Modeling and Simulation
NET 361

Lecture 3

Lecturer /Maysoon R. Isleem

Algorithm

- The key point in algorithm development is that if the queue discipline is FIFO then the truth of the expression ai < ci−1 determines whether or not job i will experience a delay.
- An equation can be written for the delay that depends on the interarrival and service times only. That is

$$\begin{aligned} c_{i-1} - a_i &= (a_{i-1} + d_{i-1} + s_{i-1}) - a_i \\ &= d_{i-1} + s_{i-1} - (a_i - a_{i-1}) \\ &= d_{i-1} + s_{i-1} - r_i. \end{aligned}$$

If $d_0 = s_0 = 0$ then $d_1, d_2, d_3, \ldots$ are defined by the nonlinear equation

$$d_i = \max\{\, 0,\; d_{i-1} + s_{i-1} - r_i \,\} \qquad\qquad i = 1, 2, 3, \ldots$$

This equation is commonly used in theoretical studies to analyze the stochastic behavior of a FIFO service node.

Algorithm

- If the arrival times a1, a2,… and service times s1, s2,… are known and if the server is initially idle, then this algorithm computes the delays d1, d2, … in a single-server FIFO service node with infinite capacity.

```
c_0 = 0.0;                                  /* assumes that a_0 = 0.0 */
i = 0;
while ( more jobs to process ) {
    i++;
    a_i = GetArrival();
    if (a_i < c_{i-1})
      d_i = c_{i-1} - a_i;                  /* calculate delay for job i */
    else
      d_i = 0.0;                            /* job i has no delay */
    s_i = GetService();
    c_i = a_i + d_i + s_i;                  /* calculate departure time for job i */
}
n = i;
return d_1, d_2, ..., d_n;
```

Algorithm

The GetArrival and GetService procedures read the next arrival and service time from a file.

Example: Algorithm 1.2.1 used to process n = 10 jobs

	i	:	1	2	3	4	5	6	7	8	9	10
read from file	a_i	:	15	47	71	111	123	152	166	226	310	320
from algorithm	d_i	:	0	11	23	17	35	44	70	41	0	26
read from file	s_i	:	43	36	34	30	38	40	31	29	36	30

For future reference, note that the last job arrived at time $a_n = 320$ and departed at time $c_n = a_n + d_n + s_n = 320 + 26 + 30 = 376$.

Output Statistics

- The purpose of simulation is insight and we gain insight about the performance of a system by looking at meaningful statistics (what statistics should be generated).

- The importance of various statistics varies on perspective:
  - Job perspective: wait time is most important

  Ex : the most important statistic might be the average delay
  - Manager perspective: utilization is critical
- Statistics are broken down into two categories:
  - Job-averaged statistics
  - Time-averaged statistics

Job-Averaged Statistics

- Job-averaged statistics: computed via typical arithmetic mean
  - **<u>Average interarrival time:</u>**

$$\bar{r} = \frac{1}{n}\sum_{i=1}^{n} r_i = \frac{a_n}{n}$$

  - The reciprocal of the average interarrival time, $1/\bar{r}$., is the arrival rate;
  - **<u>Average service time:</u>**

$$\bar{s} = \frac{1}{n}\sum_{i=1}^{n} s_i.$$

  - the reciprocal of the average service time, $1/\bar{s}$., is the service rate.

Output statistics

- For the 10 jobs in the previous example
- average interarrival time is

$$\bar{r} = a_n/n = 320/10 = 32.0 \text{ seconds per job}$$

average service is $\bar{s} = 34.7$ seconds per job
arrival rate is $1/\bar{r} \approx 0.031$ jobs per second
service rate is $1/\bar{s} \approx 0.029$ jobs per second

- The server is not quite able to process jobs at the rate they arrive on average.

Job-Averaged Statistics

- The average delay and average wait are defined as

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i \qquad \text{and} \qquad \bar{w} = \frac{1}{n}\sum_{i=1}^{n} w_i.$$

- Recall wi = di + si for all i

$$\overline{w} = \frac{1}{n}\sum_{i=1}^{n} w_i = \frac{1}{n}\sum_{i=1}^{n}(d_i + s_i) = \frac{1}{n}\sum_{i=1}^{n} d_i + \frac{1}{n}\sum_{i=1}^{n} s_i = \overline{d} + \overline{s}$$

- The point here is that it is sufficient to compute any two of the statistics $\bar{w}$, $\bar{d}$, $\bar{s}$. The third statistic can then be computed from the other two, if appropriate.

