Chapter 7
The Newsvendor Problem

Evan L. Porteus
Stanford University

The newsvendor problem has numerous applications for decision making in manufacturing and service industries as well as decision making by individuals. It occurs whenever the amount needed of a given resource is random, a decision must be made regarding the amount of the resource to have available prior to finding out how much is needed, and the economic consequences of having "too much" and "too little" are known.

Introduction

Tyler has just been put in charge of deciding how many programs to order for this week's home football game between his team, the Rainbows, and the visiting team, the Warriors. Each program ordered costs \$1.25 and sells for \$5.00 each. He found out from Nar Lee, the experienced department administrator, that 8,500 programs were ordered for the last home game and only 7,756 were sold. He thinks to himself that his problem is solved. He'll simply order 7,756 programs. But he decides to discuss the issue with his friend, Shenn, who is a little bit nerdy, but tends to tease people she likes.

As Tyler is explaining his thinking, Shenn cuts in. "The last home game wasn't that big a deal. We were playing against the Titans and they had a pretty pathetic record coming into the game. Remember, it rained a bit before the game, too, which probably kept some people at home. This game's going to be a biggie, with strong implications for each teams' chances for a postseason bowl game. Don't you think it's much more likely that you could sell more programs for this game?" Tyler agreed but felt overwhelmed with having to figure out how many more. Shenn helped him understand the difference between the unit *demand* for programs and unit *sales*: The (unit) demand consists of the number of programs that would be sold if he had an unlimited number available, whereas (unit) sales are limited by the number of programs ordered. "Suppose you order 8,000 programs. If demand turns out to be for 9,000 programs, then you will sell all 8,000 that you ordered, but no more. However, if demand is for 7,000 programs, then sales will also equal 7,000 and you will have 1,000 unsold programs."

D. Chhajed and T.J. Lowe (eds.) *Building Intuition*: *Insights From Basic Operations Management Models and Principles*.
doi:10.1007/978-0-387-73699-0,

"I get it. But I don't want to have any leftovers. That will make me look bad. I still like that 7,756 number. I'm almost sure to sell them all. They sell for $5.00 and cost only $1.25, so I'd get the full $3.75 for each of them."

Shenn looked like she was settling in for a long discussion. "Have you thought about trying to look good, rather than avoiding looking bad? If you could increase the net return from selling programs, the proceeds would be mightily appreciated by the department. I think that what my econ prof called *opportunity cost* might be useful here. Suppose you decide to order 8,000 programs. It's certainly true that if you only sell 7,000, then you will have 1,000 unsold programs that cost $1.25 each, for a total of $1,250 and if you had ordered only 7,000 you would have saved that $1,250, without changing how many you sold. However, if demand is for 9,000 programs, then you will only sell 8,000, and if you had ordered 1,000 more, you would have sold them all, for a net profit of $3.75 each, representing a total gain of $3,750. The $3,750 opportunity cost of having 1,000 too few is much more than the $1,250 opportunity cost of having 1,000 too many. It may make sense to err on the side of ordering too much rather than too little. What do you think?" Tyler had to agree, but he could feel a headache coming on. "Let's go talk to Nar about how to think about the possible demand quantities for this game."

It did not take Nar long to power up his PC and print out a histogram of program sales for the past 3 years.

Tyler was the first to comment on the chart, "Well, I can see that we've never sold more than 9,500 programs, so I don't have to worry about ordering more than that!"

Shenn jumped in, "Not so fast, Bubba-head. You're forgetting the difference between demand and sales. When we sold 9,500 programs, that is all we ordered. We likely would have sold more if we had them. About all we can say is that the demand will be random. Let's figure out how many to order for an example. We can then go back and fine-tune our assumptions. Let's assume that there are only five possible outcomes, 7,000, 8,000, 9,000, 10,000, and 11,000, with probabilities of 0.1, 0.2, 0.4, 0.2, and 0.1, respectively."

Tyler exclaimed, "I like the idea of ordering 7,000 programs, then. We get the full $3.75 margin on each, for a total return of $26,250, without any unsold programs." He was also thinking to himself that he didn't even need to remember what probabilities meant to reach this conclusion.

"But you might be able to do better if you order more. The usual way to evaluate a decision like this is to determine the *expected value* of the resulting return. We get it by determining the net profit for each possible outcome, multiplying it by the probability of that outcome, and then adding up over all outcomes. If I remember what my stats prof said, this measure corresponds to the *average* net profit that would be obtained per football game if exactly the same problem were faced many times. Let's see what we would get if we ordered 8,000 programs." It didn't take long to create the worksheet shown in Exhibit 7.1.

Tyler felt his headache reappear. Shenn explained, "This says that ordering 8,000 yields an expected net return of $29,500, a full $3,250 more than the $26,250 from ordering 7,000. If demand is 8,000 or more, we sell all 8,000, for a net return of

Exhibit 7.1. Analysis of 8,000

Demand outcome	Probability	($) Net revenue	Product
7,000	0.1	25,000	2,500
8,000	0.2	30,000	6,000
9,000	0.4	30,000	12,000
10,000	0.2	30,000	6,000
11,000	0.1	30,000	3,000
Total	1.0		29,500

Exhibit 7.2. Analysis of 9,000

Demand outcome	Probability	($) Net revenue	Product
7,000	0.1	23,750	2,375
8,000	0.2	28,750	5,750
9,000	0.4	33,750	13,500
10,000	0.2	33,750	6,750
11,000	0.1	33,750	3,375
			31,750

$30,000, which is $3,750 more than we get if we order only 7,000. And the probability of that happening is 0.9. The only time we get less is if demand is for only 7,000. We get $1,250 less in that case, but the probability of that happening is only 0.1. That is, we have a 90% chance of getting $3,750 more and a 10% chance of getting $1,250 less. The expected increase is 0.9(3,750) – 0.1(1,250) = 3,250. Looks to me like ordering 8,000 is quite a bit better than ordering 7,000. Let's see what we get if we order 9,000" (Exhibit 7.2).

