ESTIMATING THE DEMAND DURING A LEAD TIME USING ORDER STATISTICS *

By NICHOLAS BLESSEOS (Ph. D.)

Lecturer, Athens School of Economics and Business

1. Indroduction

In the design of a forecast system, it is well known that the sequence of the basic steps, data, model, smoothing and forecast, is followed.

From time to time, however, one may encounter special situations which do not fit the general mold. This article is intended to break the flow and show that there are other approaches to special problems.

We consider a topic that has been developed for inventory control or production planning applications, namely, the problem of estimating the maximum demand during a lead time, as a basis for setting the reorder levels in an inventory control system.

Reorder Level and Maximum Demand During a Lead Time.

One reason for forecasting demand in an inventory control system is furnishing the basis for deciding whether it is now time to order more material to replenish the stock of some item. There are many ways of setting up a routine control system that will flag any item when it is time to order replenishment of the stock. There are control systems, such as a max-min system, a two-bin system, an order-point, order-quantity system, a fixed-interval system, a base-stock system and so on. Each of them has its own characteristics and areas of successful application. In each system, however, there is some number, used in the control of stock replenishment, that is related to the probability distribution governing demand in the immediate future.

Consequently, setting the reorder level in an inventory control system depends on a forecast of what the future demand will be.

The reorder level is a number posted in the unit stock record. After each posting that affects the available stock (stock on hand plus any stock already

* This paper has been written, while the author was Ass. Prof. of Quantitative Business Analysis, Wright State Univ., Dayton, Ohio.

ordered but not yet received), the new balance is compared with the reorder level. If the available stock is greater than the reorder level, no action is required. If the available stock has dropped to a point equal to, or less than, the reorder level, more stock is ordered. The quantity is added to the available stock balance so that it is now larger than the reorder level. Too high a reorder level corresponds to available stock that is greater than the demand during the lead time required to replenish the stock.

Too low a reorder level corresponds to available stock that is less than the demand during the lead time. Hence, we are led inexorably to the conclusion that the right reorder level is the maximum r e a s o n a b l e demand during a lead time. Reasonable, since it may be good strategy to run short by a little bit now and again : the cost of being short occasionally is less than the cost of supporting the inventory that would prevent the shortage. The lead time covered by the forecast is the time required from the release of an order to the receipt of the material. If the lead time is known, then the general methods can be used in forecasting the total demand through that period of time.

There are many instances, however, where the lead time itself is a stochastic variable. For example, when we order material from independent vendors, particularly those that have few formal controls in their own production planning, the lead times can vary widely from order to order.

Nevertheless, in any of these situations, the correct reorder level is still the maximum reasonable demand during a lead time. One could forecast the distribution of lead times, using the same methods that are appropriate for forecasting the distribution of demand per unit time. The order point, conceptually, could be determined from the joint probability distribution of lead times and demand.

But joint probability distributions are much harder to handle than distributions that involve only one variable. It is not always possible to deal with the distributions as if they were indepedent.

The problem now can be restated : the reorder level should be the maximum reasonable demand during-a-lead-time, where the phrase «demand-during-a-lead-time» is a single variable. If the data in a forecast system are measured in units of demand-during-a-lead-time, and if the model is in the same terms, then the forecasts that result will be estimates of the demand-during-a-lead-time and we can pick the level that is the maximum reasonable level to use as a reorder level.

In what follows, a methodology how to determine the «demand-during-a-lead-time» in probability terms is described, and more general, how we can determine a confidence interval for the demand-during-a-lead-time, using ordered observations in a sample of size n, from a population whose elements are the demands of a particular item, in some past time interval (week, month, year etc.).

All the possible cases concerning the population's distribution function are examined i.e. distribution-free population, and population following discrete or continuous distribution function.

2. Methodology

2.1. Definition of the Problem

Let π_N be a population of N elements, whose elements have distinct Y-values associated with them. These Y-values can be simply ordered as

$$Y_{(1)} < Y_{(2)} < \cdots < Y_{(N)} \tag{2.1}$$

For any fixed integer t in the interval $1 \leqslant t \leqslant N$, $Y_{(t)}$ is defined as the (t/N) - th quantile of the population π_N.