Job-Averaged Statistics

- From the data in the previous example, $\bar{d} = 26.7$ $\bar{s} = 34.7$, Therefore $\bar{w} = 26.7 + 34.7 = 61.4$.
- Recall verification is one (difficult) step of model development
- Consistency check: used to verify that a simulation satisfies known equations:

Compute $\overline{w}$, $\overline{d}$, and $\overline{s}$ *independently*
Then verify that $\overline{w} = \overline{d} + \overline{s}$

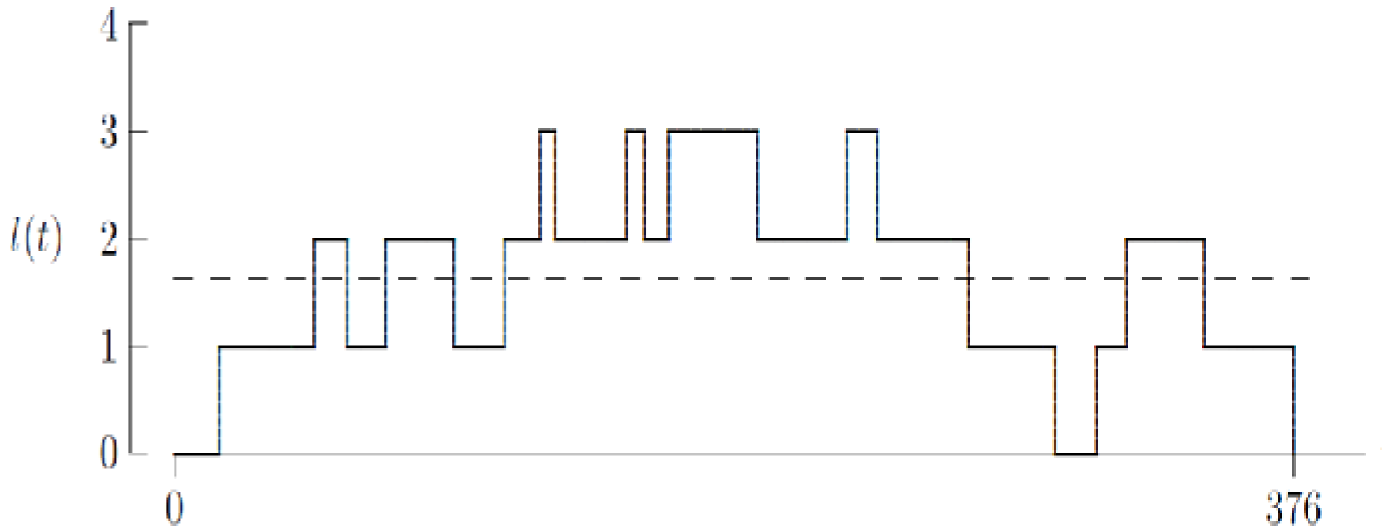
Job-Averaged Statistics

- Time-averaged statistics: defined by area under a curve (integration)
- For SSQ, need three additional functions
  - $l(t)$: number of jobs in the service node at time t.
  - $q(t)$: number of jobs in the queue at time t.
  - $x(t)$: number of jobs in service at time t.

- By definition, $l(t) = q(t) + x(t)$.
  - $l(t) = 0, 1, 2, \ldots$
  - $q(t) = 0, 1, 2, \ldots$
  - $x(t) = 0, 1$

Time-Averaged Statistics

- All three functions are piece-wise constant



- Figures for $q(\cdot)$ and $x(\cdot)$ can be deduced
  - $q(t) = 0$ and $x(t) = 0$ if and only if $l(t) = 0$

Time–Averaged Statistics

- Over the time interval $(0, T)$:

$$\text{time-averaged number in the node:} \quad \bar{l} = \frac{1}{\tau} \int_0^{\tau} l(t)dt$$

$$\text{time-averaged number in the queue:} \quad \bar{q} = \frac{1}{\tau} \int_0^{\tau} q(t)dt$$

$$\text{time-averaged number in service:} \quad \bar{x} = \frac{1}{\tau} \int_0^{\tau} x(t)dt$$

- Since $l(t) = q(t) + x(t)$ for all $t > 0$

$$\bar{l} = \bar{q} + \bar{x}$$

- Sufficient to calculate any two of $\bar{l}, \bar{q}, \bar{x}$

Time–Averaged Statistics

- The average of numerous random observations (samples) of the number in the service node should be close to $\bar{l}$.