Tyler's enthusiasm began to grow, "Let's see. If only 7,000 programs are demanded, then we net $3.75 on each of them, for $26,250 and lose $1.25 on each of the 2,000 we bought but didn't sell, for a net return of $ 26,250 – $ 2,500 = $23,750. We get $5,000 more if demand is 8,000 instead of 7,000, because we sell each of them for $5.00 and haven't bought any more. We get another $5,000 if demand increases by another 1000 to 9000. And, …, aha! If demand increases yet another 1000 to 10,000, we don't get any more, because we only have 9,000 to sell. This says our expected return is $31,750, which is $2,250 more than when we order 8,000. Can that be right?"

"Absolutely. If demand is 9,000 or more, we get $3,750 more than with ordering 8,000, and the probability of that is 0.7. If demand is 8,000 or less, we get $1,250 less, and the probability of that is 0.3. The expected gain is 0.7(3,750) – 0.3(1,250) = 2,250."

Tyler felt a bell ring in his head, "Wait, wait. Let me guess. If we order 10,000 instead of 9,000, then the expected gain will be 0.3(3,750) – 0.7(1,250) = 250, because the probability of getting the $3,750 more is the probability that demand will be 10,000 or more, which is 0.3. I can't believe that we should order that many!"

"Get used to it, baby. What about ordering 11,000?"

"Well, let's see. The expected gain will be 0.1(3,750) – 0.9(1,250) = – 750, so we would lose by making that change. We should stay at 10,000 programs. Yay!"

What is Tyler's Newsvendor Problem?

Tyler's problem is an example of what is often called the newsvendor problem. Its name derives from the context of a newsvendor purchasing newspapers to sell before knowing how many will be demanded that day. In the simplest version of the newsvendor problem, such as Tyler's, the opportunity costs of an overage or underage are proportional to the size of that overage or underage, with unit costs of c_O and c_U, respectively. For example, Tyler's unit overage cost c_O is \$1.25: It is the additional return that would be received if, in the event that too *many* programs were ordered, one *fewer* program was ordered. Tyler would have saved the \$1.25 unit cost for each program that was not needed. Similarly, Tyler's unit underage cost c_U is \$3.75: It is the additional return that would be received if, in the event that too *few* programs were ordered, one *more* program was ordered. Tyler would have sold each program that was demanded but unavailable for \$5.00. He would have had to pay the \$1.25 unit cost for each of them, netting a margin of \$3.75 per program. We will discuss more general versions of the newsvendor problem at the end of this chapter.

As far as Tyler is concerned, he is happy with the solution to his problem. However, we want to build a model that represents a general form of his problem and see if we can find a systematic way to solve it. We also would like to see if there are any insights we can glean from the analysis that can aid our intuition.

Building Intuition Tyler's decision of how many programs to order requires evaluating the *tradeoff* between having too many and too few. There is something good and bad for each outcome. If he has too many, he has an *overage*. The bad part is that he has leftover unsold programs, which he bought but are now worthless. The good part is that every customer is satisfied (able to buy a program when desired) and, hence, there are no shortages. If he has too few, he has an *underage*. The bad part is that some customers were unsatisfied (wanted to buy a program but were unable to because he ran out). He could have done better by buying more because he would have been able to sell them at a substantial margin. The good part is that he has no leftover, unsold programs.

A *newsvendor problem* has three defining characteristics: (1) there is a random amount needed of some resource, (2) a single quantity of that resource must be selected prior to observing how much is needed, and (3) all relevant economic consequences can be represented by known (opportunity) costs in terms of either the amount of overage or the underage. The optimal quantity of the resource to make available is easy to identify when the overage and underage costs are proportional to their size.

Exhibit 7.3. Probability distribution of demand

Number in list, i	Demand outcome d_i	Probability $p_i = P(D = d_i)$	Cumulative probability $P(D \le d_i)$
1	7,000	0.1	0.1
2	8,000	0.2	0.3
3	9,000	0.4	0.7
4	10,000	0.2	0.9
5	11,000	0.1	1.0

In the example that we have solved, the demand is *discrete*, which means that the demand outcomes can be put into a list, as in Exhibit 7.3. Each entry in the list is numbered, and we can therefore talk about *outcome i*, for $i = 1,2, ...,5$. It is useful to have symbols for the pertinent quantities in the model: Let d_i denote the demand (quantity) for outcome i and p_i the probability of that outcome (for each i). Let D denote the random demand that occurs. For example, the probability that demand equals 8,000 is 0.2. In symbols, $P(D = d_i) = p_i$ for each i = 1,2, ... ,5.

Let us assume henceforth that it is possible to receive a positive expected return. (We have already seen that this is true for Tyler's problem.)

When demand is discrete, then we only need to consider ordering quantities that can be demanded: It can be shown that if it is better for Tyler to order 7,001 programs than 7,000, then it is even better to order the next larger demand quantity, namely 8,000. (The expected gain by going from 7,000 to 7,001 will also be the expected gain by going from 7,001 to 7,002, and so on.)

The *optimal* amount to order is one that yields the highest expected return. The optimal amount can be determined by *marginal analysis*: We identify the expected gain of going to the next larger demand quantity from the current amount. If it is positive, we make the move, and repeat the process. In Tyler's example, if we go from 8,000 to 9,000, then the expected gain is

$$0.7(3{,}750) - 0.3(1{,}250) = 2{,}250.$$

In general, if we go from d_i to d_{i+1}, then the expected gain is

$$P(D \ge d_{i+1})(d_{i+1} - d_i)(c_U) - P(D \le d_i)(d_{i+1} - d_i)(c_o).$$

For example, if $i = 2$, then the expected gain is

$$\begin{aligned} &P(D \ge d_{i+1})(d_{i+1} - d_i)(c_U) - P(D \le d_i)(d_{i+1} - d_i)(c_o) \\ &= P(D \ge d_3)(d_3 - d_2)(c_U) - P(D \le d_2)(d_3 - d_2)(c_o) \\ &= 0.7(1{,}000)(3.75) - 0.3(1{,}000)(1.25) = 2{,}250, \end{aligned}$$

confirming what we already knew.

Finding the Optimal Solution

It can be shown that we should go from d_i to d_{i+1} if

$$P(D \leq d_i) \leq \frac{c_U}{c_O + c_U}.$$

The ratio $\gamma = c_U/(c_O + c_U)$ is called the *critical fractile* which, in Tyler's case, equals $c_U/(c_O + c_U) = 3.75/(1.25 + 3.75) = 3.75/5 = 0.75$. That is, if the inequality above holds for a particular possible demand quantity, then it is better to stock more than that quantity. For example, the probability that demand is no more than 7,000 is 0.1, which is less than 0.75, so it's better to stock more than 7,000, namely at least 8,000. Similarly, it's better to stock more than 8,000 and also 9,000. However, the probability that demand is no more than 10,000 is 0.9, which is *more* than 0.75, so it is not optimal to go to 11,000, which we already knew.

