ON THE OPTIMAL CHARACTER OF THE (s, S) POLICY IN INVENTORY THEORY[1]

By A. Dvoretzky, J. Kiefer and J. Wolfowitz

1. Introduction

It is the practice among many economists and practical people to assume that the optimal method of controlling inventories (from the point of view of minimizing expected loss or, equivalently, maximizing expected profit) is to specify two appropriate numbers, s and S say, with $0 \leqslant s \leqslant S$, to order goods when and only when the stock at hand, say x, is smaller than s, and then to order the quantity $S - x$ so as to bring the stock up to S. We shall henceforth refer to such a policy as an (s, S) policy. The authors showed in a previous paper ([**2**], pp. 194–196) that the optimal inventory policy need not be an (s, S) policy even in very simple cases. In the present paper we study in detail the problem of when an (s, S) policy is optimal for the simplest and probably the most important case of the inventory problem (see also [**1**]).

An inventory problem will be said to satisfy the (s, S) *assumption* if there exists an (s, S) policy which is an optimal ordering policy for this problem.[2] Our purpose is to obtain necessary and sufficient conditions for the validity of the (s, S) assumption for some classes of inventory problems. These conditions will be stated in terms of the distribution function of demand.

2. Statement of the problems

As stated in Section 1 we confine our study to a simple but important case. We assume that the expected loss $V(x, y)$ when x is the initial stock and y is the starting stock[3] $(0 \leqslant x \leqslant y < \infty)$ is given by

$$V(x,y) = cy + A[1 - F(y)] + \begin{cases} 0 \text{ if } y = x \\ K \text{ if } y > x. \end{cases} \tag{2.1}$$

[1] This work was sponsored by the Office of Naval Research under a contract with Columbia University and under a National Bureau of Standards contract with the University of California at Los Angeles.

[2] There may also be equally good ordering policies that are not (s, S). There are, of course, always such policies that differ from the optimal policy on sets of measure zero. However, when the (s, S) assumption is satisfied but the optimal (s, S) policy is not unique, there will be non-(s, S) optimal policies of another kind (falling, so to speak, "between" the different (s, S) policies). See the relevant remarks in Sections 3 and 4.

[3] Following the terminology of [**2**] we refer to the stock before ordering as the "initial" stock, the starting stock being obtained from the initial one by adding to it the amount ordered.

Here c, A, K are positive constants which denote, respectively, the carrying cost per unit of goods, the loss involved if it is impossible to satisfy the demand for goods, and the ordering cost. $F(z)$ is the distribution function of demand; it is the probability that the random demand D does not exceed z, i.e.,

$$F(z) = P\{D \leqslant z\}. \tag{2.2}$$

For a given initial stock x it is obviously best to order an amount $Y - x$ with $Y = Y(x)$ satisfying

$$V(x, Y) = \min_{y \geqslant x} V(x, y). \tag{2.3}$$

Since, by (2.2), $F(z)$ is nondecreasing and continuous to the right, it follows from the definition (2.1) that $V(x, y)$ is also continuous to the right and that the minimum in (2.3) always exists.

Our purpose is to characterize the distribution functions F of demand for which the above inventory problem satisfies the (s, S) assumption. There naturally arise several problems of this kind; e.g., we may require that the (s, S) assumption be satisfied for all choices of c, A and K, or we may fix c and A and require that the (s, S) assumption hold for all K, etc. Since the problem is not changed if all three parameters c, A, K are multiplied by the same positive constant, the fixing of only one parameter amounts to no restriction at all. We are thus led to the problem of characterizing the following five classes of distribution functions F:

(a) The class $C(c_0, A_0, K_0)$ of all F for which the (s, S) assumption holds when $c = c_0$, $A = A_0$, $K = K_0$.

(b) The class $C(c_0, A_0, —) = \bigcap_K C(c_0, A_0, K)$ of all F for which the (s, S) assumption holds for given $c = c_0$, $A = A_0$ and all choices of the ordering cost K.

(c) The class $C(—, A_0, K_0) = \bigcap_c C(c, A_0, K_0)$ of all F for which the (s, S) assumption holds for given $A = A_0$, $K = K_0$ and all choices of the carrying cost c.