  Same holds for $\bar{q}$ and $\bar{x}$

- Server utilization: time-averaged number in service $(\bar{x})$
- $(\bar{x})$ also represents the probability the server is busy.

Time–Averaged Statistics

- If we were to observe (sample) the number in the service node, for example, at many different times chosen *at random* between 0 and τ and then calculate the arithmetic average of all these observations, the result should be close to $\bar{l}$.
- Similarly, the arithmetic average of many random observations of the number in the queue should be close to $\bar{q}$ and the arithmetic average of many random observations of the number in service (0 or 1) should be close to $\bar{x}$.
- $\bar{x}$ must lie in the closed interval $[0, 1]$.

Definition 1.2.7 The time-averaged number in service $\bar{x}$ is also known as the server *utilization*. The reason for this terminology is that $\bar{x}$ represents the proportion of time that the server is busy.

Time–Averaged Statistics

Equivalently, if one particular time is picked *at random* between 0 and τ then $\bar{x}$ is the probability that the server is busy at that time. If $\bar{x}$ is close to 1.0 then the server is busy most of the time and correspondingly large values of $\bar{l}$ and $\bar{q}$ will be produced. On the other hand, if the utilization is close to 0.0 then the server is idle most of the time and the values of $\bar{l}$ and $\bar{q}$ will be small.

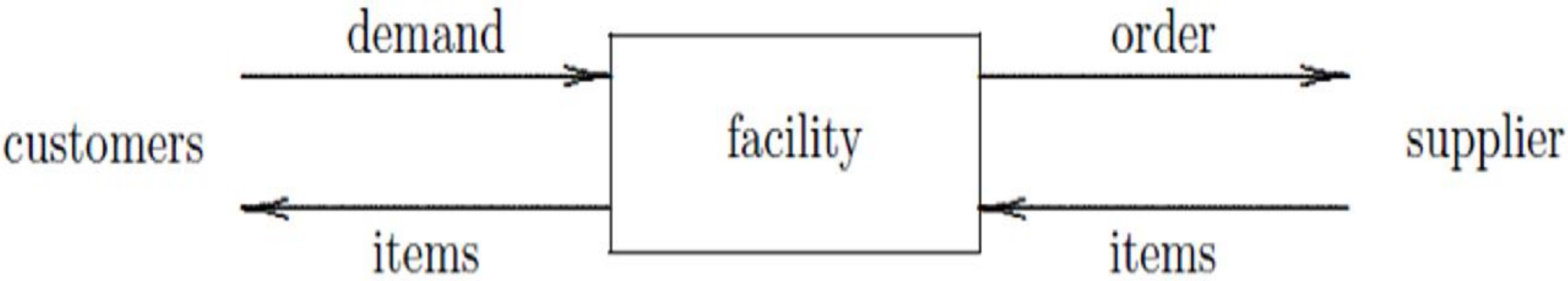
One important issue remains — how are job averages and time averages related? Specifically, how are $\bar{w}$, $\bar{d}$, and $\bar{s}$ related to $\bar{l}$, $\bar{q}$, and $\bar{x}$?

Computational Model

- As discussed previously, by using Algorithm 1.2.1 in conjunction with some statistics gathering logic it is a straight-forward programming exercise to produce a computational model of a single-server FIFO service node with infinite capacity. The ANSI C program ssq1 is an example, this program is designed with readability and extendibility considerations.

A simple inventory system

- CONCEPTUAL MODEL:



A simple inventory system

- An inventory system consists of a facility that distributes items from its current inventory to its customers in response to a customer demand that is typically random, the demand is integer-valued (discrete) because customers do not want a portion of an item.

