The inputs to program `ssq1`, the arrival times and the service times, can have any positive real value — they are *continuous* variables. In some models, however, the input variables are inherently *discrete*. That is the case with the (trace-driven) discrete-event simulation model of a simple inventory system constructed in this section. As in the previous section, we begin with a conceptual model then move to a specification model and, finally, to a computational model.

1.3.1 CONCEPTUAL MODEL

Definition 1.3.1 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, as illustrated in Figure 1.3.1. Moreover, 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.

Figure 1.3.1. Simple inventory system diagram.

As a policy, the inventory level is reviewed periodically and new items are then (and only then) 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.

- 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 $0 \leq s < S$.

* Some inventory systems distribute "items" that are not inherently discrete, for example, a service station that sells gasoline. With minor modifications, the model developed in this section is applicable to these inventory systems as well.

** An alternate to the periodic inventory review policy is a *transaction reporting inventory policy*. With this policy, inventory review occurs after *each* demand instance. Because inventory review occurs more frequently, significantly more labor may be required to implement a transaction reporting inventory policy. (The scanners at a grocery store, however, require no extra labor). The transaction reporting policy has the desirable property that, for the same value of s, it is less likely for the inventory system to experience a shortage.

A discrete-event simulation model can be used to compute the cost of operating the facility. In some cases, the values of s and S are fixed; if so, the cost of operating the facility is also fixed. In other cases, if at least one of the (s, S) values (usually s) is not fixed, the cost of operating the facility can be modified and it is natural to search for values of (s, S) for which the cost of operating the facility is minimized.

To complete the conceptual model of this simple (one type of item) inventory system we make three additional assumptions. (a) *Back ordering* (backlogging) is possible — the inventory level can become negative in order to model customer demands not immediately satisfied. (b) There is no *delivery lag* — an order placed with the supplier will be delivered immediately. Usually this is an unrealistic assumption; it will be removed in Chapter 3. (c) The *initial* inventory level is S.

Example 1.3.5, presented later in this section, describes an automobile dealership as an example of an inventory system with back ordering and no delivery lag. In this example the periodic inventory review occurs each week. The value of S is fixed, the value of s is not.

1.3.2 SPECIFICATION MODEL

The following variables provide the basis for a specification model of a simple inventory system. Time begins at $t = 0$ and is measured in a coordinate system in which the inventory review times are $t = 0, 1, 2, 3, \ldots$ with the convention that the i^{th} time interval begins at time $t = i - 1$ and ends at $t = i$.

- The inventory level at the *beginning* of the i^{th} time interval is an integer l_{i-1}.
- The amount ordered (if any) at time $t = i - 1$ is an integer $o_{i-1} \geq 0$.
- The demand quantity *during* the i^{th} time interval is an integer $d_i \geq 0$.

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) then the inventory level at the end of the interval will be negative.

The inventory level is reviewed at $t = i - 1$. If l_{i-1} is greater than or equal to s then no items are ordered so that $o_{i-1} = 0$. If instead l_{i-1} is less than s then $o_{i-1} = S - l_{i-1}$ items are ordered to replenish inventory. In this case, because we have assumed there is no delivery lag, ordered items are delivered immediately (at $t = i - 1$) thereby restoring inventory to the level S. In either case, the inventory level at the end of the i^{th} time interval is diminished by d_i. Therefore, as summarized by Algorithm 1.3.1, with $l_0 = S$ the inventory orders $o_0, o_1, o_2, \ldots$ and corresponding inventory levels $l_1, l_2, \ldots$ are defined by

$$o_{i-1} = \begin{cases} 0 & l_{i-1} \geq s \\ S - l_{i-1} & l_{i-1} < s \end{cases} \qquad \text{and} \qquad l_i = l_{i-1} + o_{i-1} - d_i.$$

Note that $l_0 = S > s$ and so o_0 must be zero; accordingly, only $o_1, o_2, \ldots$ are of interest.

Algorithm 1.3.1 If the demands $d_1, d_2, \ldots$ are known then this algorithm computes the discrete time evolution of the inventory level for a simple (s, S) inventory system with back ordering and no delivery lag.

```
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;
```

Example 1.3.1 Let $(s, S) = (20, 60)$ and apply Algorithm 1.3.1 to process $n = 12$ time intervals of operation with the input demand schedule:

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

As illustrated in Figure 1.3.2, the time evolution of the inventory level typically features several intervals of decline, followed by an increase when an order is placed (indicated by the vertical dotted line) and, because there is no delivery lag, is immediately delivered.