(d) The class $C(c_0, —, K_0) = \bigcap_A C(c_0, A, K_0)$ of all F for which the (s, S) assumption holds for given $c = c_0$, $K = K_0$ and all choices of the penalty A for not satisfying demand.

(e) The class $C = \bigcap_{c,A,K} C(c, A, K) = \bigcap_{c,A} C(c, A, —) = \bigcap_{A,K} C(—, A, K) = \bigcap_{c,K} C(c, —, K)$ of all F for which the (s, S) assumption holds for all choices of c, A and K.

It is quite easy to answer problems (a) and (b). A worthwhile solution of the other problems is less obvious. The next five sections contain the solution of the five problems mentioned above. In Section 8 the solutions to problems (d) and (e) are reformulated in a form that makes their graphical verification extremely easy. Finally Section 9 contains some remarks on the extension of the above results. In particular, the important case when D can assume only integral values is considered.

3. PROBLEM (a) AND PRELIMINARY CONSIDERATIONS

Consider the function

$$g(x) = V(x, x) = c_0 x + A_0[1 - F(x)], \qquad (x \geqslant 0). \tag{3.1}$$

Let y_0 be a nonnegative number which satisfies

$$g(y_0) = \min_{x \geqslant 0} g(x). \tag{3.2}$$

(As was remarked after (2.3) such a number always exists.) Let x_0, $(0 \leqslant x_0 \leqslant y_0)$ be defined through[4]

$$x_0 = \max_{\substack{x \in A \\ 0 \leqslant x \leqslant y_0}} x \tag{3.3}$$

where

$$A = \{x \mid x \leqslant x' \leqslant y_0 \text{ implies } g(x') \leqslant g(y_0) + K_0\}.$$

These definitions become clear from Figure 1. If we name "graph" of $g(x)$ the continuous curve obtained from the set (x, z) with $z = g(x)$ by adjoining to it, at all points of discontinuity, the vertical segment $g(x) < z \leqslant g(x - 0)$, then x_0 is the abscissa of the first point to the left of y_0 where the horizontal line $z = g(y_0) + K_0$ cuts the graph of $g(x)$.

It is clear that (s, S) with $s = x_0$, $S = y_0$ is an optimal ordering policy if and only if the following conditions hold:

(i) *For every* $x_1 > y_0$ *we have* $\min_{x \geqslant x_1} g(x) \geqslant g(x_1) - K_0$;

(ii) $g(x) \geqslant g(y_0) + K_0$ *for* $0 \leqslant x < x_0$.

Condition (i) implies that one cannot do better by ordering when the initial stock x is greater than y_0 (hence also not when $x \geqslant x_0$, since by definition of x_0 one cannot do better by ordering when $x_0 \leqslant x \leqslant y_0$). Condition (ii) implies that one cannot do better than order $y_0 - x$ when the initial stock is smaller than x_0. Conversely, if it pays to order when the initial stock is some $x_1 > y_0$, then Condition (i) cannot hold, and a similar remark applies to Condition (ii).

[4] $\{x \mid \cdots\}$ indicates the set of x having the properties $\cdots$. Usually only one x has the particular properties indicated, but there may also be an interval in which case the right end point of this interval is taken as x_0.

Remark: y_0 need not be uniquely determined by (3.2). Let y_0' be another value which satisfies (3.2) and define x_0' through (3.3), in it replacing y_0 by y_0'. Then Conditions (i) and (ii) hold if and only if the analogous conditions obtained on replacing x_0, y_0 by x_0', y_0' also hold. Hence, if Conditions (i) and (ii) are satisfied for some choice of y_0, the optimal (s, S) policy is determined up to the following: As S we may take any solution y_0 of (3.2), and then $s = x_0$ is uniquely determined through (3.3), except when there exists an interval $x_1 \leqslant x < x_0$ such that $g(x) = g(y_0) + K$ throughout this interval. If this is the case then we may take $s = x$ where x is any number satisfying $x_1 \leqslant x \leqslant x_0$.