- * Because there is a holding cost associated with items in inventory, it is undesirable for the inventory level to be too high.

- On the other hand, if the inventory level is too low, the facility is in danger of incurring a shortage cost whenever a demand occurs that cannot be met.

A simple inventory system

- As a policy, the inventory level is reviewed periodically and new items are ordered from a supplier, if necessary.
- ** When items are ordered, the facility incurs an ordering cost that is the sum of a fixed setup cost independent of the amount ordered plus an item cost proportional to the number of items ordered.
- This periodic inventory review policy is defined by two parameters, conventionally denoted s and S.

A simple inventory system

- s is the minimum inventory level | if at the time of review the current inventory level is below the threshold s then an order will be placed with the supplier to replenish the inventory. If the current inventory level is at or above s then no order will be placed.

- S is the maximum inventory level | when an order is placed, the amount ordered is the number of items required to bring the inventory back up to the level S.

- The $(s; S)$ parameters are constant in time with $0 \leq s < S$.

A simple inventory system

- Conceptual Model:

- **Periodic Inventory Review**
  - Inventory review is periodic
  - Items are ordered, if necessary, only at review times
  - (s, S) are the min,max inventory levels, $0 \leq s < S$
  - We assume periodic inventory review
  - Search for (s, S) that minimize cost
- **Transaction Reporting**
  - Inventory review after each transaction
  - Significant labor may be required
  - Less likely to experience shortage

A simple inventory system

Conceptual Model:

- **Inventory System Costs**
  - Holding cost: for items in inventory
  - Shortage cost: for unmet demand
  - Setup cost: fixed cost when order is placed
  - Item cost: per-item order cost
  - Ordering cost: sum of setup and item costs
- **Additional Assumptions**
  - Back ordering is possible
  - No delivery lag
  - Initial inventory level is S
  - Terminal inventory level is S

A simple inventory system

Specification Model

Time begins at $t = 0$

- Review times are $t = 0, 1, 2, \ldots$
- l_{i-1} : inventory level at beginning of i^{th} interval
- o_{i-1} : amount ordered at time $t = i - 1$, $(o_{i-1} \geq 0)$
- d_i : demand quantity during i^{th} interval, $(d_i \geq 0)$
- Inventory at end of interval can be negative

(Because we have assumed that back ordering is possible, if the demand during the i^{th} time interval is greater than the inventory level at the beginning of the interval (plus the amount ordered, if any)

A simple inventory system

Inventory Level Considerations:

- Inventory level is reviewed at $t = i - 1$
- If at least s, no order is placed
- If less than s, inventory is replenished to S

$$o_{i-1} = \begin{cases} 0 & l_{i-1} \geq s \\ S - l_{i-1} & l_{i-1} < s \end{cases}$$

- Items are delivered immediately
- At end of i th interval, inventory diminished by di
- $l_i = l_{i-1} + o_{i-1} - d_i$