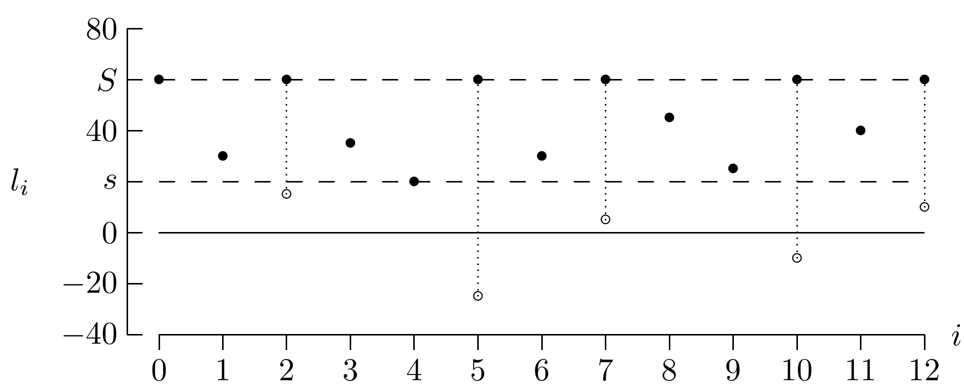


Figure 1.3.2. Inventory levels.

At the end of the last interval (at $t = n = 12$) an order for $o_n = 50$ inventory units was placed. The immediate delivery of this order restores the inventory level at the end of the simulation to the initial inventory level S, as shown in Figure 1.3.2.

1.3.3 OUTPUT STATISTICS

As with the development of program `ssq1` in the previous section, we must address the issue of what statistics should be computed to measure the performance of a simple inventory system. As always, our objective is to analyze these statistics and, by so doing, better understand how the system operates.

Definition 1.3.2 The *average demand* and *average order* are, respectively

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

Example 1.3.2 For the data in Example 1.3.1, $\bar{d} = \bar{o} = 305/12 \cong 25.42$ items per time interval. As explained next, these two averages *must* be equal.

The terminal condition in Algorithm 1.3.1 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). Therefore, if 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.*

Figure 1.3.3. Flow balance.

Average Inventory Level

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* as illustrated in Figure 1.3.4.

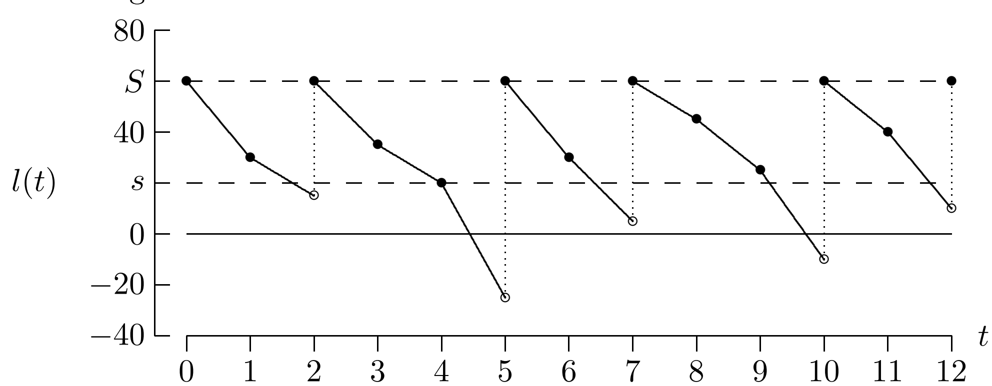
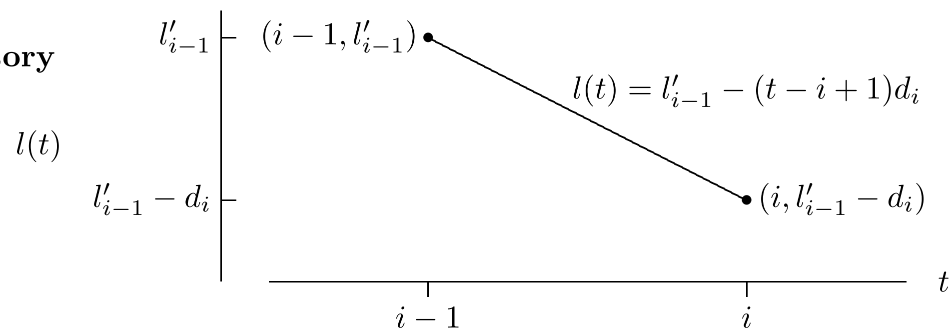


Figure 1.3.4. Piecewise linear inventory levels.