It is nice that finding the optimal amount is simple: we just go down the list of cumulative probabilities, find the last one that is less than the critical fractile, and the optimal level (to stock) is the next larger demand outcome in the list. For example, going down the list of cumulative probabilities in the rightmost column of Exhibit 7.3, the last one that is less than 0.75 is 0.7, and the optimal level is the next larger demand outcome, namely, $d_4 = 10{,}000$.

There are other ways to find the optimal amount. For example, we can start at the bottom of the list of cumulative probabilities and find the smallest one in that list that is more than the critical fractile. That demand quantity is optimal.

If one of the cumulative probabilities equals the critical fractile, then there are *alternative optimal solutions*: The demand quantity whose cumulative equals the critical fractile is optimal and so is the next larger demand quantity: Each quantity yields the same expected return. (The marginal return from going to the next larger demand quantity is zero, so there is a tie for what is best.)

Now that we know how to find the optimal solution, we do not need to restrict our consideration to simple problems with only a few possible different demand quantities. Instead of only five different quantities, we could have hundreds or thousands, and it would still be easy to find the optimal solution.

Continuous Demand Distributions

In practice, it is often convenient to use *continuous* (cumulative) probability distributions, such as the *normal distribution,* as representations for the probability distributions of demand. Technically, this requires thinking that it is possible for a fractional number of programs, such as 7,337.473, to be demanded. However, from a practical perspective, we are only *approximating* the demand distribution. We are building a model of a practical problem and virtually any model of a managerial situation makes simplifying assumptions to aid the analysis. The art of building and analyzing models in this regard is to develop an instinct for which simplifying assumptions are good ones, in the sense that they facilitate the analysis and do not mislead the results in any important way. Use of continuous distributions to approximate random quantities that are discrete is widely accepted as being a potentially good simplifying assumption.

When the demand distribution is continuous, it is also easy to determine the optimal quantity: We select the quantity Q such that

$$P(D \leq Q) = \frac{c_U}{c_O + c_U} \cdot$$

Back to Tyler's Problem

Tyler and Shenn returned their attention to Nar and, after lengthy discussion and some statistical analysis, decided that a better representation of the random demand for programs (for the coming game) is a normal distribution with mean of $\mu = 9{,}000$ and a standard deviation of $\sigma = 2{,}000$. Exhibit 7.4 illustrates the normal distribution and the critical fractile solution.

Normal Distribution

When the demand distribution is a normal distribution, then we can find the optimal order quantity either by using tables, found in statistics books, or by using built-in functions of software packages, such as Excel spreadsheets. Exhibit 7.5 makes it easy not only to find the optimal order quantity but the resulting expected overage and underage costs. The first column gives the critical fractile, γ, and the second

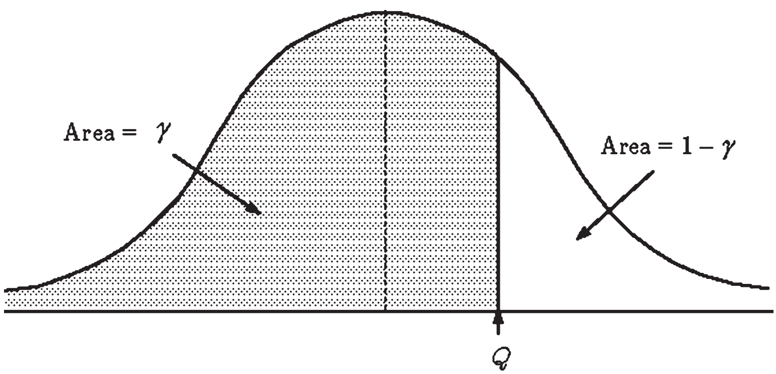

Exhibit 7.4 Critical fractile solution

Critical Fractile γ	$z = z^*(\gamma)$	$L_N(\gamma)$
0.05	-1.645	0.109
0.10	-1.282	0.195
0.15	-1.036	0.274
0.20	-0.842	0.350
0.25	-0.674	0.424
0.30	-0.524	0.497
0.35	-0.385	0.570
0.40	-0.253	0.644
0.45	-0.126	0.720
0.50	0.000	0.798
0.51	0.025	0.814
0.52	0.050	0.830
0.53	0.075	0.846
0.54	0.100	0.863
0.55	0.126	0.880
0.56	0.151	0.896
0.57	0.176	0.913
0.58	0.202	0.931
0.59	0.228	0.948
0.60	0.253	0.966
0.61	0.279	0.984
0.62	0.305	1.002
0.63	0.332	1.020
0.64	0.358	1.039
0.65	0.385	1.058
0.66	0.412	1.078
0.67	0.440	1.097
0.68	0.468	1.118
0.69	0.496	1.138
0.70	0.524	1.159

Critical Fractile γ	$z = z^*(\gamma)$	$L_N(\gamma)$
0.71	0.553	1.180
0.72	0.583	1.202
0.73	0.613	1.225
0.74	0.643	1.248
0.75	0.674	1.271
0.76	0.706	1.295
0.77	0.739	1.320
0.78	0.772	1.346
0.79	0.806	1.372
0.80	0.842	1.400
0.81	0.878	1.428
0.82	0.915	1.458
0.83	0.954	1.489
0.84	0.994	1.521
0.845	1.015	1.537
0.850	1.036	1.554
0.855	1.058	1.572
0.860	1.080	1.590
0.865	1.103	1.608
0.870	1.126	1.627
0.875	1.150	1.647
0.880	1.175	1.667
0.885	1.200	1.688
0.890	1.227	1.709
0.895	1.254	1.732
0.900	1.282	1.755
0.905	1.311	1.779
0.910	1.341	1.804
0.915	1.372	1.831
0.920	1.405	1.858