A simple random sample of size n is drawn without replacement from π_N. Denote the values associated with the sample elements, after ordering, by

$$y_{(1)} < y_{(2)} < \cdots < y_{(n)} \tag{2.2}$$

Is desired to find two-sided confidence intervals for $Y_{(t)}$ of the form $[y_{(k)}, y_{(r)}]$ where $1 \leqslant k < r \leqslant n$. From the observable random intervals, is examined first how the confidence coefficient $\Pr(y_{(k)} \leqslant (Y_{(t)} \leqslant y_{(r)})$ can be computed.

2.2. Calculation of the Confidence Coefficient

It is noted first, David (1970), that the event $(y_{(k)} \leqslant Y_{(t)})$ is the union of the disjoint compound events

$$(y_{(k)} \leqslant Y_{(t)},\ y_{(r)} \geqslant Y_{(t)}) \text{ and } (y_{(k)} \leqslant Y_{(t)},\ y_{(r)} < Y_{(t)}).$$

Thus, since $y_{(r)} < Y_{(t)}$ implies $y_{(k)} < Y_{(t)}$, follows that

$$\Pr(y_{(k)} \leqslant Y_{(t)}) = \Pr(y_{(k)} \leqslant Y_{(t)} \leqslant y_{(r)}) + \Pr(y_{(r)} < Y_{(t)}), \text{ and}$$

$$\Pr(y_{(k)} \leqslant Y_{(t)} \leqslant y_{(r)}) = \Pr(y_{(k)} \leqslant Y_{(t)}) - \Pr(y_{(r)} \leqslant y_{(t-1)}) \tag{2.3}$$

Therefore, to evaluate (2.3), it is sufficient to arrive at an expression for $\Pr(y_{(k)} \leqslant Y_{(t)})$.

Using a technique suggested in Hogg and Craig (1970), define the event :

$\{E_i\} = \{$exactly i observations in the sample have values less than or equal to $Y_{(t)}\}$.

Then, the event

$\{$At least k sample observations have values less than or equal to

$$Y_{(t)}\} = \bigcup_{i=k}^{n} \{E_i\}.$$

$$\text{Hence } \Pr(y_{(k)} \leqslant Y_{(t)}) = \sum_{i=\max(k,\, t+n-N)}^{\min(t,\, n)} \Pr(E_i) =$$

$$= \sum_{i=k}^{n} \binom{t}{i} \binom{N-t}{n-i} \Big/ \binom{N}{n}. \qquad (2.4)$$

By virture of the definition $\binom{N}{k} = 0$ if $k < 0$ or $k > N$ we can eliminate the maximum and minimum expressions in the limits of the summation. Equations (2.3) and (2.4) lead to the following,

$$\binom{N}{n} \Pr(y_{(k)} \leqslant Y_{(t)} \leqslant y_{(r)}) = \sum_{i=k}^{n} \binom{t}{i} \binom{N-t}{n-t} - \sum_{i=r}^{n} \binom{t-1}{i} \binom{N-t+1}{n-i}$$

$$= \sum_{i=k}^{r-1} \binom{t}{i} \binom{N-t}{n-i} + \sum_{i=r}^{n} \binom{t}{1} \binom{N-t}{n-i} -$$

$$- \sum_{i=r}^{n} \binom{t-1}{i} \binom{N-t+1}{n-i}$$

$$= \sum_{i=k}^{r-1} \binom{t}{i} \binom{N-t}{n-i} + \binom{t-1}{r-1} \binom{N-t}{n-r}.$$

Therefore,

$$\Pr(y_{(k)} \leqslant Y_{(t)} \leqslant y_{(r)}) = \left[\binom{t-1}{r-1} \binom{N-t}{n-r} + \sum_{i=k}^{r-1} \binom{t}{i} \binom{N-t}{n-i} \right] \Big/ \binom{N}{n}. \qquad (2.5)$$

Under the assumption that the elements of the finite population have distinct values associated with them the upper and lower confidence limits are set equal to order statistics and the exact confidence coefficient is calculated via (2.5).

