IEEM 101: Inventory control

Outline of this series of lectures:

1. Definition of inventory
2. Examples of where inventory can improve things in a system
3. Deterministic Inventory Models
    3.1. Continuous review: The Economic Order Quantity (EOQ) model
    3.2. Periodic Review
4. Stochastic Inventory models: The Newspaper Vendor Model

---

1. Definition of inventory

In most systems, ranging from a household kitchen to a production plant for Dell computers, some amount of materials is stored for use at a later time. In this series of lectures, we shall examine why we need such stored materials, and how to determine how much is a right amount to store.

Inventory: A stock of materials kept for future sale or use.

We shall be concerned with why such stocks should be kept, and what the economic implications of keeping inventory are to the organization. Based on this analysis, we shall derive some basic ***inventory control policies***: these policies dictate *when to place an order* for more materials, and *how much materials should be ordered*.

2. Examples of Use of Inventory

2.1. Ajay's Orange Juice consumption

The famous scientist Linus Pauling had a theory that vitamin C makes the body resistant against the flu virus. To get extra vitamin C, I consume approximately 300 ml of orange juice, or OJ, every day. I can buy a 300 ml carton of OJ every day, or I can buy a 1.8 L large carton, and consume it over the next 6 days. In the latter case, I will have to hold an inventory of some amount of OJ in my refrigerator during the week. Also, although six 300ml cartons cost a little more than one 1.8L carton (though they both provide the same amount of OJ totally. So option 1 (300 ml carton daily) has two cons – *low volume surcharge*, and high *ordering cost* in the form of daily travel time to the shop. On the other hand, option 2 has a con – higher refrigerator bills (inventory holding costs).

Further, there are some risks associated with the options: under option 1, if I decide to drink some OJ on some night, I must drive all the way to the nearest 24-hour store, incurring a high *out-of-stock cost*. On the other hand, with option 2, if on some days I consume extra oranges, and decide not to drink OJ, then there is a danger that my carton of OJ may expire before it is consumed -- thus I may have to bear *over-stock costs*. These risks are due to *demand variability*.

The question for me is to somehow quantify the merits/demerits of each scheme, and come up with the best strategy of how much OJ to buy, and when. This decision is called my *policy* regarding my OJ inventory.

2.2. Hospital blood bank

Hospitals in every city maintain a reasonable supply of donated blood, for use in transfusions during operations or for treatment of injured humans. This inventory level must be managed with some care: the shelf-life of stored blood components varies (approx 7 weeks for red-blood cells, 5 days for platelets, and ~ 1year for plasma). Thus over-stocking will only result in tremendous wastage. However, the cost of under-stocking is very high, since a shortage may result in lost lives. The demand has high variability at individual hospitals. This type of problem is often tackled by building multi-level inventories. One example could be to store a few days worth of blood at each hospital in a city, while maintaining a larger bank at a common city-level blood bank. Thus the variability of demand at individual hospitals can be somewhat counter-acted by the shared store of blood at the city-bank. Such multi-level inventory models are fairly complex to analyse, so we shall skip the analysis in this course.

2.3. A computer manufacturing factory

Consider a computer manufacturer, e.g. Dell that assembles different components of computers to construct computers. We focus on a simple sub-assembly operation: each computer comes with a pair of speakers. The speaker assembly line has few operations, and typically can produce 800 speakers per hour. The main computer assembly line, however, only puts out 100 computers per hour. The situation is shown in Figure 1.

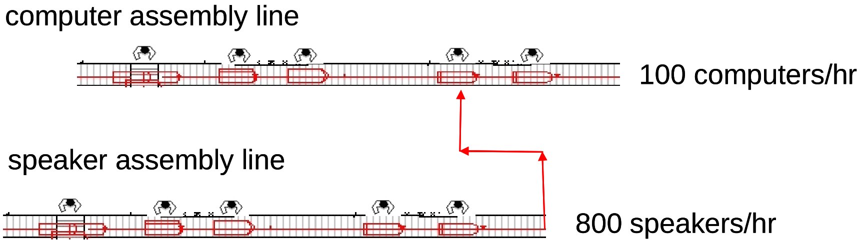


Figure 1. Assembly operations in a computer factory

The factory manager has two options: she can reduce the number of workers in the speaker assembly line – each remaining worker will then need to perform more assembly tasks, therefore the rate at which the speakers are produced will fall. Unfortunately, by increasing the number of activities for each worker, we also reduce their overall efficiency (since one worker, many tasks → low division of labour). On the other hand, if the speaker assembly line is working at optimal efficiency, it produces too many speakers (we only need 100 pairs, i.e. 200 per hour). One solution is to run the speaker line for an hour (making 400 pairs), and then stop this line for the next three hours, perhaps using the line to make some other components in this time. By repeating this cycle, we can maintain balance in the demand (400 pairs in 4 hours) and supply (400 pairs in 1+3 hours) of speakers. This solution obviously results in a non-zero inventory of speakers. The optimal policy for the plant manager is based on figuring out how long it takes to setup the speaker assembly line at the start of each cycle.

2.4. Retail shops

One of the most obvious examples of where inventory is essential is retail shops. Stores like Park-and-Shop maintain high level of inventory in the store, so that every customer can get the item they want (in fact, this level is so high in their stores that it is often inconvenient to walk around the store). Further, they also maintain a large inventory of items in their on-site warehouse (storage room). Thus retail chains often manage multiple-levels of inventory, ranging from on-shelf inventory, on-site inventory, and perhaps district/city level warehouses, and perhaps province level central warehouses.

2.5. Manufacturing Shops

Our final example shows in interesting case where maintaining inventory can result in utilization of a service facility (or a machine) at higher efficiency. Consider a part being manufactured on a machine. We require, on the average, 6 minute per part (though for individual parts, the time can be anywhere between 4 minutes and 8 minutes). Thus if we work on this machine for a full shift of 8 hours, we expect to produce approximately 80 parts (average of 10 parts per hour)

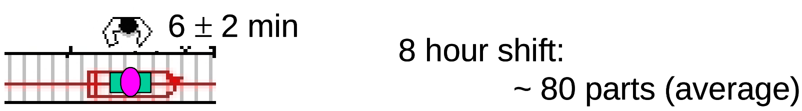


Figure 2. Variable processing time

Now let's assume that the part requires two operations, one after the other. For example, the first operation may be to perform some machining operations on a part, and the second operation may be to use a file to smooth out the rough edges left after to machining. Again, the time for *each* operation ranges between 6±2 minutes, with an average time of 6 minutes. If such a production line is working over a period of one 8-hour shift, we shall find that the average number of parts produced in this case is *much less than 80*! To understand why this happens, imagine that we get a sequence of parts requiring the times as follows: (time for operation 1, time for operation2): (7, 5), (6, 7), (5, 6), (6, 6) etc. Assume that when the first worker gets part 1 (takes 7 minutes), the second worker completed a previous part in, say, 6 minutes. Figure 4 shows what happens at each station as time goes on:

It is evident that each worker is idle intermittently at some stage or the other. Although the average task time for each worker is 6 minutes per part, due to the forced idle times, the overall efficiency of the workers, and of the system, is less than 100%. Thus, on an average, every 8 hour shift will produce fewer than 80 parts.