Critical Fractile γ	$z = z^*(\gamma)$	$L_N(\gamma)$
0.925	1.440	1.887
0.930	1.476	1.918
0.935	1.514	1.951
0.940	1.555	1.985
0.945	1.598	2.023
0.950	1.645	2.063
0.955	1.695	2.106
0.960	1.751	2.154
0.965	1.812	2.208
0.970	1.881	2.268
0.975	1.960	2.338
0.980	2.054	2.421
0.985	2.170	2.525
0.986	2.197	2.549
0.987	2.226	2.575
0.988	2.257	2.603
0.989	2.290	2.633
0.990	2.326	2.665
0.991	2.366	2.701
0.992	2.409	2.740
0.993	2.457	2.784
0.994	2.512	2.834
0.995	2.576	2.892
0.996	2.652	2.962
0.997	2.748	3.050
0.998	2.878	3.170
0.9985	2.968	3.253
0.9990	3.090	3.367
0.9995	3.290	3.555
0.9999	3.719	3.952

Exhibit 7.5 Newsvendor solution for the normal distribution

gives the stock level, expressed as the number $z^*(\gamma)$ of standard deviations above the mean. In Tyler's problem, $\gamma = 0.75$, so we see that $z^*(0.75) = 0.674$. Thus, the optimal stock level is

$$Q^* = \mu + z^*(\gamma)\,\sigma = 9{,}000 + 0.674(2{,}000) = 10{,}348.$$

The third column in Exhibit 7.5 shows the cost factor $L_N(\gamma)$, which gives the resulting expected overage and underage costs as follows:

$$c_O\,\sigma\,L_N(\gamma).$$

In Tyler's problem, we have

$$c_O\,\sigma\,L_N(\gamma) = 1.25(2{,}000)(1.271) = 3178,$$

which is the optimal expected opportunity cost. But Tyler is not satisfied with this number, because he can not relate it to the expected return, which is what counts. The formula for the expected return in Tyler's problem is

$$3.75\mu - c_O \sigma L_N(\gamma) = 3.75(9{,}000) - 1.25(2{,}000)(1.271) = 33{,}750 - 3{,}178 = 30{,}572,$$

which we will now explain.

The 3.75μ is the expected return Tyler would get if he could defer deciding how many programs to order until after seeing how many programs would be demanded. In this case, for each quantity demanded, exactly that number would be ordered, and the full margin of \$3.75 would be received on each of them. The resulting expected return would be 3.75 times the expected demand. This is sometimes called the *expected return under perfect information* because it corresponds to having a magic elf that has perfect foresight in that he can perfectly see ahead to what the random demand will be and whisper the amount to Tyler before he places his order. The expected overage and underage costs is also called the *expected value of perfect information (EVPI)* because it is how much better Tyler would do if he had the services of that magic elf: Because he must decide on how many programs to order before knowing what the random demand will be, the overage and underage costs would be eliminated if he had perfect information.

Insights

There are useful insights that can aid your intuition when facing similar problems. They are easiest to understand when the probability distribution of demand is continuous, because in that case there is no chance that demand will exactly equal the amount ordered. Either $Q > D$ which means an overage occurs or $Q < D$, which means an underage occurs.

Both the unit overage and underage costs, c_O and c_U, are strictly positive. (We will clarify below that the problem is improperly formulated otherwise.) The critical fractile $\gamma = c_U/(c_O + c_U)$, which is between zero and one, gives the *optimal overage probability*, $P(D < Q) = P(D \leq Q)$, namely the optimal probability that you will have too much. Similarly, $1 - \gamma$ is the optimal probability that you will have too little. The solution can be interpreted in gambling parlance: It is optimal to run odds of c_U to c_O of having too much, which means that the probability of having too much is $c_U/(c_U + c_O)$. Odds do not depend on their units, so we can also say it is optimal to run odds of c_U/c_O to 1 of having too much. For example, if the (unit) underage cost is 3, 4, or 9 times the overage cost, then it is optimal to run a 75, 80, or 90% chance, respectively, of having too much. It is intuitive that the higher the relative cost of having too little is, the higher the optimal probability of having too much should be. The newsvendor model gives us more than that, namely, the exact way in which the probability should change.

One insight is that *never having leftover programs is bad, not good.* Tyler thought he would look good by not having any leftover, unsold programs, but our analysis shows that he should run a 75% chance (3–1 odds) of having leftovers. Tyler's boss needs to understand this, too, so he or she doesn't implicitly put pressure on Tyler to order too few and thereby receive lower expected returns.

A related insight is that *never having unsatisfied customers* (who tried to buy a program but were unable to, because the programs were sold out) *is also bad*. Tyler should run a 25% chance of having unsatisfied customers.

Appeasing Unsatisfied Customers

Tyler took a marketing class once and he remembered that the mantra in that class was never to have unsatisfied customers. He could feel the headache returning. Unsatisfied customers might decide that they would stop attending games in the future, or would not bother trying to buy a program at future games. They might have such a bad taste in their mouths that they decide not to donate any money to the athletic department. These all have negative financial consequences that were not considered in our previous analysis. If Tyler can assess the expected financial loss from each such unsatisfied customer, then he can redo the analysis with the new considerations. For example, suppose he determines that each unsatisfied customer will yield an expected loss, of the types discussed, of \$7.50. Then the underage cost changes to become $c_U = 3.75 + 7.50 = 11.25$. In the event of an underage (programs selling out), then, for each additional program, not only would it get sold, yielding a margin of \$3.75, but the expected shortage cost of \$7.50 would be eliminated. The new critical fractile is $c_U/(c_O + c_U) = 11.25/12.50 = 0.90$, which means that more programs should be ordered, the optimal probability of having no unsatisfied customers is 90%, and the optimal probability of having unsatisfied customers has been cut down to 10% from 25%. The optimal opportunity costs increase and the optimal expected return decreases.

Shenn, who took the same marketing class with Tyler but paid more attention to it, suggests that the analysis should not be left there. If there is some way to appease the unsatisfied customers that costs less than \$7.50, then the athletic department would be better off doing that. For example, suppose all unsatisfied customers would be so happy to receive a \$2 voucher to be redeemed at any food and drink concession at the game that they would carry no lingering resentment about not being able to buy a program. Then the underage cost could be reduced to $c_U = 3.75 + 2 = 5.75$, Tyler would order fewer programs than under the \$7.50 shortage cost, and the expected return would increase. There are a couple of practical problems that must be solved in cases like this: The first is determining exactly how much is needed to appease an unsatisfied customer. (The fact that this amount might be random (differ across different potential unsatisfied customers) is not a problem, as long as we can estimate the *expected* amount.) The second is to assure that only truly unsatisfied customers are given the appeasement. For example, Tyler does not want any unsatisfied customers telling their friends and family at the game that all they need to do is pretend they want to buy a program and they will receive a \$2 voucher. (The expected shortage cost could be much more than $2 for each (truly) unsatisfied customer in that case.)