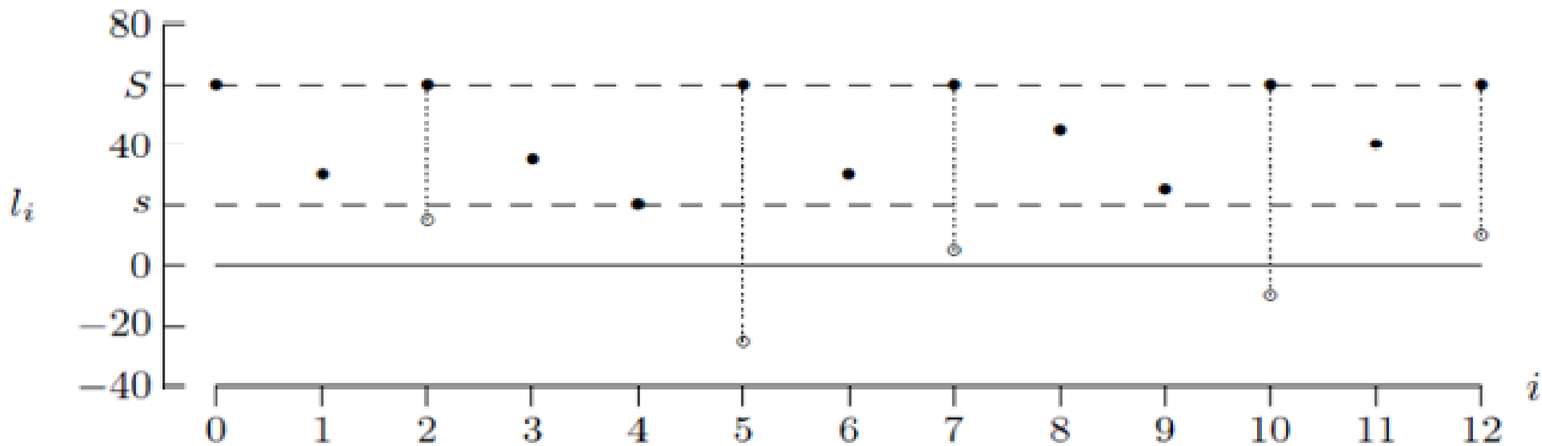
Time evolution of inventory level

```
l_0 = S;        /* the initial inventory level is S */
i = 0;
while (more demand to process ) {
    i++;
    if (l_{i-1} < s)
        o_{i-1} = S - l_{i-1};
    else
        o_{i-1} = 0;
    d_i = GetDemand();
    l_i = l_{i-1} + o_{i-1} - d_i;
}
n = i;
o_n = S - l_n;
l_n = S;        /* the terminal inventory level is S */
return l_1, l_2, ..., l_n and o_1, o_2, ..., o_n;
```

SIS with sample demands

- Let (s, S) = (20, 60) and consider n = 12 time intervals

i	:	1	2	3	4	5	6	7	8	9	10	11	12
d_i	:	30	15	25	15	45	30	25	15	20	35	20	30



Output statistics

- we must address the issue of what statistics should be computed to measure the performance of a simple inventory system.
- our objective is to analyze these statistics and, by so doing, better understand how the system operates.
- Average demand and average order

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i \qquad \bar{o} = \frac{1}{n}\sum_{i=1}^{n} o_i .$$

*For the example
$\bar{d} = \bar{o} = 305/12 \simeq 25.42$ items per time interval.

Flow Balance

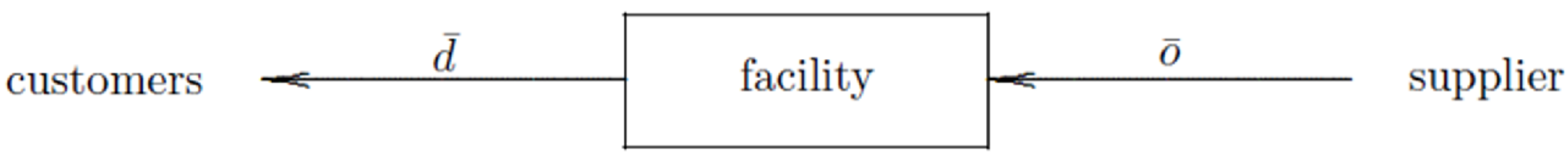
The terminal condition is that at the end of the n^{th} time interval an order is placed to return the inventory to its initial level. Because of this terminal condition, independent of the value of s and S, the average demand $\bar{d}$ and the average order $\bar{o}$ must be equal. That is, over the course of the simulated period of operation, all demand is satisfied (although not immediately when back ordering occurs).

the inventory level is the same at the beginning and end of the simulation then the average "flow" of items into the facility from the supplier, $\bar{o}$, must have been equal to the average "flow" of items out of the facility to the customers, $\bar{d}$. With respect to the flow of items into and out of the facility, the inventory system is said to be *flow balanced*.