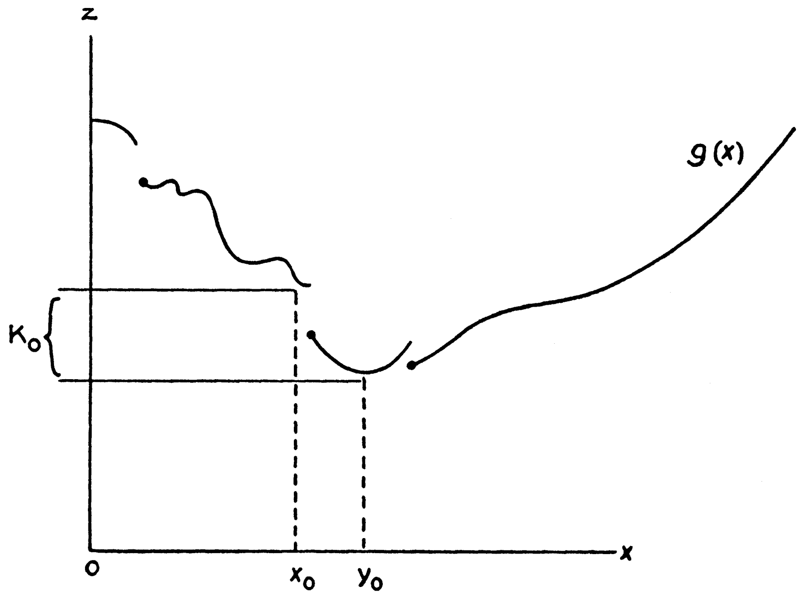

FIGURE 1

It is convenient to think of (i) as a forward condition and of (ii) as a backward condition. The former states that to the right of the minimum there are no relative minima (or maxima) of depth exceeding K_0; the latter states that in the interval $(0, x_0)$ the two sets

$$\{x \mid g(x) > g(y_0) + K_0\} \qquad \text{and} \qquad \{x \mid g(x) < g(y_0) + K_0\}$$

are separated, the first lying to the left of the second.

It is useful for the sequel to restate Conditions (i) and (ii) in a different form. To this end we note that

$$1 - \frac{g(x)}{A_0} = F(x) - \frac{c_0}{A_0}\, x. \tag{3.4}$$

Consider now the points (x, z) with $z = F(x)$, $(x \geqslant 0)$ and a straight line of slope c_0/A_0 completely above this set of points. As the line is moved downward parallel to itself it will pass through a position where it has, for the first time, a point of contact with the set (x, z).[5] Let $[y_0, F(y_0)]$ be any such point of contact. (This way of defining y_0 is

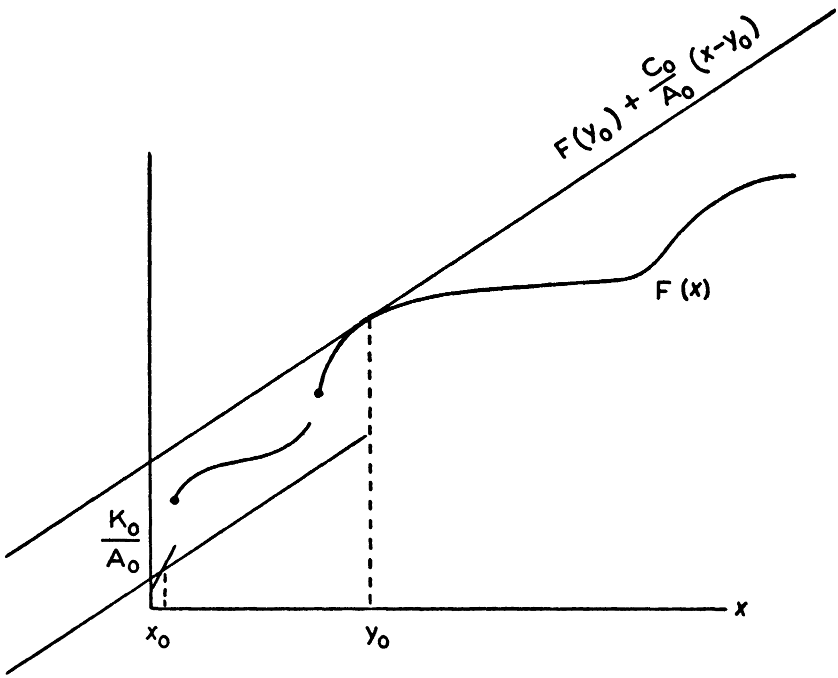

FIGURE 2

equivalent to the previous one in view of (3.4).) Then (see Figure 2) Conditions (i) and (ii) become respectively:

Condition (α). *For every* $x_1 > y_0$ *we have*

$$F(x_1) + \frac{c_0}{A_0}(x - x_1) \geqslant F(x) - \frac{K_0}{A_0} \qquad \textit{for all } x \geqslant x_1;$$

i.e., for $x > x_1$, $F(x)$ is below (more precisely: not above) the straight line of slope c_0/A_0 through $(x_1, F(x_1) + K_0/A_0)$.