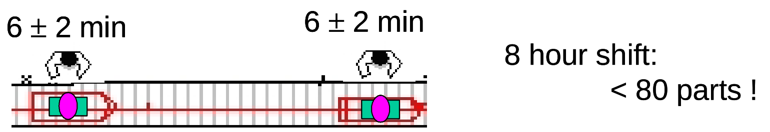


Figure 3. Two-worker assembly line, variable process times

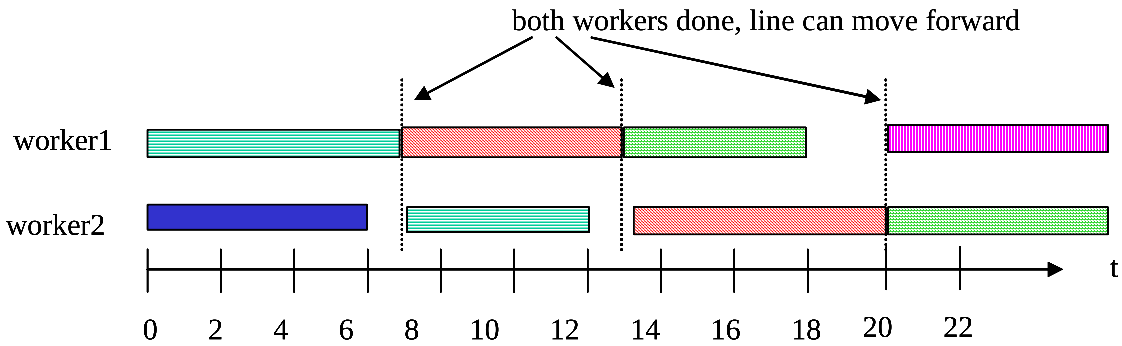


Figure 4. A chart showing time relation of different tasks, with task times for worker1 and worker2 as: (.., 6) (7, 5), (6, 7), (5, 6), (6, 6),…

Again, consider the same 2-worker line, but this time, we put a bin of input parts for each worker, as shown in Figure 5. The bin before worker 1 contains incoming parts. After worker-1 has completed the operation, he puts the semi-finished part into the bin in front of worker-2. Each worker, as soon as he completes his current job, can pick up a fresh part from his bin. Initially, both bins are non-empty. As you can imagine, if the bins are filled with enough parts, then, on an average, worker-1 is filling the bin of worker-2 at approximately 80 parts per 8-hour-shift. Likewise, worker-2 has no idle time, and therefore, on an average processes 80 parts per 8-hour shift. Here, the bins act as a "buffer", called a Work-In-Process inventory (*WIP inventory*), which can increase the efficiency of the system. This is another use of inventories, and again, the optimal policy should tell us how much WIP inventory would be sufficient to maintain a given level of throughput rate of part production.

These examples highlight the different and important uses of inventory. From an economical point of view, larger inventories are bad: several costs are associated with storage of materials – space must be purchased or rented, storage conditions must be controlled (e.g. temperature, humidity), some parts may get old or obsolete, capital is spent on inventory that could have been earning interest income, we may need to buy insurance if we are afraid the stocks will get damaged/stolen, etc. Clearly, we need to find an optimum level of inventory if we need to balance the pros with the cons.

We will now study a few simple models of inventory control. First, we look at deterministic models, where the demand is known; then we look at a stochastic model where we don't really know the demand level exactly.

3. Deterministic Inventory Models

Before starting, we introduce some terminology, which are commonly used in almost all inventory theory books, and will be used to develop out own models.

Ordering cost: Each time we place an order to buy some materials, we incur two types of costs – the ***setup cost*** (which can be seen as the order processing costs) and the ***production cost***, or per-unit item costs.

K ($ per order): ***setup cost***. This can include the cost of transporting the purchased materials to our storage location; it may even include cost of the time spent by our purchase officer to issue the purchase order. We will often assume that setup cost is a constant sum of money spent per order.

c ($ per item): ***production cost***. The production cost is the cost to purchase the material we want; if we buy Q units of material, and each unit costs c $, then the total production (or procurement) cost = cQ.

h ($ per unit item per unit time): ***holding cost***. This is the cost of storing each unit of material. It is typically expressed in $ per unit time. Estimating holding costs is a complex activity, since such costs include the cost of capital tied up, space rental, protection, insurance, and so on. We will assume in our models that the holding cost is constant (you should be aware that this assumption is not always valid in real life).

p ($ per unit item per unit time): ***shortage penalty***. This is easier to imagine from a retailer's inventory point of view. If a customer wants to buy an item that is not available at the retail shop. The customer may decide to go elsewhere to buy it – this leads to a loss of sale, and the retailer incurs a loss; or the customer may decide to wait for it, but we may still lose some goodwill; it is possible to associate a $-value to such loss of goodwill (since it results in future loss of sales.

R ($ per item): ***revenue***. In some models, we need to incorporate the money received when an item is sold (from the inventory), bringing a revenue of R dollars. We will not need this in our models.

Salvage value ($ per item). If an item stays in inventory for a period of time, its sale value may change. If the item deteriorates or its demand falls, then the salvage value is lower than the value at the beginning of the period. This is often true for perishable items (e.g. vegetables sold at the fresh food market).

α: ***discount rate***. This takes into account the time value of money. Capital that is spent on purchase of inventory could be earning interest. The time value of the tied-up capital depends largely on the discount rate. Consider this example. A computer manufacturer has an annual revenue from sales of computers = US$ 30 billion. Assume that the cost of materials is approximately US$10 billion. At any time, the company has 4 weeks of materials in inventory. Assuming a year to be approximately 50 weeks, this means that the value of inventory is approximately US$ 800 million. Consider that the company can improve their operations so that they only need to hold 2 weeks worth of inventory (i.e. $400 million). This results in a saving of US$ 400 million that can be invested in some secure investment, e.g. bonds, to earn 5% per annum. The reduction of the inventory levels leads to an extra annual income of 5% of 400 million = US$ 20 million. If the discount rate goes down by 1% to 4%, the corresponding loss of earnings would be US$ 4 million per annum!