Flow Balance

Average demand and order must be equal

- Ending inventory level is S
- Over the simulated period, all demand is satisfied
- Average "flow" of items in equals average "flow" of items out



- The inventory system is flow balanced

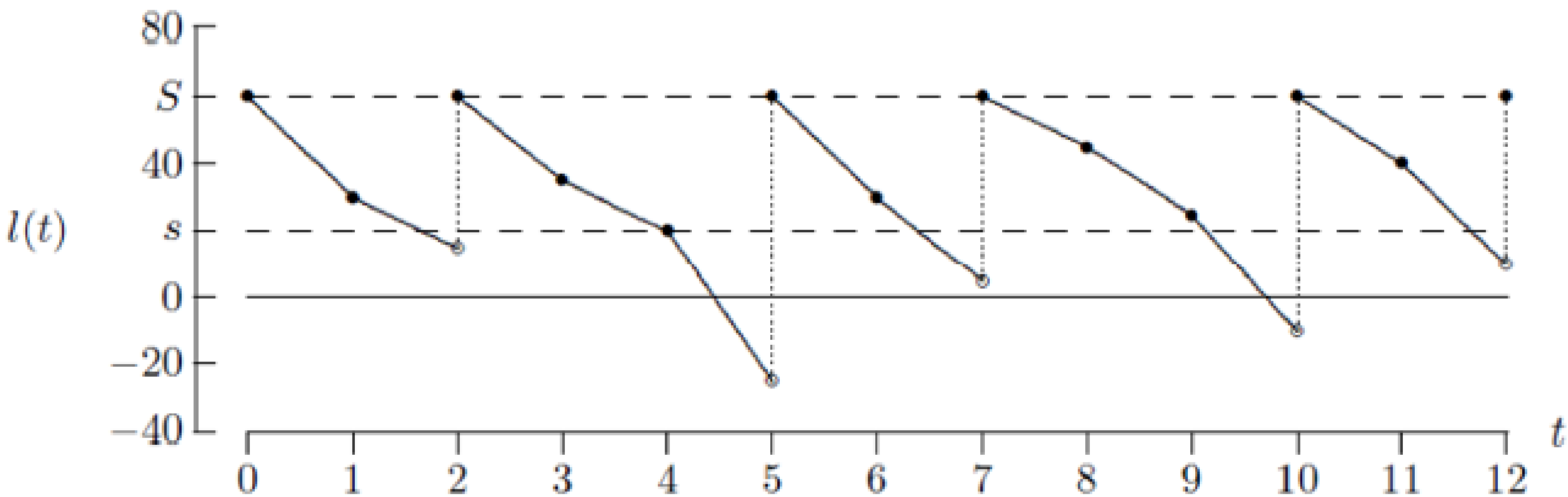
Constant Demand Rate

- The holding cost and shortage cost are proportional to time-averaged inventory levels .

- To compute these averages it is necessary to know the inventory level for all t, not just at the inventory review times.

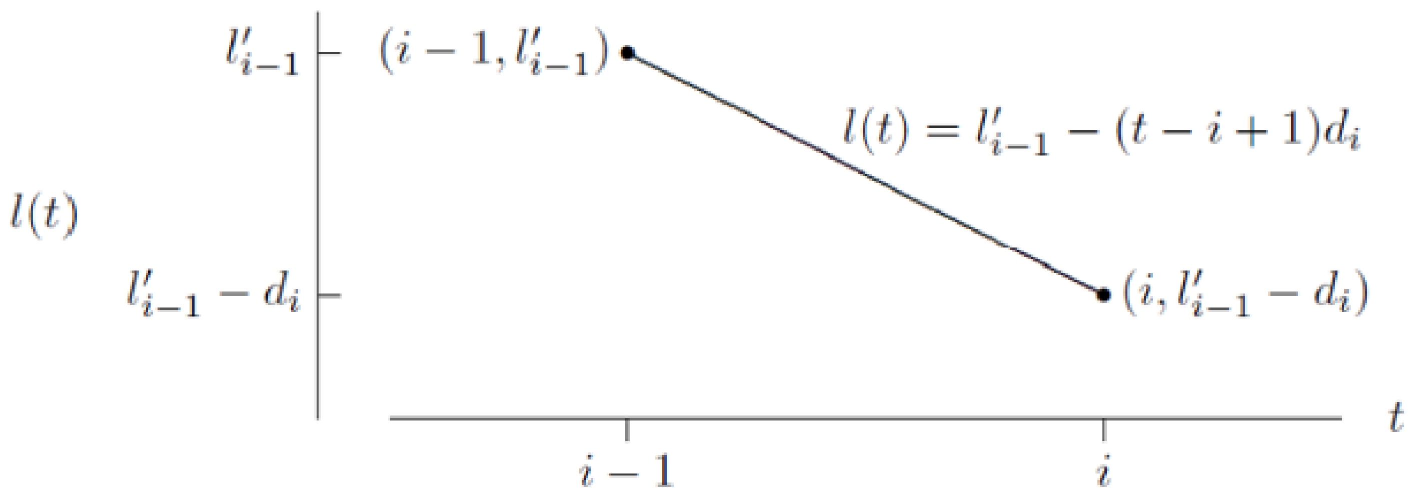
- Therefore, we assume that the demand rate is constant between review times so that the continuous time evolution of the inventory level is piecewise linear