Condition (β). *There exists a number* x_0, $0 \leqslant x_0 \leqslant y_0$, *such that*

$$F(x) \begin{Bmatrix} \leqslant \\ \geqslant \end{Bmatrix} F(y_0) + \frac{c_0}{A_0}(x - y_0) - \frac{K_0}{A_0} \textit{ for } \begin{cases} 0 \leqslant x < x_0 \\ x_0 \leqslant x \leqslant y_0 \end{cases};$$

[5] This follows from the fact that F is bounded, nondecreasing and continuous to the right.

i.e. the graph of $F(x)$ *to the left of* y_0 cannot be cut more than once by the straight line of slope c_0/A_0 through y_0, $(F(y_0) - K_0/A_0)$.

4. PROBLEM (b)

From Conditions (i) and (ii) it is easy to obtain a characterization of $C(c_0, A_0, —)$. All that is required is that (i) and (ii) hold for all $K_0 > 0$.

Condition (i) for all $K_0 > 0$ is obviously equivalent to: $g(x)$ is monotone nondecreasing for $x > y_0$. Similarly, Condition (ii) becomes: $g(x)$ is monotone nonincreasing for $x < y_0$.

Summing up we have: *F belongs to the class $C(c_0, A_0, —)$ if and only if $g(x)$ is* **U**-*shaped.*[6]

It follows in particular that $F(x)$ is continuous for $x \geqslant y_0$. It is also obvious that whereas S in the optimal (s, S) policy may be taken independently of K_0, this is not the case for s (except when $g(x)$ degenerates to a monotonic function in which event we may always take $s = S = 0$).

We also rephrase Conditions (α) and (β) for this case. They become respectively:

Condition (α'). *For every* $x_1 > y_0$ *we have*

$$F(x_1) + \frac{c_0}{A_0}(x - x_1) \geqslant F(x) \qquad \textit{for all } x > x_1;$$

i.e., for $x > x_1$, $F(x)$ cannot be above the straight line of slope c_0/A_0 through $[x_1, F(x_1)]$.

Condition (β'). *Every straight line of slope* c_0/A_0 *can cut the graph of* $F(x)$, $0 \leqslant x \leqslant y_0$ *at most once.*[7]

It should, perhaps, be remarked that Condition (α') implies that the probability density[8] of D cannot exceed c_0/A_0 for $x \geqslant y_0$.

5. PROBLEM (c)

In this section we characterize, in an easily verifiable manner, the class $C(—, A_0, K_0)$. For this purpose, as well as for the following sections, it is necessary to introduce some notions and notations.

Given any distribution function $F(y)$ $(0 \leqslant y < \infty)$, we denote by

[6] A function $g(x)$ is **U**-shaped if it is monotonically nonincreasing up to some value of x, and thereafter is monotonically nondecreasing ($g(x)$ may degenerate to a monotonically nondecreasing function).

[7] The possibility of there being more than one point of the graph on the straight line is not excluded, but in such a case all the points with intermediary abscissae must also be on the same straight line.

[8] Or the upper right and upper left probability densities wherever the probability density does not exist.

$F^*(y)$, $0 \leqslant y < \infty$, the *minimal concave* function $\geqslant F(y)$. $F^*(y)$ is the function defined for $y \geqslant 0$ through

$$(5.1)\quad F^*(y) = \sup_{0 \leqslant y' \leqslant y < y''} \left[\frac{y'' - y}{y'' - y'} F(y') + \frac{y - y'}{y'' - y'} F(y'')\right].$$

It is obvious from (5.1) that $F^*(y)$ is concave and monotonically nondecreasing. For all $y \geqslant 0$ we have $F^*(y) \geqslant F(y)$ and there are arbitrarily large y for which $F^*(y) = F(y)$. Furthermore, if for some value of y we have $F^*(y) > F(y)$, then $F^*(y)$ is linear in some neighborhood of y.