Comparative Statics (Sensitivity Analysis)

In the previous discussion, we saw that as the underage cost increases, the critical fractile, and hence the optimal probability of an overage and the optimal order quantity, increases. As the overage cost increases, the opposite happens.

If either the overage or underage cost increases, then the optimal expected costs increase, and for situations such as Tyler's, the optimal expected return decreases. Of course, if either the underage or overage cost decreases, the opposite happens.

Suppose for the remainder of this section that the demand distribution is normal. If the critical fractile equals 0.5, then it is optimal to stock the mean demand (zero standard deviations above the mean). If the critical fractile is above 0.5, then it is optimal to stock more than the mean and if the standard deviation increases, then the optimal order quantity increases. However, if the critical fractile is below 0.5, then it is optimal to stock *less* than the mean and if the standard deviation increases, then the optimal order quantity *decreases*. Regardless of the value of the critical fractile, the optimal expected opportunity costs increase as the standard deviation increases. The intuition here is that the effect of more uncertainty (increasing the standard deviation) on the optimal order quantity depends on which is higher, the overage or underage cost, but the effect on the costs is always bad.

If the mean demand increases by one (and no other changes are made), then the optimal order quantity also increases by one, but the optimal expected opportunity costs remain the same. The expected return increases. Tyler has the incentive to increase the mean (demand) and reduce the standard deviation.

One Stocking Decision for a Perishable Product

The newsvendor model applies to problems of determining a single order quantity when unsold units of the product ordered will be obsolete and cannot be offered for sale again. Tyler's problem fits this description because the programs ordered for one game cannot be legitimately offered as the program for any subsequent game. Newspapers and magazines are similar.

The newsvendor model can be useful for determining stock levels of perishable produce at a grocery store: For example, if a grocery store gets new lettuce every day and throws out the old, unsold lettuce (or gives it to a food bank or otherwise diverts the leftover unsold lettuce), then the grocery store faces a newsvendor problem each day for lettuce. If you are a customer of such a grocery store and that store almost never runs out of lettuce, then either the margin on lettuce is extremely high, or the grocer estimates the financial cost of a customer finding no lettuce to be extremely high, or both. Incidentally, it is heard on the street that grocery stores get a significant amount of their profits from their produce sections. The newsvendor model does not capture all of the dynamics of selling produce at a grocery store. For example, if the grocer sells out of one type of lettuce, a customer may buy a different type of lettuce, as a substitute. To the extent that this kind of substitution takes place, the effective cost of a shortage is less and, thus, the optimal stock level can be lower. (If there is a type of lettuce that is the heavy favorite to be selected as a substitute for other types when they are out of stock, the way vanilla ice cream is the favorite substitute for other flavors, then the optimal stock level for that type of lettuce may be higher than if the effect of substitution were not considered.)

One Stocking Decision for a Nonperishable Product

The newsvendor model can be useful for determining how much to order of a product that requires a long lead time to produce and will be sold only over a relatively short season. In this case, there is still only one stocking decision. Much fashion-oriented clothing falls into this category: It takes many months to procure materials, manufacture the clothing, and transport it to the location/country where it will be sold. Unsold merchandise at the end of the season is generally sold to another organization at a deep discount. In contrast to restocking lettuce, which is done daily, this decision would be made once for the selling season (such as the Christmas season). The overage cost would be the unit materials, assembly, and transportation costs less end of season scrap value.

An Improper Formulation

Something is wrong with the formulation of the above problem if the overage cost is negative: This would happen if the end of season scrap value is assessed as more than the sum of the unit materials, assembly, and transportation costs. In this case, it would be optimal to buy an infinite amount, and each of the (infinite) number of unsold goods would net more scrap value than what they cost, yielding an infinite total return. In actuality, it is likely that if there are only a few unsold units, then they can be scrapped at a reasonably high value, perhaps more than what they cost. As the number of unsold units increases, the average scrap value decreases. A model of this situation is discussed further at the end of this chapter.

Stocking Decisions Under Buy-Backs

Although the newsvendor model is nominally designed for determining the stock level of a product that becomes obsolete by the end of the selling season, it can also be useful for determining the stock levels of a product that still retains a great deal of value at the end of the selling season. In particular, here we assume that the supplier offers *buy-backs,* which means that the supplier will buy back any unsold units. For example, Tyler may also be responsible for deciding how many cases of soda to order for the drink stands to sell during the game, and unsold cases can be returned to the supplier for a substantial credit. Suppose that each case of soda costs \$4 from the supplier including delivery and that every unsold case of soda can be returned for a net of \$3 per case (\$3.50 less \$0.50 for labor and transportation getting the soda back to the supplier). If each case sells for \$24, then the unit underage cost would be \$20. (In case of an underage, Tyler could have bought one more case for \$4, sold it for \$24, and netted \$20.) The unit overage cost would be \$1. (If there is an overage, Tyler could have bought one fewer case for \$4, which would therefore not be available to be returned for \$3, for a net gain of \$1.) The critical fractile is 20/(20 + 1) = 0.953, so Tyler should run a 95.3% chance of having unsold soda and a 4.7% chance of running out.

Shenn again offers an opportunity for improvement by suggesting that Tyler not return unsold cases of soda until the end of the football season, because, as long as the unsold cases can be stored safely (without deterioration) between games for less than \$1 each, then Tyler should do that. For example, if a case of soda can be put into storage and put back into position for sale at the next game for \$0.60, then that approach would save \$0.40 per case. The unit overage cost becomes \$0.60. (If there is an overage, he could have bought one fewer case for \$4. That benefit would be cancelled out because it would therefore not be available to be sold at the next game, and, thus, a new case would have to be purchased then for \$4 delivered. The benefit would come from not having to spend the \$0.60 to store it between games.) The critical fractile increases and the expected costs decrease. Shenn warns him that this logic is valid only when he is assured that all unsold cases would require replacing at the next game if they were not available. For example, if Tyler has 1000 cases of unsold soda at the end of one big game and the next game is a small one where only 800 cases would be purchased if there were no leftover soda, then 200 of the unsold cases did not require replacing at the next game and a different unit overage cost applies to them. The overage costs would not be proportional to the size of the overage. This case of nonlinear overage costs is discussed at the end of this chapter.