3.1. Continuous review: The Economic Order Quantity (EOQ) model

The simplest situation faced by a retailer or manufacturer is when there is a constant rate of usage of some materials, and their stocks must be replenished from time to time. We shall first assume that the rate at which the material is used up, or purchased, is constant – namely there is a constant, uniform demand = a items per unit time. A setup cost, K, is charged each time we place an order. The inventory holding cost is h dollars per unit time per unit of material.

Uniform demand, No shortages

Our objective is to find the optimum quantity that we must order each time we place an order. The main constraint is that we must not allow for an out-of-stock situation – in other words, stocks must be replenished as soon as current inventory reaches zero (or earlier).

First assume that we receive supplies as soon as we place an order (i.e. zero lead time). Say we order an amount = Q. This is our initial inventory, of which a units gets used up in every unit of time. That is, after t units of time, remaining inventory = $Q - at$. Thus we need to reorder when stocks become zero, i.e. $Q - at = 0$, which is at $t = Q/a$. Obviously, if Q is the optimal order quantity, then we must just repeat this cycle of ordering and use, as shown in Figure 5.

Let's call the duration between $t=0$ and $t=Q/a$ as one *cycle* (since under optimal conditions, this pattern repeats indefinitely). The total expenses incurred in each cycle are made up of three components:

The setup cost = K;

The materials cost = cQ (we bought Q units at c dollars each);

The holding costs = h X (average inventory level during the cycle) X (length of the cycle)

$= h \times (Q - 0)/2 \times (Q/a) = hQ^2/2$

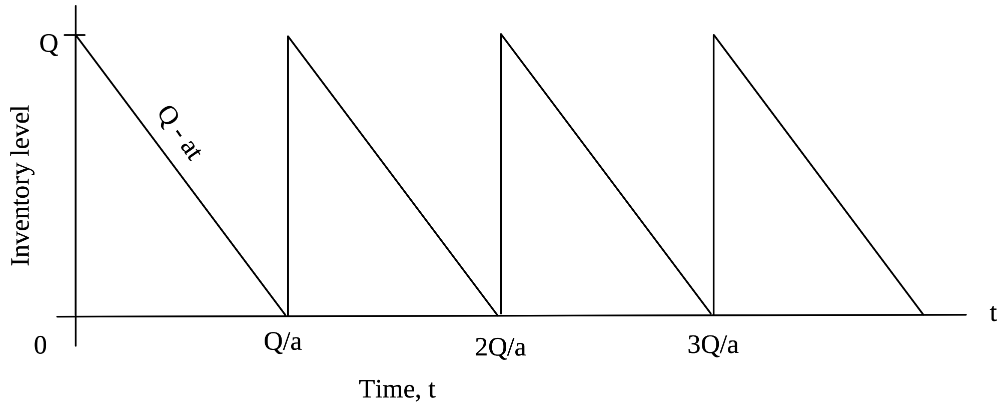


Figure 5. Uniform demand, no shortages allowed

Thus total cost per cycle = $K + cQ + \dfrac{hQ^2}{2a}$.

Cost per unit time = total cost/cycle length = $T(Q) = \dfrac{K + cQ + \dfrac{hQ^2}{2a}}{Q/a} = \dfrac{aK}{Q} + ac + \dfrac{hQ}{2}$

The best order quantity, Q, is that quantity which minimizes the cost per unit time, and to determine this, noting that T(Q) is a smooth function, we must have:

$dT/dQ = 0 = -\dfrac{aK}{Q^2} + \dfrac{h}{2}$, which gives the optimum order quantity = $Q^* = \sqrt{\dfrac{2aK}{h}}$.

Q* is often called the ***Economic Order Quantity***.

Now we can relax the assumption that order lag time = 0. From figure 5, it is clear that we should place an order so as to receive Q^* units at time $t^* = \dfrac{Q^*}{a} = \sqrt{\dfrac{2K}{ah}}$. If the delay between placing an order and receiving the shipment = d, then we should place the order at $t = t^* - d$.

Uniform demand, Shortages allowed

Now let's consider the case where some shortages are allowed. This is especially true for retail outlets, where some items may be out of stock, but the retailer may convince the customer to buy-now, collect-later. Obviously, some percentage of customers will reject this suggestion, and prefer to buy from another shop; on the other hand, some customers may accept the delay, especially if the retailer can give a bonus – e.g. free home delivery. This may happen if you buy, for example, a refrigerator from a large appliance store such as Fortress. We denote the *shortage cost* = *p* dollars per item per unit time.

Again, there is a trade-off: some percentage of sales is lost, and perhaps some additional costs are incurred due to the out-of-stock condition. On the other hand, the average inventory level that is held at the store is lower, and therefore the average holding costs are lower. The question is: what is the best policy in this case?

Assume that we receive Q units in our warehouse. At this time, we must be holding zero inventory, and in fact, must have been out-of-stock for some time – because if we have positive inventory, receiving new goods will lead to higher holding costs. In the period that we were out of stock, we have sold (but not delivered) several units. We immediately deliver these items, and so out of Q units, we are left with an inventory level of, say, S. this inventory gets used at the rate of a items per unit time, so after t = S/a, we are left with zero inventory. From this point of time until t = Q/a, we shall accumulate orders without delivery. Finally, at t = Q/a, we receive the next lot of Q units, and repeat this cycle.

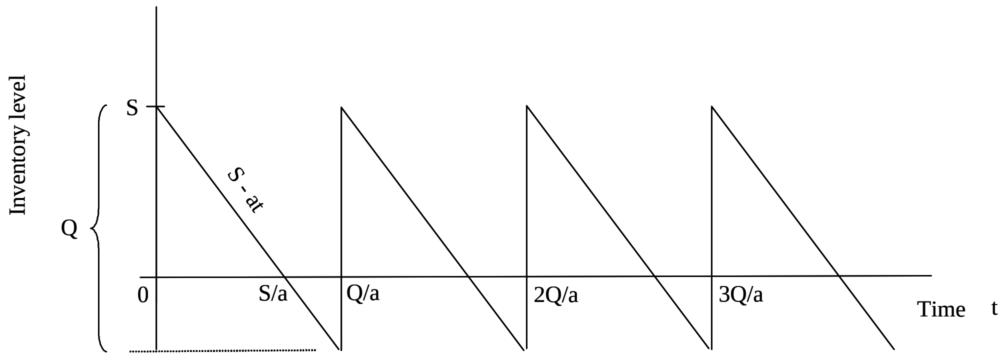


Figure 6. Uniform demand, shortages allowed

The objective is to find the optimum order quantity, Q, and the maximum level of shortage (= Q –S) allowed under the optimum policy.

Once again, as seen in figure 6, our inventory level exhibits a cyclical behaviour; in the figure, the shortage levels are indicated by a negative inventory as a convenient notation.