A *line of support* of a concave function $G(y)$, $0 \leqslant y < \infty$, at y_1 is a straight line $z = l(y)$ satisfying $l(y) \geqslant G(y)$ for $y \geqslant 0$ with equality holding for $y = y_1$.

We can now state the following result:

THEOREM 1. *F belongs to the class $C(—, A_0, K_0)$ if and only if, for every y_1, $(0 \leqslant y_1 < \infty)$ and every line of support $l(y)$ of $F^*(y)$ at y_1, the set of points $T^- = \{y \mid F(y) < l(y) - K_0/A_0, 0 \leqslant y < y_1\}$ is to the left of the set $T^+ = \{y \mid F(y) > l(y) - K_0/A_0, 0 \leqslant y < y_1\}$.*

PROOF. *Necessity.* Assume the condition is violated for y_1. Let $l(y)$ be a line of support of $F^*(y)$ at y_1. By the remarks following (5.1) there exists a number y_0, $y_0 \geqslant y_1$, such that $F(y_0) = F^*(y_0) = l(y_0)$. (If $F(y_1) = F^*(y_1)$ we take $y_0 = y_1$, otherwise $F^*(y) = l(y)$ for y sufficiently near to y_1 and there exists a largest value y for which the last equality holds; we take y_0 as this largest y.) If we define c_0 by $c_0/A_0 =$ slope of $l(y)$, then the above y_0 has the properties which define y_0 in Section 3 (see discussion preceding Condition (α) in Section 3).

If the condition of the theorem is violated for y_1 and $l(y)$, it is of course also violated for y_0 and $l(y)$. But this then means that Condition (β) of Section 3 is not satisfied for $c = c_0$ defined above, and thus no (s, S) policy is optimal for this choice of c.

Sufficiency. The proof of necessity shows that the condition of Theorem 1 is equivalent to the statement that Condition (β) holds for the given values A_0 and K_0 and every choice of c_0. To complete the proof it remains to show that Condition α is also satisfied.

First we prove that the condition of Theorem 1 implies

$$(5.2)\qquad F(x) \geqslant F^*(x) - K_0/A_0 \qquad (x \geqslant 0).$$

Indeed, let (5.2) be false for $x = x_1$. Then $F^*(x)$ is linear in the neighborhood of x_1. Let (y', y_1) be the largest segment containing x_1 in which $F^*(x)$ is linear, and denote by $l(y)$ the straight line which coincides with $F^*(x)$ on this segment. Then $l(y)$ is a line of support of $F^*(x)$ at $x = y_1$, and in the notation of Theorem 1 we have $y' \in T^+$, $x_1 \in T^-$.

Since $0 \leqslant y' < x_1 < y_1$ this contradicts the condition of the theorem and (5.2) is thus established.

To complete the proof it remains to show that (5.2) implies Condition (α) of Section 3 for every choice of c_0. As remarked above, the straight line of slope c_0/A_0 through $[y_0, F(y_0)]$ is a line of support of $F^*(y)$ at y_0. Let $x_1 > y_0$ and denote by $l_1(y)$ the line of support of $F^*(y)$ at $y = x_1$. Obviously $F(x) \leqslant l_1(x)$ for all x, hence by (5.2) $F(x) \leqslant F(x_1) + K_0/A_0 + (c_1/A_0)(x - x_1)$ where c_1/A_0 is the slope of $l_1(x)$. But F^* is concave and $x_1 > y_0$, hence $c_1 \leqslant c_0$ and therefore we have for $x > x_1$

$$F(x) \leqslant F(x_1) + \frac{K_0}{A_0} + \frac{c_0}{A_0}(x - x_1).$$

This is precisely Condition (α).

Remark: It is easy to see that (5.2) is equivalent to Condition (α), but this is not required for our proof. It is very easy to verify whether (5.2) holds; since (5.2) is a necessary condition for $F \in C(—, A_0, K_0)$, it may be worthwhile to check whether it is fulfilled before trying to see if the condition of Theorem 1 is satisfied.

6. PROBLEM (d)

In this section we characterize the class $C(c_0, -K_0)$.