Improper Formulations and Their Resolution

An interesting situation arises if there are no handling, transportation, or other transaction costs and cases of soda can be returned for their purchase price. An initial analysis would yield an overage cost of zero. (In the case of an overage, Tyler could have bought one fewer case for \$4, which would therefore not be available to be returned for \$4, for a net gain of \$0.) The critical fractile would be $c_U / c_U = 1$, which means that Tyler should run absolutely no risk of having a shortage. If demand is really normally distributed, then there is no finite number Q such that $P(D \leq Q) = 1$, so Tyler would have to stock an infinite amount. This is obviously not a practical solution. Something is wrong with the formulation. One resolution is to find that there must be some cost that was left out of the analysis. For instance, Tyler would not have enough room to store all the cases of soda in the world at the stadium, so space would have to be rented, which would cost something and create a nonzero overage cost, and, therefore, a finite optimal stock level. There are many other possible resolutions. For example, the supplier would realize that it would be a bad idea to absorb the handling and transportation costs on tons of soda that almost surely would be returned. The supplier would be inclined to require Tyler to share in some of the cost, and even seek at least a small gain for all soda that is returned.

Another resolution of the improper formulation is to incorporate the opportunity cost of capital. For example, for some small expensive items, such as jewelry, the handling and transportation costs are very small compared to the cost of the item. The overage cost should also include any holding cost that is incurred for unsold items, which in this case, includes the interest costs of the money needed to purchase the items. Even if the unsold items can be sold back at the same price they were bought at, having those items for a period of time requires the commitment of capital, which could have been used for other productive purposes during that period.

Further Insights on Buy-Backs

Tyler figures that the more the supplier pays for buy-backs, the better off he will be. That is true if nothing else changes. However, in situations such as campus bookstores, the supplier can take nearly all the profits of the business. Professors decide what textbook to require for their courses, only one publisher prints each book, and there is a published retail price, so the bookstore has no ability to change the textbook, source it from a different supplier, or charge a higher price. The publishers then charge the bookstore a price that is nearly as high as the retail price. They also offer a buy-back rate that is nearly as high as its wholesale price. This combination means that the bookstore will make a little bit of money selling the books that are demanded, and not run any risk of having expensive, unsold textbooks after the demand has been realized. The publishers make virtually all the money. Without the buy-back system, the bookstore would have a high overage cost, leading to a low critical fractile, and a high probability that some students would not be able to buy the book, an untenable campus outcome. The bookstore would be pressured to stock more books than is optimal, and doing so might lead to losing money on a regular basis. The bookstore might even have to be closed. All these negative outcomes would likely lead to the revelation that the publishers are taking almost all of the profits. Pressure might be created on the publishers to reduce their wholesale prices. By using buy-backs, they keep their situation off the radar screen.

Publishers often go one step farther, by requiring the bookstores to send only the covers of the unsold copies of the books they buy back. This practice suggests that the shipping and handling costs incurred in sending the book back to the publisher are more than the marginal cost of printing it in the first place. (The publisher finds it cheaper to print more copies in the first place rather than have the complete books sent back for redirection to another bookstore later.) The publishers insist on getting the covers sent back, because they do not want to offer the temptation to the bookstores of requesting reimbursement (buy-backs) for books that they actually sold. The publishers often even include a message to consumers inside the cover that the book is “illegal” if it does not include its cover. (The publishers don’t want the bookstores to send the covers back, get reimbursed for them, and then sell what’s left of the books to consumers.)

Multiple Replenishment Problem for a Nonperishable Product

A useful way of understanding whether the newsvendor model is useful for determining replenishment stock levels of a nonperishable product over a sequence of selling periods is whether we can safely assume that each period’s problem starts with zero initial inventory. For instance, if unsold units available at the beginning of the next period can be returned for the same price as new ones can be purchased, then we can conceptually think of returning any unsold units and then buying them back (at the same price) as needed to get to the optimal stock level for the next period.

Similar logic is valid even if unsold units cannot be returned. Suppose that the problem is *stationary*, which means that each period is alike in the sense of the demand and the costs and revenues, and has an *infinite horizon*, which means that there are an infinite number of periods. Then the optimal stock level will be the same in each period. Thus, any unsold units at the end of a period can be used to reduce the number that are purchased at the beginning of the next period. Inventory management in this case is called *one for one* replenishment: Each unit that is sold in a period is replenished at the beginning of the next period. The overage cost is the holding cost, which includes both *financial holding costs*, which include interest costs, and *physical holding costs*, which include handling, storage, temperature control, insurance, and other costs incurred for each unit stocked for the period. The newsvendor model gives the optimal stock level. (See Chap. 9 in this volume for a discussion and use of the same idea.)

Capacity Management

The newsvendor model can be useful in determining decisions other than stock levels. For example, Tyler’s uncle, Rusty, works for an automobile manufacturer that must decide on how much *capacity* to have to manufacture each of its models for the approaching model year. The capacity set is the quantity Q and D is the (unknown, random) annual demand for that model. The overage cost includes the savings from providing one unit less capacity, and the underage cost includes the lost profit from not being able to make one more car of that type. In reality, there are other considerations that must be included in making a capacity decision like this one, such as the option to operate the plant at one, two, or three shifts a day, and/or weekends. Plants also often are designed so that multiple models can be manufactured on the same assembly line, which provides the flexibility to make more of the hot selling models and fewer of the duds. It may also be possible to make adjustments to the capacity level during the year. All these considerations can be addressed in a model, and a newsvendor model can be a good starting point for the analysis.