Total cost per period = setup cost + material cost + holding cost + shortage costs

From figure 6, it is easy to see that we only hold inventory in the first S/a units of time per cycle, and we must pay shortage costs during the last (Q/a – S/a) = (Q – S)/a units of time of each cycle.

Total cost per unit time is easily computed as before:

Total cost per period / length of period = $T = \dfrac{K + cQ + \dfrac{hS^2}{2a} + \dfrac{p(Q-S)^2}{2a}}{Q/a}$

Simplifying, we get: $T(S, Q) = \dfrac{aK}{Q} + ac + \dfrac{hS^2}{2Q} + \dfrac{p(Q-S)^2}{2Q}$

T is a function of S and Q, and it is well behaved. From basic calculus, its minimum point must satisfy the conditions: $\dfrac{\partial T}{\partial S} = 0, \;\; and \;\; \dfrac{\partial T}{\partial Q} = 0$.

$\dfrac{\partial T}{\partial S} = \dfrac{hS}{Q} - \dfrac{p(Q-S)}{Q} = 0$,

which gives, since $Q \neq 0$, $\quad (Q - S) = hS/p \quad$ [1]

that we can re-arrange as: $\quad Q = S(h+p)/p \quad$ [2]

$\dfrac{\partial T}{\partial Q} = -\dfrac{aK}{Q^2} - \dfrac{hS^2}{2Q^2} + \dfrac{p(Q-S)}{Q} - \dfrac{p(Q-S)^2}{2Q^2} = 0$, which gives (multiply both sides by $2Q^2$, and simplify): $2aK + hS^2 + 2p(Q-S)^2 = 2p(Q-S)Q$

In this, we substitute the expressions for (Q – S) and Q from [1] and [2] above, giving:

$2aK + hS^2 + \dfrac{2h^2S^2}{p} = 2h(h+p)S^2$, or

$2apK + phS^2 + 2h^2S^2 = 2h(h+p)S^2$, from which:

$S^* = \sqrt{\dfrac{2aK}{h}} \sqrt{\dfrac{p}{p+h}}$; substituting this into [2] above yields the optimum order quantity as: $Q^* = \sqrt{\dfrac{2aK}{h}} \sqrt{\dfrac{p+h}{p}}$. These values define the optimum policy for the case when shortages are allowed, and demand is constant.

Uniform demand, no shortages, volume discounts

Another variation of the basic EOQ model is when the unit price of the materials is not constant, but changes with the size of the order quantity. This is typically the case when a wholesaler offers volume discounts. The typical form of this type of discounting is as follows:

Order quantity	cost per item
$Q < A_1$	c_1
$A_1 \le Q < A_2$	c_2
$A_2 \le Q < A_3$	c_3
…	

Since we are concerned with volume discounts, $c_1 > c_2 > c_3 > \ldots$. To analyse this situation, consider the total cost per unit time for a particular material cost, say c_i. we can write: $T_i = \frac{aK}{Q} + ac_i + \frac{hQ}{2}$. The first and third terms on the RHS do not change – for any different value of per-unit cost, the difference between two functions, T_i and T_j, at any given value of Q, is: $a(c_i - c_j)$. Since this is constant, a change in unit cost merely shifts the Total cost curve up or down according as unit cost increases or decreases. Thus, for the situation of volume discounts, we can graphically represent the cost curves as shown in Figure 7.

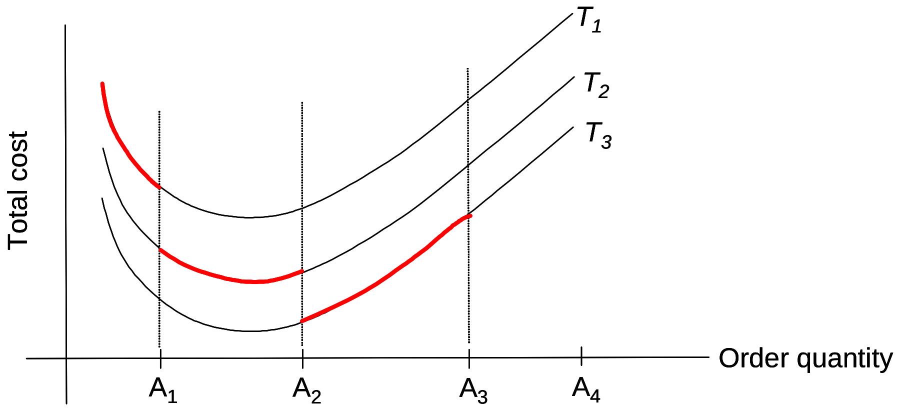


Figure 7. Constant demand, no shortage, volume discounts

The figure shows the total cost curves corresponding to the different unit costs. The actual cost curve in this case is a piecewise continuous curve, and is shown in red color in the figure. To find the optimum value of order quantity, we only need to find the minimum point on this piecewise curve. To do so, we find the minimum of each segment, and the global minimum is the least value among these.

3.2. Periodic Review

We now look at the problem of optimal inventory policy when the demand rate is not uniform. To simplify the analysis, we shall break up the time axis into discrete periods, and assume that the demand for each period is known. This type of situation is not uncommon, since many companies register sales data only in the last week of each month, and therefore their demand data is really capturing the monthly demand rate – in this case, the periods are 1-month long.

Further, we shall assume that the demand for each period is different (that is, the demand for each month in the planning horizon is known). We shall depict the material requirements, or demand, for each period, as: $r_1, r_2, \ldots, r_n$. No out-of-stock situation is allowed in this model, and the other costs are the same: setup cost = *K*, production cost = *c* $ per unit. Finally, the holding cost will now be modelled as *h* dollars per unit *per period* (note the difference from the earlier models, where *h* was the holding cost per item per *unit time*). The reason for this choice is merely simplicity of notation.

The first thing to note is that since the demand (consumption) rate is not the same at different times, therefore the pattern for our purchase of goods will not be a repeated cycle as in our previous models. In the current case, the length of the periods is given, not derived from the analysis. Therefore, the purchasing pattern is dictated by the variable demand rate.

To simplify our analysis, we shall again impose a condition that we order at the beginning of a period. In other words, we seek solutions that specify how much quantity of materials will be purchased at the beginning of each period in the planning horizon. It is sometimes better not purchase anything at all in the beginning of some periods. The following example shows why.

Example

Planning horizon:	2 periods
Demand:	r1 = 100, r2 = 10
Unit cost:	c = 1
Holding cost:	h = 1
Setup cost:	K = 50

There are two possible solutions: order the entire amount at the start of period 1, or order just enough for period 1 at the start, and then order enough for the 2nd period at the start of period 2. These are shown in Figure 8 (a) and (b) respectively.