In the case of non-distinct values for the population elements, the evaluated confidence coefficient is a lower bound for the true confidence coefficient.

Now consider the situation where our sample is drawn from a population with continuous cumulative distribution function F(x). More specifically, let $x_{(1)} \leqslant x_{(2)} \leqslant \ldots \leqq x_{(n)}$ represents the values of a random sample of size n from a population with continuous c.d.f F(x). It is known, David (1970), that

$$\Pr(x_{(k)} \leqslant X_{(p)} \leqslant x_{(r)}) = I_p(k, n-k+1) - I_p(r, n-r+1)$$

$$= \sum_{i=k}^{r-1} \binom{n}{i} p^i (1-p)^{n-i}, \qquad (2.6)$$

where X_p denotes the p-th quantile of the population, defined as $\int_{-\infty}^{X_p} f(x)\, dx = p$, and Ip(k,n-k + 1) is Karl Pearson's incomplete beta function

$$I_p\,(n_1,n_2) = \int_0^p \frac{\Gamma\,(n_1+n_2)}{\Gamma(n_1)\,\Gamma(n_2)}\, u^{n_1-1}\,(1-u)^{n_2-1}\,du \qquad 0 < u < 1 \qquad (2.7)$$

From equation (2.6) it is obvious that the random variable $v_t = F\,(x_{(t)}) = Pr\,(x \leqslant x_{(t)})$, follows the Beta distribution function,

$$f\,(v_t) = Be\,(t, n-t+1) = \begin{cases} \dfrac{\Gamma\,(n+1)}{\Gamma\,(t)\,\Gamma\,(n-t+1)}\, v_t^{\,t-1}\,(1-v_t)^{n-t} & 0 \leqslant v_t \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\text{with expected value } E(v_t) = \frac{t}{n+t}$$

$$(2.8)$$

$$\text{and variance } Var\,(v_t) = \frac{t\,(n-t+1)}{(n+1)^2\,(n+2)}\,.$$

In the case where our distribution function is discrete we have the following.

Let $x_{(1)} \leqslant x_{(2)} \leqslant \ldots \leqslant x_{(n)}$ denote an ordered random sample of size n on a variate X which may take on the values 0, 1, 2, . . . with probabilities p (0), p (1), p (2), . . . respectively, where $p\,(\alpha) \geqslant 0$ and $\sum_\alpha p\,(\alpha) = 1$.

When X takes on only a finite number of values (say 0,1,. . . .,M) we interprete $p\,(M + \alpha) = 0, \alpha = 1, 2 \ldots$. Let $P\,(x) = \sum_{\alpha=0}^{x} p\,(\alpha)$ be the distribution function of X. Let p be a fixed real number such that $0 < p < 1$. Define β to be that integer such that $P\,(\beta-1) < p \leqslant P\,(\beta)$. Now,

$$\begin{aligned} Pr\,(x_{(k)} \leqslant \beta \leqslant x_{(r)}) &= Pr\,(x_{(k)} \leqslant \beta) - Pr\,(x_{(r)} < \beta) \\ &= Pr\,(x_{(k)} \leqslant \beta) - Pr\,(x_{(r)} \leqslant (\beta-1)) \qquad (2.9) \end{aligned}$$

Khatri (1963) gives

$$Pr\,(x_{(k)} \leqslant \beta = k \binom{n}{k} \int_0^{P(\beta)} \omega^{k-1}\,(1-\omega)^{n-k}\,d\omega = I_{P(\beta)}\,(k, n-k+1).$$

Therefore,

$$Pr\,(x_{(k)} \leqslant \beta \leqslant x_{(r)}) = I_{P(\beta)}\,(k, n-k+1) - I_{P(\beta-1)}\,(r, n-r+1). \qquad (2.10)$$

This term is greater than or equal to $I_p\,(k, n-k+1) - I_p\,(r, n-r+1)$ the corresponding result for the continuous case.