Definition 1.3.3 If the demand rate is constant between review times, then at any time t in the i^{th} time interval the inventory level is $l(t) = l'_{i-1} - (t - i + 1)d_i$, as illustrated in Figure 1.3.5.

Figure 1.3.5. Linear inventory level in time interval i.



In this figure and related figures and equations elsewhere in this section, $l'_{i-1} = l_{i-1} + o_{i-1}$ represents the inventory level *after* inventory review. Accordingly, $l'_{i-1} \geq s$ for all i. (For the figure in Example 1.3.1, the ∘'s and •'s represent l_{i-1} and l'_{i-1} respectively).

The equation for $l(t)$ is the basis for calculating the time-averaged inventory level for the i^{th} time interval.* There are two cases to consider. If $l(t)$ remains non-negative over the i^{th} time interval then there is only a time-averaged *holding level* integral to evaluate:

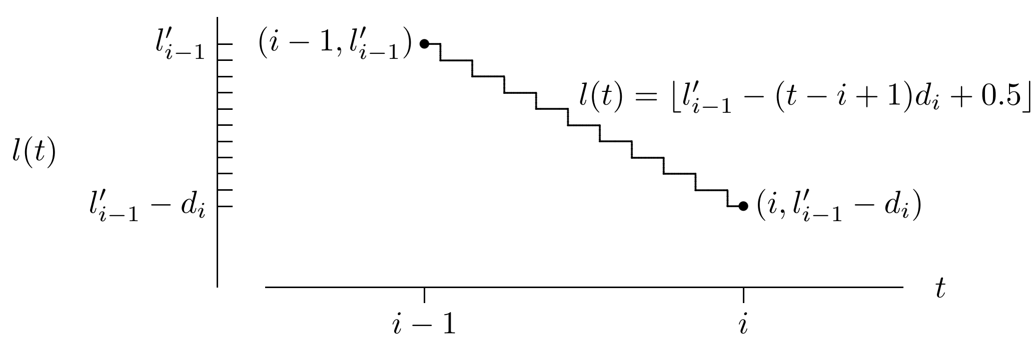
$$\bar{l}_i^+ = \int_{i-1}^{i} l(t)\,dt.$$

If instead $l(t)$ becomes negative at some time τ interior to the i^{th} interval then, in addition to a time-averaged holding level integral, there is also a time-averaged *shortage level* integral to evaluate. In this case the two integrals are

$$\bar{l}_i^+ = \int_{i-1}^{\tau} l(t)\,dt \qquad \text{and} \qquad \bar{l}_i^- = -\int_{\tau}^{i} l(t)\,dt.$$

* Because the inventory level at any time is an *integer*, the figure in Definition 1.3.3 is technically incorrect. Instead, rounding to an integer value should be used to produce the inventory level time history illustrated in Figure 1.3.6. ($\lfloor z \rfloor$ is the floor function; $\lfloor z + 0.5 \rfloor$ is z rounded to the nearest integer.)

Figure 1.3.6. Piecewise constant inventory level in time interval i.



It can be shown, however, that rounding has *no* effect on the value of $\bar{l}_i^+$ and $\bar{l}_i^-$.

No Back Ordering

The inventory level $l(t)$ remains non-negative throughout the i^{th} time interval if and only if the inventory level at the end of this interval is non-negative, as in Figure 1.3.7.

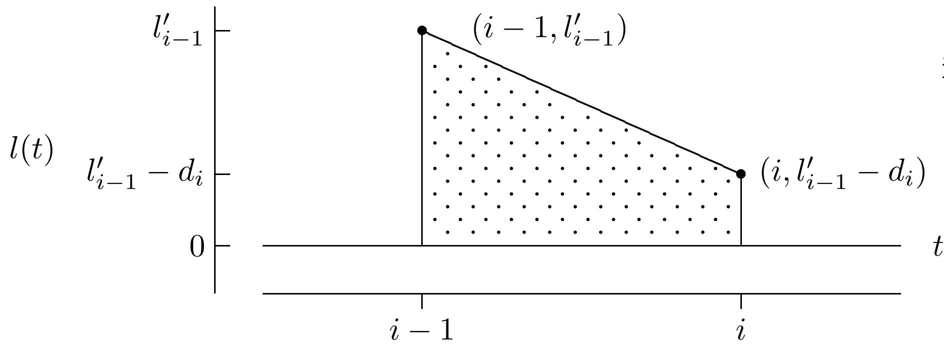


Figure 1.3.7. Inventory level in time interval i with no backordering.