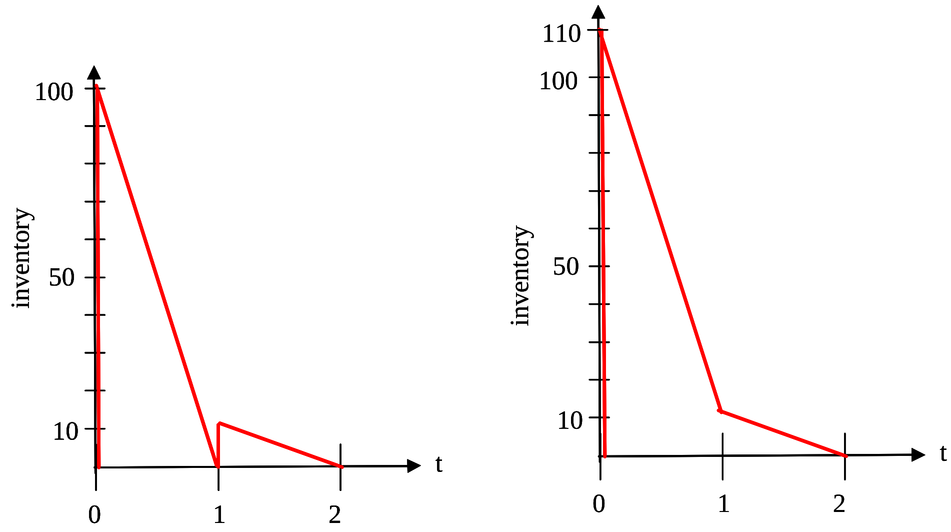


Figure 8 (a) Option 1 (b) Option 2

The total cost for option 1

= $\Sigma_{i=1,2}$ (setup-cost$_i$ + purchasing-cost$_i$ + holding-cost$_i$)

= (50 + 1x100 + 1x100/2) + (50 + 1x10 + 1x10/2) = 265

The total cost for option 2:

= $\Sigma_{i=1,2}$ (setup-cost$_i$ + purchasing-cost$_i$ + holding-cost$_i$)

= (50 + 1x(100+10) + (1x100/2 + 1x10)) + (0 + 0 + 1x10/2) = 225

Clearly, it is better to order once, at the start of period 1, for the entire planning horizon.

Before we develop a simple technique for computing the optimal policy in such problems, let us look at a useful property of the optimal solution:

Property 1. For fixed setup cost, constant production cost per item, and constant holding cost per item per unit time, an optimum policy will require production only when the inventory level reaches zero at the start of a period.

Proof: Stated another way, the property says that if we decide to produce (or procure) some items at the start of period *j*, then the optimum inventory level at the end of period *(j-1)* must be zero. Intuitively, the reason for this as follows: we are carrying inventory because we want to avoid paying the setup cost. Suppose that we decide to pay the setup cost at a point of time, with an order quantity of *Q*; assume that at that moment, we are holding x units of inventory. So we have been paying the holding costs on these *x* units of inventory in the previous period. If, instead, we had ordered up to *x* units less in our previous order, then (a) would not have had any shortage (or else how could we be holding x units now?), and (b) we would have saved paying the holding costs on these units. At the start of this period, we can just order *(Q+x)* units, without incurring any extra setup costs.

This situation is shown graphically in Figure 9. Here, the blue colour graph shows some non-optimal policy. There is a finite amount produced at the start of period *i*, and subsequently at the start of period *j*. This policy cannot be optimal since we could have produced a smaller amount at the start of period *i*, such that we reach the start of period *j* with exactly zero inventory, as shown by the red colour graph. Note that from period *i* to *(j-1)*, the red graph has a lower inventory level at any given time, and so the holding costs are lower in this case.

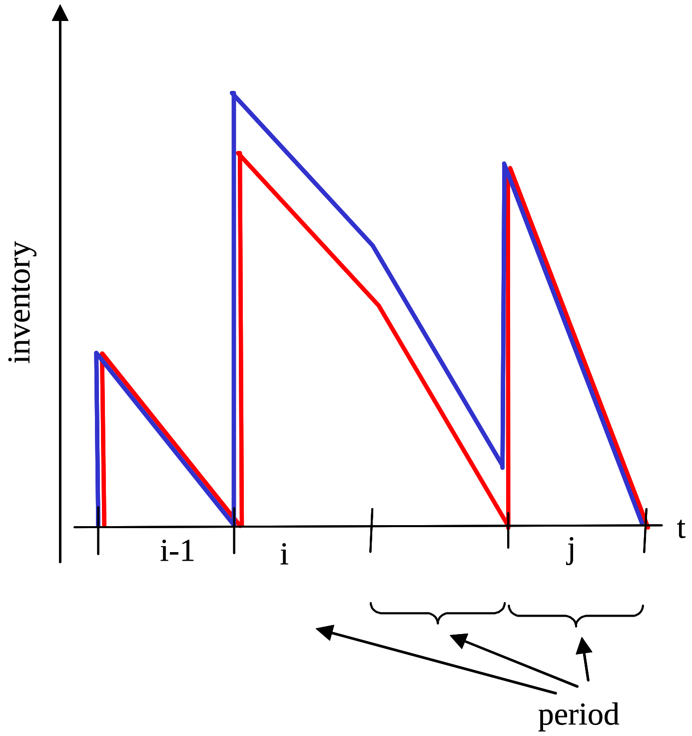


Figure 9. Optimally, inventory level is zero at each ordering point

This property is useful to us. We know that we are going to produce only at the start of some periods. Therefore, the amount we need to produce at the start of period *j* is either r_j, or $(r_j + r_{j+1})$, or …, or $(r_j + \ldots + r_n)$.

In other words, if we order at the start of period *i*, and subsequently order at the start of period *j*, then the optimal order quantity at the start of period *i* is $(r_i + \ldots + r_{j-1})$. So if we know the ordering points, we know the optimal policy. Unfortunately, there are n possible ordering points (start of each of the n periods); we must order at the start of period 1; at each of the remaining n-1 start points of the periods, we may or may not order. Therefore there are 2^{n-1} possible policies, one of which must be optimal. This is a large number of policies to evaluate – but fortunately, we can find the optimal solution with much less effort.

Method to solve periodic review models

As consequences of Property 1, the quantity produced (ordered) at start of period i $\in$ { 0, r_i, $r_i + r_{i+1}$, …, $r_i + \ldots + r_n$}

Suppose that the optimal amount produced at the start of period 1 = Q_1* covers k periods

Then we only need to solve for the optimum solution for a smaller problem:

Starting from period $k+1$, with demands = r_{k+1}, r_{k+2}, … r_n

Let:

C_i = Cost of *optimum policy* for periods *i,,n* when the inventory level at the start of period *i = 0*.