Constant Demand Rate



Inventory Level As A Function Of Time

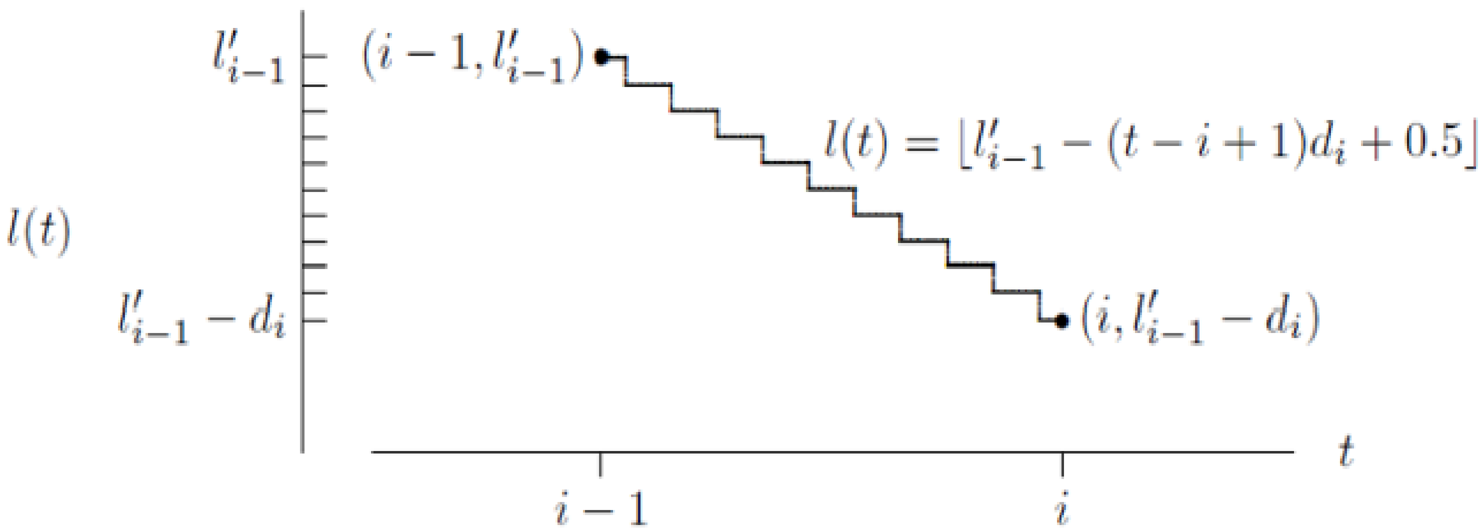
- The inventory level at any time t in i^{th} interval is
- $l(t) = l'_{i-1} - (t - i + 1)\, di$
- if demand rate is constant between review times



- $l'_{i-1} = l_{i-1} + o_{i-1}$ represents inventory level after review

Inventory Decrease Is Not Linear

- Inventory level at any time t is an integer
- l(t) should be rounded to an integer value
- l(t) is a stair-step, rather than linear, function



Time-Averaged Inventory Level

- l (t) is the basis for computing the time-averaged inventory level
  - Case 1: If l (t) remains non-negative over ith interval

$$\bar{I}_i^+ = \int_{i-1}^{i} I(t)dt$$

  - Case 2: If l (t) becomes negative at some time

$$\bar{I}_i^+ = \int_{i-1}^{\tau} I(t)dt \qquad \bar{I}_i^- = -\int_{\tau}^{i} I(t)dt$$

- $\bar{l}^+_i$ is the time-averaged holding level
- $\bar{l}^-_i$ is the time-averaged shortage level