Therefore, there is no shortage during the i^{th} time interval if and only if $d_i \le l'_{i-1}$. In this case the time-averaged holding level integral for the i^{th} time interval can be evaluated as the area of a trapezoid so that

$$\bar{l}_i^+ = \int_{i-1}^{i} l(t)\,dt = \frac{l'_{i-1} + (l'_{i-1} - d_i)}{2} = l'_{i-1} - \frac{1}{2}d_i \qquad \text{and} \qquad \bar{l}_i^- = 0.$$

With Back Ordering

The inventory level becomes negative at some point τ in the i^{th} time interval if and only if $d_i > l'_{i-1}$, as illustrated in Figure 1.3.8.

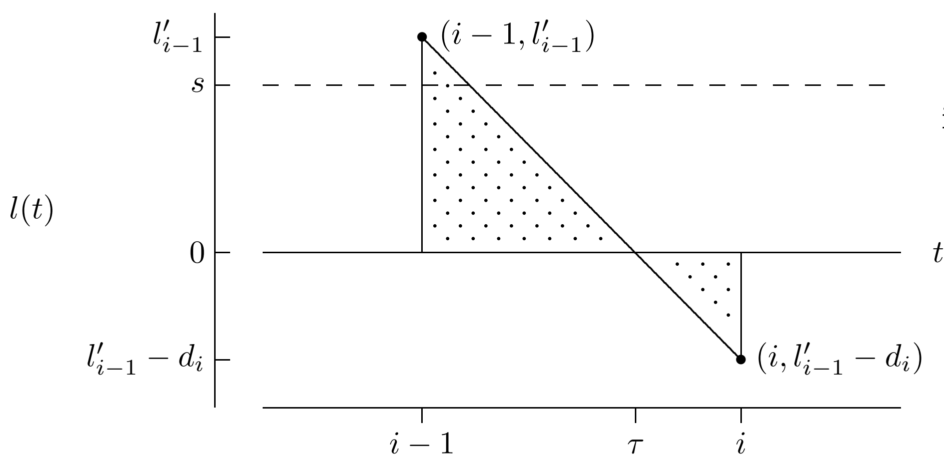


Figure 1.3.8. Inventory level in time interval i with backordering.

By using similar triangles, it can be shown that $\tau = i - 1 + (l'_{i-1}/d_i)$. In this case, the time-averaged holding level integral and shortage level integral for the i^{th} time interval can be evaluated as the area of a triangle so that

$$\bar{l}_i^+ = \int_{i-1}^{\tau} l(t)\,dt = \cdots = \frac{(l'_{i-1})^2}{2d_i} \qquad \text{and} \qquad \bar{l}_i^- = -\int_{\tau}^{i} l(t)\,dt = \cdots = \frac{(d_i - l'_{i-1})^2}{2d_i}.$$

The time-averaged holding level and shortage level for each time interval can be summed over all intervals with the resulting sums divided by the number of intervals. Consistent with Definition 1.3.4, the result represents the average number of items "held" and "short" respectively, with the average taken over all time intervals.

Definition 1.3.4 The *time-averaged holding level* and the *time-averaged shortage level* are, respectively

$$\bar{l}^+ = \frac{1}{n}\sum_{i=1}^{n} \bar{l}_i^+ \qquad \text{and} \qquad \bar{l}^- = \frac{1}{n}\sum_{i=1}^{n} \bar{l}_i^- .$$

It is potentially confusing to define the time-averaged shortage level as a *positive* number, as we have done in Definition 1.3.3. In particular, it would be a mistake to compute the time-averaged inventory level as the sum of $\bar{l}^+$ and $\bar{l}^-$. Instead, the *time-averaged inventory level* is the difference

$$\bar{l} = \frac{1}{n}\int_0^n l(t)\,dt = \bar{l}^+ - \bar{l}^- .$$

The proof of this result is left as an exercise.