Then:

$$C_1 \equiv \text{Min} \begin{cases} \text{cost of making } r_1 \text{ at } t{=}0 + C_2 \\ \text{cost of making } (r_1+r_2) \text{ at } t{=}0 + C_3 \\ \ldots \\ \text{cost of making } (r_1+r_2+r_k) \text{ at } t{=}0 + C_{k+1} \\ \ldots \\ \text{cost of making } (r_1+r_2+\ldots+r_n) \text{ at } t{=}0 + C_n \end{cases}$$

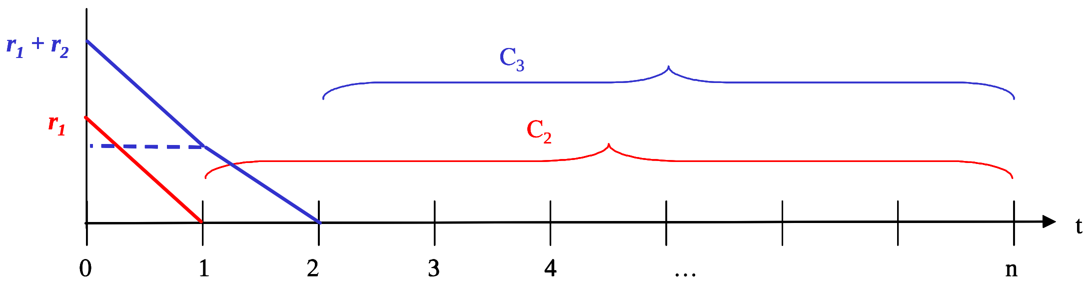


Figure 10. Graphical illustration of first two options for finding C_1

Let us try to compute the cost of the different options in finding the best C_1.

Cost of making r_1 at t=0 + C_2 = K + c r_1 + h r_1 / 2 + C_2

Cost of making (r_1+r_2) at t=0 + C_3 = K + c(r_1+r_2) + (h r_1 / 2 + hr_2 +hr_2 / 2) + C_3

K + c(r_1+r_2) + h(r_1 + 3r_2)/2 + C_3

and in general,

Cost of making (r_1+r_2+ r_k) at t=0 + C_{k+1} = K + c(r_1+…+r_k) + h/2 $\Sigma_{1,,k}$ (2i -1)r_i + C_{k+1}

Therefore, to compute C_1, we must first know the values of C_2, …, C_n

Likewise, to compute C_2, we must know C_3, …, $C_{\underline{n}}$, and so on.

From this observation, we can see that a backward planning approach can be used to compute the best policy:

First find C_n, then use it to find C_{n-1}, and so on, until we can find C_1.

Step 1.

C_n: only option is to produce r_n at the start of period n:

Cost = C_n = K + c r_n + h/2 r_n

Step 2.

To compute C_{n-1}, we must compare between two choices: either we produce r_{n-1} now and the remaining demand later, or we produce (r_{n-1} + r_n) now. The corresponding costs are easily computed to give:

$$C_{n-1} = \min \begin{cases} \text{produce } r_{n-1} + C_n \longrightarrow K + c\, r_{n-1} + h/2\, r_{n-1} + C_n \\ \text{produce } r_{n-1}+r_n \longrightarrow K + c\,(r_{n-1}+r_n) + h/2\,(r_{n-1} + 3r_n) \end{cases}$$

Note that both the possibilities can now be evaluated since all terms, including *C_n*, are now known.

By continuing in this fashion, we can compute C_{n-2}, and then C_{n-3}, … until we find C_1.

The above models provide us with some insight about how inventory policies are determined when the demand is known in advance, over the entire planning horizon. Very often in real life, the demand is not known with certainty. In such cases, we must still make decisions about our inventory policy and ordering schedules. Since demand is not known with certainty, we cannot be certain if our plan is optimal. Nevertheless, we don't want to make extremely poor decisions – in other words, we want to make decisions that maximize *the likelihood* of minimizing the costs. It turns out that such models are much more complicated, and therefore not suitable to be taught in an introductory course. However, there is one special example that can be studied with an understanding of basic statistics, and it is quite interesting to study. We look at this model in the next section.

4. Stochastic Inventory models: The Newspaper Vendor Model

4.1. Some Statistical Background

Before we begin our study of the newsvendor problem, here is a brief review of simple statistical results. We begin with a simple example.

Example. A fruit seller makes a profit of $4 for each undamaged mango that he sells. However, some mangoes get slightly damaged due to handling errors, and he must discount them, making only a profit of $1 on them. Suppose that he sells 100 mangoes in a day – how much profit will he expect to make?

Clearly, the answer depends on how many mangoes are damaged. Assume that on an average, around 20% of mangoes are damaged. In this case, his profit can be computed as: 0.2 x 100 x ($1) + 0.8 x 100 ($4) = 20 + 320 = $340.

In other words, his expected profit per mango = $3.4.

We can easily generalize this result as follows:

Definition:

If the probabilities of obtaining amounts $a_1, a_2, \ldots, a_k$ are $p_1, p_2, \ldots, p_k$ respectively, then the expected value (or the ***expectation***) is given by:

$$a_1 p_1 + a_2 p_2 + \ldots + a_k p_k = \Sigma_{i = 1,,k}\, a_i p_i$$

Example. A florist sells wedding bouquets that he prepares each morning. Each bouquet is sold at $50, and the cost of making it is $35. Any bouquet that is not sold by the end of the day is a total loss. Based on his past experience, the probability of number of bouquets he can sell is shown in the table below.

number of bouquets	3	4	5	6	7	8	9
probability	0.05	0.12	0.20	0.24	0.17	0.14	0.08

Given this data, how many bouquets should he make each morning so as to *maximize the expected profit*?

Suppose that he makes 3 bouquets. That the demand is 3 or more is certain (probability = 1) – so he will certainly sell all three, with a profit of 3x50 – 3x35 = $45.

Suppose he makes 4 bouquets. The likelihood that only 3 bouquets will be sold is 0.05, while the likelihood that the demand is 4 or more is 0.95 (why?). Hence, his expected profit = 0.05 x 3 x 50 + 0.95 x 4 x 50 – 4 x 35 = $57.50

Suppose he makes 5 bouquets. The chance that he only sells 3 is exactly 0.05; the chance that he sells exactly 4 is 0.12. And the chance that the demand will be for 5 or more, in which case he can sell all 5, is (1 – 0.05 – 0.12) = 0.83. Therefore his expected profit is:

0.05 x 3 x 50 + 0.12 x 4 x 50 + 0.83 x 5 x 50 – 5 x 35 = $64.