THEOREM 2. *F belongs to the class $C(c_0, - K_0)$ if and only if for every $y_1 (0 \leqslant y_1 < \infty)$ and every line of support $l(y)$ of $F^*(y)$ at y_1, the set of points $\{y \mid F(y) < l(y - K_0/c_0), 0 \leqslant y < y_1\}$ is to the left of the set $\{y \mid F(y) > l(y - K_0/c_0), 0 \leqslant y < y_1\}$.*

PROOF. *Necessity.* Given any $A = A_0 > 0$ let $l(y)$ be a line of support of $F^*(y)$ with slope c_0/A_0 and let y_1, $0 \leqslant y_1 < \infty$, be such that $F^*(y_1) = F(y_1) = l(y_1)$. Then

$$l\left(y - \frac{K_0}{c_0}\right) = F(y_1) + \left(y - \frac{K_0}{c_0} - y_1\right)\frac{c_0}{A_0} = F(y_1) + \frac{c_0}{A_0}(y - y_1) - \frac{K_0}{A_0}.$$

Thus the condition of Theorem 2 is equivalent to Condition (β) holding for all A_0 and is therefore necessary.

Sufficiency. We need only to show that Condition (α) holds for all A_0.

It is easily seen from the derivation of (5.2) in the preceding section, that the condition of the theorem implies

$$F(y) \geqslant F^*\left(y - \frac{K_0}{c_0}\right) \quad \text{for} \quad y \geqslant \frac{K_0}{c_0}. \tag{6.1}$$

From (6.1), the monotonicity and concavity of F^*, and the fact that $F(y_0) = F^*(y_0)$, we have, for $x_1 \geqslant y_0$,

$$F^*(x_1) - F(x_1) \leqslant F^*(x_1) - \max\,[F^*(y_0), F^*(x_1 - K_0/c_0)]$$

$$= F^*(x_1) - F^*[\max (y_0, x_1 - K_0/c_0)]$$

$$\leqslant F^*(y_0 + K_0/c_0) - F^*(y_0) \leqslant (K_0/c_0)\cdot(c_0/A_0) = K_0/A_0 .$$

Hence, again using the concavity of F^*, we have, for $x \geqslant x_1 \geqslant y_0$,

$$F(x) - F(x_1) \leqslant F^*(x) - F^*(x_1) + F^*(x_1) - F(x_1)$$

$$\leqslant (c_0/A_0)(x - x_1) + K_0/A_0 .$$

Thus Condition (α) is satisfied and the proof is complete.

7. PROBLEM (e)

From either of Theorems 1 and 2 we deduce immediately the following

THEOREM 3. *F belongs to the class C if and only if it is concave for $y \geqslant 0$.*

Thus $F \in C$ if, and only if, left and right probability densities exist for $y > 0$ and each of these is monotonically nonincreasing (the two densities are, of course, equal except at an at most countable set of values of y).

The condition of Theorem 3 is never satisfied when D is a discrete random variable, e.g., in the important case when D is always an integral multiple of some unit. In this case it is reasonable, however, also to restrict the stock and the amounts ordered and not to allow them to assume arbitrary positive values. For a further discussion we refer to Section 9 below.

8. GEOMETRIC FORMULATIONS

A straight line $l(y)$ *crosses* $F(y)$ *at most once to* the left of $y_1(>0)$[9] if the interval $(0, y_1)$ can be divided into two intervals[10] such that $F(y) \leqslant l(y)$ in one of them and $F(y) \geqslant l(y)$ in the other. Using this terminology it is very easy to give a graphical interpretation of Theorems 1 and 2.

Theorem 1 becomes:

$F \in C(—, A_0, K_0)$ if and only if every support line of $F^*(y) - K_0/A_0$ crosses $F(y)$ at most once to the left of the point of support.

Similarly Theorem 2 may be reformulated:

$F \in C(c_0, —, K_0)$ if and only if every support line of $F^*(y - K_0/c_0)$ crosses $F(y)$ at most once to the left of the point of support.

The reformulation of Theorem 2 needs some justification. Let $l(y)$ be a line of support of $F^*(y - K_0/c_0)$ at y_1. Then the condition of Theorem 2 is equivalent to the statement that $l(y)$ crosses $F(y)$ at most once to the left of $y_1 - K_0/c_0$. We must prove that no trouble can be caused

[9] Or to the left of the point $[y_1, l(y_1)]$.