Example 1.3.3 For the data in Example 1.3.1, $\bar{l}^+ = 31.74$ and $\bar{l}^- = 0.70$. Therefore, over the 12 time intervals, the average number of items held was 31.74, the average number of items short was 0.70, and the average inventory level was 31.04.

1.3.4 COMPUTATIONAL MODEL

Algorithm 1.3.1 is the basis for program `sis1` presented at the end of this section — a trace-driven computational model of a simple inventory system.

Program sis1

Program `sis1` computes five statistics: $\bar{d}$, $\bar{o}$, $\bar{l}^+$, $\bar{l}^-$ and the order frequency $\bar{u}$, which is

$$\bar{u} = \frac{\text{number of orders}}{n}.$$

Because the simulated system is flow balanced, $\bar{o} = \bar{d}$ and so it would be sufficient for program `sis1` to compute just one of these two statistics. The independent computation of both $\bar{o}$ and $\bar{d}$ is desirable, however, because it provides an important consistency check for a (flow balanced) simple inventory system.

Example 1.3.4 Program `sis1` reads input data corresponding to $n = 100$ time intervals from the file `sis1.dat`. With the inventory policy parameter values $(s, S) = (20, 80)$ the results (with *dd.dd* precision) are

$$\bar{o} = \bar{d} = 29.29 \qquad \bar{u} = 0.39 \qquad \bar{l}^+ = 42.40 \qquad \bar{l}^- = 0.25.$$

As with program `ssq1`, in Chapter 3 we will free program `sis1` from its reliance on external data by using randomly generated demand data instead.

1.3.5 OPERATING COST

Definition 1.3.5 In conjunction with the four statistics $\bar{o}$, $\bar{u}$, $\bar{l}^+$ and $\bar{l}^-$, a facility's cost of operation is determined by four constants:

- c_{item} — the (unit) cost of a new item;
- c_{setup} — the setup cost associated with placing an order;
- c_{hold} — the cost to hold one item for one time interval;
- c_{short} — the cost of being short one item for one time interval.

Case Study

Consider a hypothetical automobile dealership that uses a weekly periodic inventory review policy. The facility is the dealer's showroom, service area and surrounding storage lot and the items that flow into and out of the facility are new cars. The supplier is the manufacturer of the cars and the customers are people convinced by clever advertising that their lives will be improved significantly if they purchase a new car from this dealer.

Example 1.3.5 Suppose space in the facility is limited to a maximum of, say $S = 80$, cars. (This is a small dealership.) Every Monday morning the dealer's inventory of cars is reviewed and if the inventory level at that time falls below a threshold, say $s = 20$, then enough new cars are ordered from the supplier to restock the inventory to level S.*

- The (unit) cost to the dealer for each new car ordered is $c_{\text{item}} = \$8000$.
- The setup cost associated with deciding what cars to order (color, model, options, etc.) and arranging for additional bank financing (this is not a rich automobile dealer) is $c_{\text{setup}} = \$1000$, independent of the number ordered.
- The holding cost (interest charges primarily) to the dealer, per week, to have a car sit unsold in his facility is $c_{\text{hold}} = \$25$.
- The shortage cost to the dealer, per week, to not have a car in inventory is hard to determine because, in our model, we have assumed that *all* demand will ultimately be satisfied. Therefore, any customer who wants to buy a new car, even if none are available, will agree to wait until next Monday when new cars arrive. Thus the shortage cost to the dealer is primarily in goodwill. Our dealer realizes, however, that in this situation customers may buy from another dealer and so, when a shortage occurs, he sweetens his deals by agreeing to pay "shorted" customers \$100 cash *per day* when they come back on Monday to pick up their new car. This means that the cost of being short one car for one week is $c_{\text{short}} = \$700$.

---

* There will be some, perhaps quite significant, delivery lag but that is ignored, for now, in our model. In effect, we are assuming that this dealer is located adjacent to the supplier and that the supplier responds immediately to each order.

Definition 1.3.6 A simple inventory system's average operating costs *per time interval* are defined as follows:

- item cost: $c_{\text{item}} \cdot \bar{o}$;
- setup cost: $c_{\text{setup}} \cdot \bar{u}$;
- holding cost: $c_{\text{hold}} \cdot \bar{l}^{+}$;
- shortage cost: $c_{\text{short}} \cdot \bar{l}^{-}$.

The average total cost of operation per time interval is the sum of these four costs. This sum multiplied by the number of time intervals is the total cost of operation.