Continuing in this fashion, we can compute his expected profit for any number of bouquets made, as shown in the table below:

number of bouquets	3	4	5	6	7	8	9
probability	0.05	0.12	0.20	0.24	0.17	0.14	0.08
Expected profit	45	57.5	64	60.5	45	21	-10

Clearly, he should make 5 bouquets every day.

Let's say that we are interested in a measure of something (e.g. quantity demanded in a given period), and that this measure can take different values, with different probabilities. We can use a variable to represent the thing we are measuring. Since the variable can have a different value each time we measure it, we call it a ***random variable***. Let's say that for each value that the random variable can take, we associate a probability. Such an association provides us with a "mapping" – for each value, *x*, of the random variable, there is a value, called the probability of *x*, or *p(x)*. Note that for any given value of the variable, there is exactly one probability – so this mapping is really a function. It is conventional to call this mapping, p(x), a ***probability distribution function***.

In the above example, the demand is discrete – our random variable (number of bouquets) can only have integer values. However, in many cases, the random variable can take any real value (for example, the height of people in a city) within a given range. In such cases, i.e. when we have a ***continuous random variable***, we extend the idea of probability distributions to a ***probability density function***. For a continuous random variable, the probability that it takes a specific value is zero (think of selecting a random real number between 0 and 1: there are infinite numbers in this range, so the likelihood that you come up with any one of them is $1/\infty = 0$). However, the likelihood that the random variable lies within a certain range, no matter how small the range, is more meaningful. A common probability distribution is the normal distribution, which is the familiar bell shaped graph as the one in Figure 11 below.

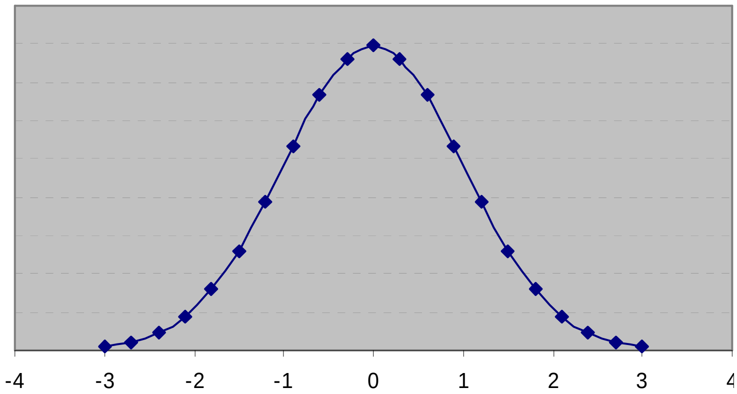


Figure 11. A normal distribution curve

Formally, the probability density function, f(x), is defined is such a way that the probability that the random variable lies between some range, $a \leq x \leq b$, is given by:

$P(a\leq x \leq b) = \int_a^b f(x)\,dx$.

Using this convention, the probability that the random variable has a value $\leq$ k for some k is $\int_{-\infty}^{k} f(x)\,dx$. And so we can define a cumulative distribution function, F(x) as the probability that the random variable has a value $\leq$ x as: $F(x) = \int_{-\infty}^{x} f(x)\,dx$. This notation is convenient, since $P(a\leq x \leq b) = F(b) - F(a)$.

Just as we can compute the expected value of a discrete random variable using the formula $E(x) = \Sigma_{i\ =\ 1,,k}\ x_i p_i$, similarly, we can get the expected value of a continuous random variable as: *expected value of x* = $E(x) = \int_{-\infty}^{\infty} f(x)\,dx$.

It is useful to know a little bit more about the normal distribution. The normal probability density function is given by:

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\ \sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \qquad -\infty \leq x \leq \infty$$

Not a pleasant looking function, but we shall need to deal with it directly. The standard normal distribution curve (similar to the one in Figure 11), has a mean, μ = 0, and a standard deviation, σ = 1. To find the cumulative distribution function, F(x), for it, we would need to integrate this function from $-\infty$ to x; this is done using numerical methods, and the values of F(x) corresponding to the standard normal distribution are available as a standard table in most textbooks on statistics.

The table are useful for any data that is normally distributed (no matter what its mean and standard deviation are) because of the following property:

Property:

If a random variable, x, is distributed normally with mean = μ and standard deviation = σ, then the corresponding standard random variable is given by $z = (x - \mu)/\sigma$, and z is normally distributed, with a $\mu = 0$ and $\sigma = 1$.

The newsvendor model

We shall introduce the problem by means of an example. Mrs. Kandell runs a Christmas tree shop, and must place orders for Christmas trees in September. The trees are delivered and sold in December, and the exact number of trees she will sell is unknown until then. Each tree costs $25, and is sold for $55 if she can sell it before Dec 25. However, after Dec 25, if the tree remains unsold, she can only get $15 for it. By the time she has better information about her actual demand, it is too late to order more trees. The table below gives the probability associated with the number of trees she can expect to sell (such data is generated using past information for her shop and other shops in the neighbourhood, for example).

If she orders too many, then the cost of overstock = co = $25 - $15 = $10 per tree.

If she orders too few, then the cost of understock (lost sales) = cu = $55 - $25 = $30 per tree.

Sales	22	24	26	28	30	32	34	36
Probability	.05	.10	.15	.20	.20	.15	.10	.05

How many trees should she order ?

Let us first state the assumptions of the model we shall construct for this problem:

(1) We need to plan for a single period

(2) The demand is unknown

(3) p(y) = probability that the demand will be equal to y units is known

(4) For now, we shall assume that there is no setup cost

Let:

D total demand before Christmas

$F(x)$ the demand distribution (known)

So:

$D > Q$ ➔ stockout, at a cost of: $c_u\,(D - Q)^+ = c_u \max\{D - Q, 0\}$

and

$D < Q$ ➔ overstock, at a cost of: $c_o\,(Q - D)^+ = c_o \max\{Q - D, 0\}$

The total cost = G(Q) = $c_u\,(D - Q)^+ + c_o\,(Q - D)^+$.

Note that at most one of the two terms of G(Q) is positive. Further, since the demand, D, is a random variable, therefore G(Q) is also a random variable, and we can compute its expected value.

The Expected cost, E(G(Q)) = E($c_u\,(D - Q)^+ + c_o\,(Q - D)^+$) = c_u E$(D - Q)^+$ + c_o E$(Q - D)^+$

$$= \sum_{x=0}^{\infty}[c_u(x-Q)^+ + c_o(Q-x)^+]P(x) \;=\; \sum_{x=Q}^{\infty}[c_u(x-Q)^+]P(x) + \sum_{x=0}^{Q}[c_o(Q-x)^+]P(x)$$

where P(x) is the probability distribution of the demand D, in other words, P(x) is the probability that D=x.

Mrs. Kandell's ***objective*** is to *minimize her expected costs*.