[10] One of the two may be degenerate.

by adjoining the interval $(y_1 - K_0/c_0, y_1)$. It is obviously sufficient to show this for y_1 with $F^*(y - K_0/c_0) < l(y)$ for $y > y_1$. But then $F^*(y)$ is not linear throughout any neighborhood of $y_1 - K_0/c_0$. Hence $F(y_1 - K_0/c_0) = F^*(y_1 - K_0/c_0) = l(y_1)$, and therefore $F(y) \geqslant l(y)$ for $y_1 - K_0/c_0 \leqslant y \leqslant y_1$ as required.

It is very easy to check the above conditions graphically by means of a simple linkage mechanism such as, for example, is shown in the following figure:

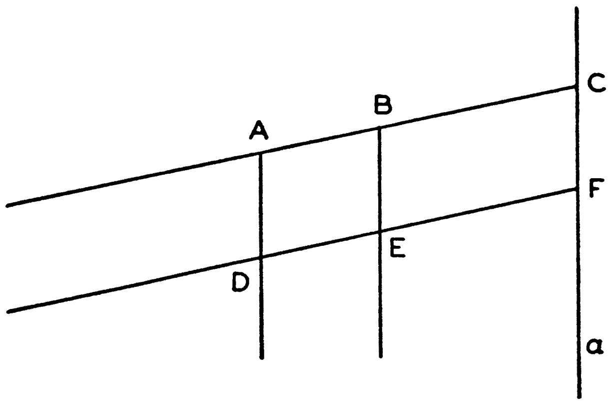

FIGURE 3

Here $\overline{AB} = \overline{DE}$, $\overline{BC} = \overline{EF}$; the lines ABC and DEF pivot around A, B, D, E, and C, F slide freely on the bar a. To check the condition of Theorem 1, for example, one adjusts D and E so that $\overline{AD} = \overline{BE} = K_0/A_0$ and sets the line a parallel to the ordinate axis. Then one moves the line ABC through the positions where it is a line of support of $F^*(y)$ and observes where the parallel line DEF cuts $F(y)$. (It is not necessary for this purpose to draw $F^*(y)$.)

9. THE DISCRETE CASE

The method of Sections 3–6 can be applied to characterize many other classes of demand functions for which the (s, S) assumption holds. Usually, however, the conditions obtained are clumsy and not easily verifiable.

There is, however, one remark that should be made. Let us restrict ourselves to the most important case when the commodity considered comes in integral multiples of some unit. We may then assume that the demand D can take only integral values, and similarly the stock at hand and the amount ordered are integers. In this case all our conditions must be satisfied only for integral values of the arguments. Thus the condition of Section 4 becomes: $F \in C(c_0, A_0, —)$ *if and only* *if the sequence* $g(n)$, $n = 0, 1, 2, \cdots$, *is decreasing up to some* n_0 *and nondecreasing for* $n \geqslant n_0$.

Similarly we must worry about Conditions (α) and (β) only for integral values of the arguments. Thus in Theorems 1–3 it is necessary to consider $F(y)$ only for integral values of y. An equivalent method of doing this is the following:

Theorems 1–3 *become valid for integral valued commodities if* $F(y)$ *is replaced by*[11]

$$F^0(y) = ([y] + 1 - y)F([y]) + (y - [y])F([y] + 1),$$

where $[y]$ *denotes the largest integer* $\leqslant y$.

Theorem 3 perhaps deserves a special formulation in this case:

THEOREM 4. $F \epsilon C$ *for integral valued commodities if and only if the sequence* $\beta_n = \text{Prob}\ (D = n)$ *is nonincreasing for* $n \geqslant 1$.

Hebrew University
Cornell University
Cornell University and University of California at Los Angeles

REFERENCES

[1] ARROW, K. J., T. HARRIS, AND J. MARSCHAK, "Optimal Inventory Policy," ECONOMETRICA, Vol. 19, July, 1951, pp. 250–272.

[2] DVORETZKY, A., J. KIEFER, AND J. WOLFOWITZ, "The Inventory Problem: I, Case of Known Distributions of Demand," ECONOMETRICA, Vol. 20, April, 1952, pp. 187–222.

[11] This amounts to redefining $F(y)$ for nonintegral y by linear interpolation between the two integers nearest y.