Example 1.3.6 From the statistics in Example 1.3.4 and the constants in Example 1.3.5, for our auto dealership the average costs are:

- the item cost is $\$8000 \cdot 29.29 = \$234,320$;
- the setup cost is $\$1000 \cdot 0.39 = \$390$;
- the holding cost is $\$25 \cdot 42.40 = \$1060$;
- the shortage cost is $\$700 \cdot 0.25 = \$175$.

Each of these costs is a per week average.

Cost Minimization

Although the inventory system statistic of primary interest is the average total cost per time interval, it is important to know the four components of this total cost. By varying the value of s (and possibly S) it seems reasonable to expect that an optimal (minimal average cost) periodic inventory review policy can be determined for which these components are properly balanced.

In a search for optimal (s, S) values, because $\bar{o} = \bar{d}$ and $\bar{d}$ depends only on the demand sequence, it is important to note that the item cost is *independent* of (s, S). Therefore, the only cost that can be controlled by adjusting the inventory policy parameters is the sum of the average setup, holding, and shortage costs. In Example 1.3.7, this sum is called the average *dependent cost*. For reference, in Example 1.3.6 the average dependent cost is $\$390 + \$1060 + \$175 = \$1625$ per week.

If S and the demand sequence is fixed, and if s is systematically increased, say from 0 to some large value less than S, then we expect to see the following.

- Generally, the average setup cost and holding cost will increase with increasing s.
- Generally, the average shortage cost will decrease with increasing s.
- Generally, the average total cost will have a 'U' shape indicating the presence of one (or more) optimal value(s) of s.

Example 1.3.7 is an illustration.

Example 1.3.7 With S fixed at 80, a modified version of program `sis1` was used to study how the total cost relates to the value of s. That is, the cost constants in Example 1.3.5 were used to compute the average setup, holding, shortage, and dependent cost for a range of s values from 1 to 40. As illustrated in Figure 1.3.9, the minimum average dependent cost is approximately \$1550 at $s = 22$.