Of course, it is possible to analyze this situation as a discrete model, just as we did for the example above. However, this time we shall take a different approach, and solve the problem assuming that the demand D is a *continuous* random variable, and that the probability of demand can be expressed in the form of a probability density function, f(x).

Therefore, we approximate the expression of E(G(Q)) as:

$$g(Q) \;=\; E(G(Q)) \;=\; \int_{x=0}^{Q} c_0\,(Q-x)\,P(x)\,dx + \int_{x=Q}^{\infty} c_u\,(x-Q)\,P(x)\,dx$$

The model that optimizes the expected cost is found by the usual way, namely solving for dg(Q)/dQ = 0. to do this, we need to use a result of calculus that allows us to find the derivative of functions that are the result of an integration. For our discussion here, I will skip the details (extra notes will be provided for those who are interested).

The result of solving for dg(Q)/dQ = 0 is that the amount of order that minimizes our expected costs is achieved when the cumulative distribution function, F(Q) = $c_u/(c_u + c_o)$. Recall that $F(Q) = \int_0^Q P(x)\,dx$. Most books on inventory theory call $c_u/(c_u + c_o)$ as the critical ratio, and usually depict it as the Greek letter $\beta = c_u/(c_u + c_o)$.

Let's apply this result to the problem faced by Mrs Kandell. The table below shows the probability of the different demand quantities; from this, the cumulative distribution is easily computed, as shown.

D	22	24	26	28	30	32	34	36
Probability	0.05	0.1	0.15	0.2	0.2	0.15	0.1	0.05
***F*(*D*)**	0.05	0.15	0.3	0.5	0.7	0.85	0.95	1

In her case, β = 30/(10 + 30) = 30/40 = 0.75.

From the above table, we can see that the cumulative distribution function reaches 0.75 somewhere between a demand of 30 and 32 trees, i.e. roughly 31 trees. Thus the optimum number of trees that Mrs Kandell should order is 31.

Notice that if we computed the expected value of the demand, E(D) = 22x 0.05 + 24 x 0.1 + … + 36 x 0.05 = 29. In other words, the optimum solution is suggesting her to purchase slightly more number of trees than the expected demand. Why is this?

The answer has to do with the value of the critical ratio. Let's examine the extreme cases to understand what effect β has on the inventory decision. Consider that b ≈ 0, which occurs when the under-stocking cost is negligible compared to the over-stock cost. In this situation, the model suggests that we buy the least possible number of units (because the overstock cost is high, so we don't want to take even the slightest chance that we will purchase more than the actual demand). On the other hand, assume that b ≈ 1, in which case the overstock cost is negligible; now, F(D) reaches 1 only when we order the maximum possible demand, i.e. 36 units. Since the overstock cost is negligible, it makes sense for us to order the maximum possible trees, since even if the real demand turns out to be lower, we will not incur a significant cost).

In conclusion, the value of the critical ratio can give us an indication of (a) how much we will deviate from the expected demand, and (b) which way will we deviate from the expected demand, i.e. shall we order more, or less than the expected demand.

In our example above, the demand distribution was assumed to be continuous, although it had been stated initially as discrete. On the other hand, in real life examples, it is often very difficult to come up with accurate probabilities of different levels of demand. This is even more difficult when the random variable can really take arbitrarily large number of values. In such cases, data of the past demand can be used as a sample, and its mean/standard deviation can be used as a fair representation of the true mean/standard deviation of the variable. Further, it is possible to approximate the distribution with one of the standard probability distributions (e.g. normal, poisson, uniform, etc.) If these choices (of the mean, standard deviation and the distribution type) are representative of the data, then one can conveniently look up the values of F(x) from the standard tables available for such distributions. Let's look at an example of this type.

Example

A clothing designer will introduce a new fashion dress, selling for $2000/piece. The production costs are $1000/piece. At the end of the season, all remaining stock will be sold at a 70% discount.

Assume that the demand follows a normal distribution, with a mean 500 and standard deviation of 100.

(a) What is the recommended order quantity?

(b) What is the probability that at least some customers will ask to purchase the product after it is sold out, assuming quantity obtained in part (a) is ordered?

Solution

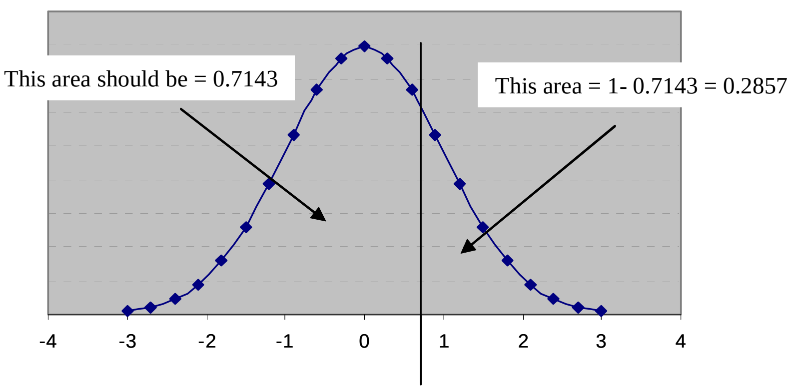
(a) We need to use the normal distribution function, and the related data (which is available in standard tables) to solve this problem.

If we cannot sell the dress in this season, we will sell it at 70% discount, for $ 600, therefore the overstock cost, co = $1000 - $ 600 = $ 400.

The under-stock cost = c_u = $2000 - $1000 = $1000.

β = 1000/(1000+400) = 0.7143.

We now need to look at the table providing values for F(x) for the standard normal distribution curve, corresponding to the situation shown below.



From the standard tables, corresponding to the point on the standard normal distribution corresponding to area = 0.7143 is approximately at μ+0.565σ. For our normal distribution, this corresponds to the point would therefore be at 500 + 100 x 0.565 = 556.5 , or nearly 557 units.

Therefore we should order 557 units.

(b) For this part, we need to find the area under the curve for all points beyond $\mu+0.565\sigma$, i.e. P(D > 557), which is 1 – F((557 – 500)/100) = 1 – F(0.57) = 0.2843.

Concluding remarks

The study of inventory and the control of inventory levels is a crucial activity for all enterprises. Although real-life situations are made complex due to randomness in the demand, there are many cases where the demand pattern approximately follows one of the models that we have studied. Most modern inventory theory uses tools from statistical analysis: we only studied the simplest stochastic model here. However, even with this simple example, we can appreciate the fact that some amount of analysis of the problem can lead to the discovery of interesting properties (e.g. the critical ratio) that can assist in making better decisions.

Main reference:

Operations Research, Hillier and Leiberman

Lecture notes, Prof Liming Liu, Dept of IEEM, HKUST (for the Newsvendor model)