Furthermore,

$$\Pr(x_{(k)} < X_p < x_{(r)}) = \Pr(x_{(k)} \leqslant \beta-1) - \Pr x_{(r)} \leqslant \beta)$$
$$= I_{P(\beta-1)}(k, n-k+1) - I_{P(\beta)}(r, n-r+1)$$
$$\leqslant I_P(k, n-k+1) - I_p(r, n-r+1)$$

3. Numerical Illustration

From the demand values of a particular item in the last 100 days, we take randomly the following ten sample values for the demand :

$$10, 20, 7, 15, 30, 4, 8, 41, 35, 12. \qquad (3.1)$$

For administrative purposes, we like to know, for the demand-during-a-lead-time in the future, what is the chance that the 25-th ordered demand value out of 100 in total, will be between the values 7 and 12.

After ordering our sample observations, 4,7,8,10,12,15,20,30,35,41 note that the demand values 7 and 12 are actually the second and the fifth ordered sample values, i.e. $y_{(2)} = 7$ and $y_{(5)} = 12$. The problem now is to find the

$$\Pr(y_{(2)} \leqslant Y_{(.25)} \leqslant Y_{(5)}), \text{ using the expression (2.5).}$$

Hence

$$\Pr(y_{(2)} \leqslant Y_{(.25)} \leqslant y_{(5)}) = \left[\binom{25-1}{5-1}\binom{100-25}{10-5} + \sum_{i=2}^{4}\binom{25}{i}\binom{100\text{-}25}{10\text{-}i}\right] \Big/ \binom{100}{10}$$

$$\simeq 0.7138$$

Consider now, that the demand-during-a-lead-time in the future is the median of our 99 population values; (quantiles like population median are of great interest to many practitioners). We like to find, what is the chance, that this value will lie between the sample demand's values 8 and 30.

Then, using again the expression (2.5), with $N = 99$, $n = 10$ $t = 50$ and $y_{(3)} = 8$, $y_{(8)} = 30$, yields

$$\Pr(y_{(3)} \leqslant Y_{(.50)} \leqslant y_{(8)} = \left[\binom{49}{7}\binom{49}{2} + \sum_{i=3}^{7}\binom{50}{i}\binom{49}{10-i}\right] \Big/ \binom{100}{10}. \qquad (3.2)$$

It is obvious that expressions like (3.2), and even more complicated, are not easy to handle, even using a good calculator. For this reason a Fortran program, which calculates the confidence coefficient $\Pr(y_{(k)} \leqslant Y_{(t)} \leqslant y_{(r)})$, was written for use on the UNIVAC 1108 computer.

Suppose now, that the ten sample observations in (3.1) come from a population with a continuous cumulative distribution function. Then for the same problem, the chance that the 25-th population quantile lies between $x_{(2)}$ and $x_{(5)}$, is given by the expression (2.6) as follows :

$$\Pr\left(x_{(2)} \leqslant X_{(.25)} \leqslant x_{(5)} \right) = I_{(.25)}(2,9) - I_{(.25)}(5,6)$$

$$= 1 - I_{(.75)}(9,2) - 1 + I_{(.75)}(6,5)$$

$$= I_{(.75)}(6,5) - I_{(.75)}(9,2)$$

$$\simeq 0.93 - 0.25 \simeq 0.68.$$

4. Conclusion

Based on the simple example in the previous section, we conclude that it is easier to work with population with continuous cummulative distribution function, using just the Incomplete Beta function's tables, than with distribution-free population using the combinatorial expression in (2.5).

But on the other hand, the results are much better on the second case (using (2.5)), since actually we don't have any restriction on the population's distribution, like the continuity's one.

REFERENCES

David, H. A. (1970). Order Statistics, John Wiley and Sons, Inc., New York.

Hogg, R. V. and Craig, A.T. (1970). Introduction to Mathematical Statistics, 3rd edition. The Macmillan Company, New York.

Khatri, C. G. (1963), «Distribution of Order Statistics for Discrete Case». Ann. Inst. Stat. Math. 14 : 167 - 171.

Pearson, K. B. (1934). Tables of the Incomplete Beta-Function, Cambridge University Press, Cambridge, England.