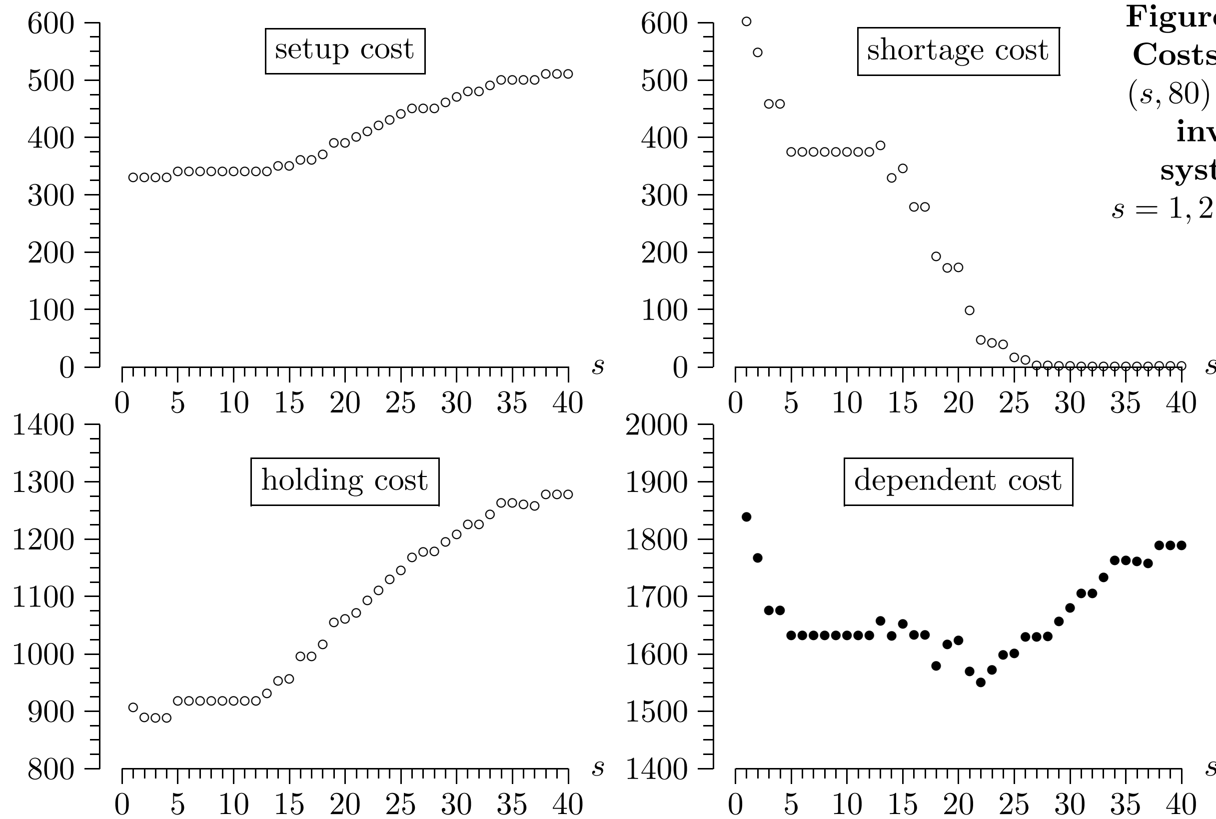


Figure 1.3.9. Costs for an $(s, 80)$ simple inventory system for $s = 1, 2, \ldots, 40$.

As in the case study concerning the ice cream shop, the "raw" simulation output data is presented. In this case, however, because the parameter s is inherently integer-valued (and so there is no "missing" output data at, say, $s = 22.5$) no interpolating or approximating curve is superimposed. [For a more general treatment of issues surrounding simulation optimization, see Andradóttir (1998).]

1.3.6 EXERCISES

Exercise 1.3.1 Verify that the results in Example 1.3.1 and the averages in Examples 1.3.2 and 1.3.3 are correct.

Exercise 1.3.2 (a) Using the cost constants in Example 1.3.5, modify program `sis1` to compute all four components of the total average cost per week. (b) These four costs may differ somewhat from the numbers in Example 1.3.6. Why? (c) By constructing a graph like that in Example 1.3.7, explain the trade-offs involved in concluding that $s = 22$ is the optimum value (when $S = 80$). (d) Comment on how well-defined this optimum is.

Exercise 1.3.3 Suppose that the inventory level $l(t)$ has a constant rate of change over the time interval $a \leq t \leq b$ and both $l(a)$ and $l(b)$ are integers. (*a*) Prove that

$$\int_a^b l(t)\,dt = \int_a^b \lfloor l(t) + 0.5 \rfloor\,dt = \frac{1}{2}(b-a)\big(l(a) + l(b)\big).$$

(*b*) What is the value of this integral if $l(t)$ is truncated rather than rounded (i.e., if the 0.5 is omitted in the second integral)?

Exercise 1.3.4 (*a*) Construct a table or figure similar to Figure 1.3.7 but for $S = 100$ and $S = 60$. (*b*) How does the minimum cost value of s seem to depend on S? (See Exercise 1.3.2.)

Exercise 1.3.5 Provided there is no delivery lag, prove that if $d_i \leq s$ for $i = 1, 2, \ldots, n$, then $\bar{l}^- = 0$.

Exercise 1.3.6 (*a*) Provided there is no delivery lag, prove that if $S - s < d_i \leq S$ for $i = 1, 2, \ldots, n$ then $\bar{l}^+ = S - \bar{d}/2$. (*b*) What is the value of $\bar{l}^-$ and $\bar{u}$ in this case?

Exercise 1.3.7 Use Definitions 1.3.3 and 1.3.4 to prove that the average inventory level equation

$$\bar{l} = \frac{1}{n}\int_0^n l(t)\,dt = \bar{l}^+ - \bar{l}^-$$

is correct. *Hint*: use the $(\cdot)^+$ and $(\cdot)^-$ functions defined for any integer (or real number) x as

$$x^+ = \frac{|x| + x}{2} \qquad \text{and} \qquad x^- = \frac{|x| - x}{2}$$

and recognize that $x = x^+ - x^-$.

Exercise 1.3.8 (*a*) Modify program `sis1` so that the demands are first read into a circular array, then read out of that array, as needed, during program execution. (*b*) By experimenting with different starting locations for reading the demands from the circular array, explore how sensitive the program's statistical output is to the order in which the demands occur.

Exercise 1.3.9[a] (*a*) Consider a variant of Exercise 1.3.8, where you use a conventional (non-circular) array and randomly *shuffle* the demands within this array before the demands are then read out of the array, as needed, during program execution. (*b*) Repeat for at least 10 different random shuffles and explore how sensitive the program's statistical output is to the order in which the demands occur.