Airline Overbooking

Another application of the newsvendor model is to airline overbooking of flights. Tyler has another uncle, Howie, who works for an airline that faces, for each flight, the possibility that not all potential passengers who made reservations will show up for the flight. If one of Howie's flights has 100 economy seats, all 100 are reserved, but only 90 show up to board the flight, then the 10 empty seats generate no revenue, but would have if Howie had allowed 10 extra reservations to be made for the flight and all of them had shown up. The available quantity 100 of seats is fixed in advance by the type of airplane assigned to the flight. The decision Q is the number of overbooked seats to allow. The (random) demand $D(Q)$ is the number of no-shows, passengers with reservations who do not show up for the flight, as a function of the number Q of overbooked seats. If $Q > D(Q)$, then the flight is overbooked at the time of boarding: The number of no-shows is less than the number of overbooked seats, so there are some passengers who cannot board the flight. What the airlines do in this situation is offer vouchers redeemable on future flights to confirmed passengers who agree to take a later flight. Their initial offer is sometimes a free domestic flight within their system. Passengers who accept the offer are assumed to be appeased, creating no unforeseen negative consequences. If not enough passengers accept the offer, the airline will raise the offer, as in an auction, essentially paying the lowest price it can to reduce the number of passengers on the flight to its capacity.

The simplest version of this problem occurs when $D(Q)$ is independent of Q, which means that the distribution of the number of no-shows is not affected by the number of overbooked seats. In the next paragraph, we will briefly discuss the more realistic scenario of there being a dependence of D on Q. The situation of overbooked confirmed passengers corresponds to $Q > D$, which, from the perspective of the newsvendor model, is an "overage" because the quantity set (number of overbooked seats) is higher than (over) the demand (number of no-shows). Thus, the unit "overage" cost here is the expected appeasement cost that the airline would save if it had overbooked one fewer seat and therefore had one fewer overbooked passenger at boarding. (It consists of the expected cost of adding one reservation to the later flight plus the expected cost of the voucher.) The unit "underage" cost is the expected additional return the airline would obtain by increasing the number of overbooked seats by one and therefore had one more paying customer on the flight. Both of these costs can be quite large and similar in magnitude. If they were equal, the critical fractile would be 0.5, which means the airline should run an equal risk of having empty seats on the flight and having overbooked passengers at boarding. If passengers are simply appeased with a voucher for a free flight of the same type in the future (along with the seat on a later flight), and the later flight is underbooked, then the cost of appeasing passengers would be less than the additional revenue from a paying passenger, because the cost of the seat on the later flight is insignificant and the voucher has the appeased passenger simply flying for free on another flight, which might not be overbooked. Therefore, the expected cost of the voucher is less than a paid fare. In this case, the critical fractile, which is the optimal probability of having overbooked passengers at boarding, is greater than 0.5.

If we assume more realistically that the number of no-shows does indeed depend on the number of overbooked seats (D depends on Q), then, provided the relationship is not perverse, marginal analysis still yields the optimal number of seats to overbook. The expected gain from overbooking one more seat is found as follows. If that ticket holder shows up and there is room on the flight (at boarding), then the airline will receive the additional revenue from the ticket. If that ticket holder shows up and there is no room on the flight, then the gain will equal the additional revenue received from the ticket and the loss will equal the opportunity cost of the seat that will be taken on the later flight plus the opportunity cost of the voucher that will appease the customer. There is nothing to stop the airline from overbooking as many seats as possible until these costs exceed the gain from the ticket sold. The financial consequences of the ticket holder not showing up must also be worked into the equation. For example, if the ticket is fully refundable, then the airline loses the revenue from the ticket and incurs the costs of processing both the initial ticket purchase and the refund. Airlines often offer much cheaper fares for nonrefundable tickets that involve a substantial penalty to change to another flight. This arrangement presumably reduces the incidence of no-shows.

The overbooking tradeoffs discussed here also are relevant to hotels, restaurants, and other services that take reservations.

Flexible Medical Savings Accounts

This application of the newsvendor model is to personal finance. Some US employers allow their employees to specify an amount to be taken out of their pay check, before taxes, that can be used to cover medical expenses that are not covered by the employee's medical (and dental) insurance. At the urging of Shenn, Tyler is considering signing up to have \$10 per month from his pay allocated to this account. He would therefore select \$120 as the amount in his account for the year. He has good insurance coverage, but he faces a number of potential medical expenses that are not covered by any of his insurance plans. For example, he must co-pay for each office visit and a percentage of any medical procedure required. He also has no coverage for prescription eye glasses or for certain dental procedures. He is not sure what the total of these uncovered expenses will be for the year, but, with the help of Shenn, he has estimated that the probability distribution for the total is normal with a mean of \$200 and a standard deviation of \$20. If his uncovered expenses are more than \$120, then he will get the full \$120 back. Tyler's marginal income tax rate is 25%, so, because the \$120 was not taxed, Tyler has saved \$30: If Tyler had not signed up for the program, he would have had to pay that \$120 out of his after-tax earnings. So he would have had to earn \$160 in pretax earnings to yield the \$120 needed to pay the uncovered expenses. Under the program, he only needed to use \$120 of his pre-tax earnings to cover these expenses, so he can keep \$40 more in pretax earnings, which converts into \$30 in after-tax earnings. The downside of the program is that if Tyler's uncovered expenses are less than the \$120 he put aside, then he loses the unused funds.

This is a newsvendor model: Q is the amount put into the program for the year, and D is the total amount of uncovered expenses. In case of an overage, $Q > D$, then if he had reduced Q by one dollar, he could have kept that pretax dollar, which converts into \$0.75 in after-tax dollars. So the unit overage cost is $c_O = 0.75$. In case of an underage, $Q < D$, then if he had increased Q by one dollar, he would have lost the \$0.75 in after-tax dollars, but would have saved \$1 in after-tax dollars that he needed to use to pay the extra uncovered expenses. Thus, the (net) unit underage cost is $c_U = 0.25$.

The critical fractile, the optimal probability of having too much in the account, is therefore

$$\gamma = \frac{c_U}{c_U + c_O} = \frac{0.25}{0.25 + 0.75} = 0.25,$$

which, using Exhibit 7.5, means that Tyler should put

$$Q^* = \mu + z^*(\gamma)\,\sigma = 200 - 0.674(20) = 187$$

into his account.

Other Applications

There are numerous other applications of the newsvendor model, such as setting cash reserves by a bank or an individual for a checking account, selecting spare parts for a product at the end of the life of its production line, deciding how much scrubber capacity to have available to clean noxious fumes being produced, college admissions, water reservoir management, staff sizing in a service business, and prepurchased maintenance calls. See Wagner (1975), Shogan (1988), and Denardo (2002) for more.

