$

W_q = average time a machine spends waiting for service

$$W_q = \frac{L_q}{\bar{\lambda}}$$

EXAMPLE 12 Patrol Cars

The Gotham Township Police Department has 5 patrol cars. A patrol car breaks down and requires service once every 30 days. The police department has two repair workers, each of whom takes an average of 3 days to repair a car. Breakdown times and repair times are exponential.

1 Determine the average number of police cars in good condition.

2 Find the average down time for a police car that needs repairs.

3 Find the fraction of the time a particular repair worker is idle.

Solution This is a machine repair problem with $K = 5$, $R = 2$, $\lambda = \frac{1}{30}$ car per day, and $\mu = \frac{1}{3}$ car per day. Then

From (52),

$$\pi_1 = \binom{5}{1}\left(\frac{1}{10}\right)\pi_0 = .5\pi_0$$

$$\pi_2 = \binom{5}{2}\left(\frac{1}{10}\right)^2\pi_0 = .1\pi_0$$

$$\pi_3 = \binom{5}{3}\left(\frac{1}{10}\right)^3\frac{3!}{2!2}\pi_0 = .015\pi_0 \tag{58}$$

$$\pi_4 = \binom{5}{4}\left(\frac{1}{10}\right)^4\frac{4!}{2!(2)^2}\pi_0 = .0015\pi_0$$

$$\pi_5 = \binom{5}{5}\left(\frac{1}{10}\right)^5\frac{5!}{2!(2)^3}\pi_0 = .000075\pi_0$$

Then $\pi_0(1 + .5 + .1 + .015 + .0015 + .000075) = 1$, or $\pi_0 = .619$. Now (58) yields $\pi_1 = .310$, $\pi_2 = .062$, $\pi_3 = .009$, $\pi_4 = .001$, and $\pi_5 = 0$.

1 The expected number of cars in good condition is $K - L$, which is given by

$$K - \sum_{j=0}^{j=5} j\pi_j = 5 - [0(.619) + 1(.310) + 2(.062) + 3(.009) + 4(.001) + 5(0)]$$

$$= 5 - .465 = 4.535 \text{ cars in good condition}$$

2 We seek $W = \frac{L}{\bar{\lambda}}$. From (55),

$$\bar{\lambda} = \sum_{j=0}^{j=5}\lambda(5 - j)\pi_j = \frac{1}{30}(5\pi_0 + 4\pi_1 + 3\pi_2 + 2\pi_3 + \pi_4 + 0\pi_5)$$

$$= \frac{1}{30}[5(.619) + 4(.310) + 3(.062) + 2(.009) + 1(.001) + 0(0)]$$

$$= 0.151 \text{ car per day}$$

or

$$\bar{\lambda} = \lambda(K - L) = \frac{4.535}{30} = 0.151 \text{ car per day}$$

Since $L = 0.465$ car, we find that $W = \frac{0.465}{.0151} = 3.08$ days.

3 The fraction of the time that a particular repair worker will be idle is $\pi_0 + 0.5\pi_1 = .619 + .5(.310) = .774$.

If there were three repair people, the fraction of the time that a particular server would be idle would be $\pi_0 + (\frac{2}{3})\pi_1 + (\frac{1}{3})\pi_2$, and for a repair staff of R people, the probability that a particular server would be idle is given by

$$\pi_0 + \frac{(R-1)\pi_1}{R} + \frac{(R-2)\pi_2}{R} + \cdots + \frac{\pi_{R-1}}{R}$$

Using LINGO for Machine Repair Model Computations

The LINGO function **@PFS**(K*λ/μ,R,K) will yield L, the expected number (in the steady state) of machines in bad condition. The FS stands for Finite Source. Thus, for Example 12, **@PFS**(5*(1/30)/(1/3),2,5) will yield .465.