Historical Background

The first generally accepted framing and discussion of what we now call the newsvendor problem appeared in Edgeworth (1888) in the context of a bank setting the level of its cash reserves to cover demands from its customers. His focus was on use of the normal distribution to estimate the probability that a given level is large enough to cover the total demand.

Morse and Kimball (1951) coin the term "newsboy" in introducing this problem, use marginal analysis, and implicitly provide the critical fractile solution. Arrow et al. (1951) formulate a more general problem and also implicitly provide the critical fractile solution within a special case of their solution. Whitin (1953) explicitly provides the critical fractile solution.

The problem began to be known as the *Christmas tree problem* and the *newsboy problem* in the 1960s and 1970s. In the 1980s, researchers in the field sought vocabulary, such as the "newsperson" model, that would not be gender specific. Matt Sobel's suggestion, the *newsvendor problem,* is now in general use.

So far, all discussion has been limited to the case of proportional costs. When Arrow et al. (1951) quoted the defeated Richard III as exclaiming, "A horse, a horse, my kingdom for a horse," they not only suggested that the unit underage (shortage) cost might be very high, but that it might get larger as a function of the size of the shortage. Early work, such as Dvoretzky et al. (1952) and Karlin (1958), included such nonproportional (nonlinear) costs in their analyses. Karlin (1958) also allowed for the possibility that the overage cost might be a nonlinear function of the overage.

Suppose that both the underage and overage cost functions are *convex*, which can be understood as being bowl shaped with the property that no matter how they are tilted, while being kept right side up, their only flat spots can be at the bottom. (The proportional cost case is covered.) Then the expected opportunity costs, as a function of the quantity selected, is convex, so any flat spot we find on it will be at the bottom and therefore minimize the expected costs. (Flat spots are found using calculus by computing the derivative and setting it to zero.) From a computational perspective, a solution is easily found, although not nearly as easily as computing a critical fractile γ based on the given unit costs and then finding the Q such that $P(D \le Q) = \gamma$.

Bellman et al. (1955) provide the explicit critical fractile solution for the multiple replenishment problem of a nonperishable product in the stationary, infinite horizon case. Veinott (1965) and Sobel (1981) extend the analysis to the nonstationary and finite horizon cases and to other settings.

Gallego (1997) provides a realistic formulation of the salvage value of unsold units: He assumes that there is a random quantity that can be sold at a given unit salvage value. It is also possible to include possibly many different markets in which the unsold units can be salvaged. Each market is characterized by the unit salvage value that it offers and the (possibly random) maximum number that can be salvaged at that price. This approach avoids the surprise of having many more unsold units to salvage than were anticipated.

When the purchase costs for the resource are proportional to the amount purchased, then it is straightforward to reformulate the costs as proportional to the overage and the underage. However, in many realistic situations, the purchase costs are not proportional. For example, there may be economies of scale, in that the average purchase cost may decrease in the size of the purchase. Considerations such as this have led to the development of stochastic inventory theory, with the seminal contribution of Arrow et al. (1958). Zipkin (2000) and Porteus (2002) give expositions of the current status of that theory, with Harrison et al. (2003) giving an introduction to supply chain management, a much broader field that developed rapidly with the help of researchers in inventory management. There has been an analogous development of capacity management, with Manne (1961) providing an early impetus, and Van Mieghem (2003) giving a recent review. The development of revenue management is more recent, with Talluri and van Ryzin (2004) providing an exposition of its recent status. In short, the newsvendor model is a key component of several different modern disciplines that address practical problems.

Acknowledgments The author wishes to thank Matthew Sobel for his comments and suggestions leading to a much improved document and Linda Bethel for helping convert the author's LaTex document into Word.

References

Arrow, K., T. Harris, J. Marschak (1951) "Optimal Inventory Policy," *Econometrica* **19** 250–272.

Arrow, K., S. Karlin, H. Scarf (1958) *Studies in the Mathematical Theory of Inventory and Production*, Stanford University Press, Stanford.

Bellman, R., I. Glicksberg, O. Gross (1955) "On the Optimal Inventory Equation," *Management Science* **2** 83–104.

Denardo, E. (2002) *The Science of Decision Making: A Problem-Based Approach Using Excel,* John Wiley, New York.

Dvoretzky, A., J. Kiefer, J. Wolfowitz (1952) "The Inventory Problem: I. Case of Known Distributions of Demand," *Econometrica* **20** 187–222.

Edgeworth, F. (1888) "The Mathematical Theory of Banking," *Journal of Royal Statististical Society* 113–127.

Gallego, G. (1997) "Uncertain Demand at Salvage Value," *Notes for IEOR 4000: Production Management*, Columbia University, New York.

Harrison, T., H. Lee, J. Neale 2003 *The Practice of SupplyChain Management: Where Theory and Application Converge*, Kluwer Academic Publishers.

Karlin, S. (1958) "One Stage Inventory Models with Uncertainty," in Arrow, K., S. Karlin, H. Scarf (eds) *Studies in the Mathematical Theory of Inventory and Production,* Stanford University Press, Stanford, 109–134.

Manne, A. (1961) "Capacity Expansion and Probabilistic Growth," *Econometrica* **29** 632–649.

Morse, P., G. Kimball (1951) *Methods of Operations Research*, Technology Press of MIT, Cambridge.

Porteus, E. (2002) *Foundations of Stochastic Inventory Theory*, Stanford University Press, Stanford.

Shogan, A. (1988) *Management Science*, Prentice-Hall, Englewood Cliffs.

Sobel, M. (1981) "Myopic Solutions of Markovian Decision Processes and Stochastic Games," *Operations Research* **26** 995–1009.

Talluri, K., G. van Ryzin (2004) *The Theory and Practice of Revenue Management*, Kluwer Academic Publishers.

Van Mieghem, J. (2003) "Capacity Management, Investment, and Hedging: Review and Recent Developments," *Manufacturing & Service Operations Management* **5** 269–302.

Veinott, A., Jr. (1965) "Optimal Policy for a Multi-product, Dynamic, Nonstationary, Inventory Problem," *Management Science* **12** 206–222.

Wagner, H. (1975) *Principles of Operations Research*, Prentice-Hall, Englewood Cliffs.

Whitin, S. (1953) *The Theory of Inventory Management*, Princeton University Press, Princeton.

Zipkin, P. (2000) *Foundations of Inventory Management*, McGraw-Hill/Irwin, New York.