A Note on Assemble-to-Order Systems with Batch Ordering

Jing-Sheng Song
Graduate School of Management, University of California, Irvine, California 92697
jssong@uci.edu

We study an assemble-to-order inventory system. The stocks are held for components, with final products assembled only when customer orders are realized. Customer orders form a multivariate compound Poisson process, component replenishment leadtimes are constant, and demands in excess of inventory on hand are backlogged. The component inventories are controlled by \((R, nQ)\) policies. We show that under certain general conditions the inventory position vector has a uniform equilibrium distribution. This result generalizes the corresponding single-item theory considerably. It allows us to express the key performance measures of the system, such as order fill rates and average order-based backorders, as the averages of their counterparts in the base-stock systems.

(Assemble-to-Order Systems; Performance Evaluation; Multivariate Compound Poisson Process; \((R, nQ)\) Policy; Markov Chain; Uniform Distribution)

1. Introduction and the Model

Inspired by industrial practices, there have been several recent research developments in analyzing the assemble-to-order (ATO) systems; see Song (1997, 1998), Song et al. (1999), and the references therein. In an ATO system, components are made or purchased to stock, but final products are assembled only when customer orders are realized. This approach makes it possible to realize great product variety and quick introduction of new products, while the component inventories enable fast and reliable fulfillment of customer orders. The stocking problems faced in mail-order systems and service-part inventories have the same structure. In fact, any multi-item inventory system in which customer orders may consist of several items in different amounts can be viewed as an ATO system. Because of the analytical complexities, however, the work in this area has primarily focused on base-stock systems. That is, the component inventories are replenished according to base-stock (order-up-to) policies. The purpose of this note is to show that some of the results developed previously can be easily extended to ATO systems that employ more general inventory control policies, the \((R, nQ)\) policies.

We assume that the component inventories are reviewed continuously. For each item (component) \(i\), there is a base lot-size \(Q_i\) (a positive integer), such as a truck load or a full container of parts. Whenever the inventory position (net inventory plus inventory on order) of item \(i\) falls to or below the reorder point \(R_i\), an order of size \(nQ_i\) is placed, where \(n\) is the smallest integer so that the inventory position after ordering is above \(R_i\). When \(Q_i = 1\), the policy reduces to a base-stock policy.

The \((R, nQ)\) policies are widely used in industry. For single-item inventory systems, this type of policy is optimal when the order quantity is restricted to be integral multiples of the base lot-size; see Veinott (1965) for periodic-review systems and Chen (1996) for continuous-review systems. Also, it is nearly optimal for single-item systems with a fixed plus variable order cost. (In this case, the true optimal policy is of \((s, S)\) type. When there is no fixed order cost, a base-stock policy is optimal.) See Zheng and Chen...
Thus, the type of policy considered here is appropriate for systems in which there is a fixed base lot-size or there are economies of scale in component replenishments.

Except for the generalized inventory policies, the other features of our model are identical to those in Song (1997, 1998). In particular, the demand for the final products forms a Poisson process with overall demand rate \( \lambda \). Each demand (or product, or customer order) is specified by a fixed kit of items required, and the amount requested for each item within the kit is a positive random variable. For any subset of items \( K \) of \( \{1, \ldots, J\} \), we say a demand is of type \( K \) if it requires \( Z_i^K \) units of item \( i \) in \( K \) and 0 units outside \( K \).

(Throughout this note, we use subscripts to indicate item (or component) types and superscripts order (or product) types.) Here, \( Z_i^K \) is a positive integer random variable. We do not impose any restrictions on the joint distribution of \( Z^K = (Z_i^K, i \in K) \). In other words, \( Z_i^K \) may be arbitrarily dependent for all \( i \in K \). We assume that there is a fixed probability \( q^K \) that an order is of type \( K \). Thus, type-\( K \) demands form a Poisson stream with rate \( \lambda^K = q^K \lambda \). Also, each demand’s type is independent of the other demands’ types and of all other events. Let

\[
\mathcal{K} = \text{set of all demand types} = \{K: q^K > 0\}.
\]

Notice, although \( \mathcal{K} \) may contain all the nonempty subsets of \( \{1, \ldots, J\} \), this rarely happens in any practical instance with large \( J \).

The replenishment leadtime for component \( i \) is a constant \( L_i \), \( i = 1, \ldots, J \). The final assembly time of any product is negligible relative to the component leadtimes. Demands that cannot be filled immediately are backlogged. When an order arrives and we have some of its items in stock but not all, we can either ship the in-stock items or put them aside as committed inventory. However, a customer request is considered backlogged unless it can be satisfied completely. Demands are filled on a First-Come-First-Served (FCFS) basis. When there are backorders, they are also filled on a FCFS basis. Notice that the demand process for any item \( i \) again forms a compound Poisson process whose rate \( \lambda_i \) is the sum of that of the individual demand processes \( \lambda_i = \sum_{K \in \mathcal{K}} \lambda^K \), and the batch size is a mixture of the individual batch sizes.

We are concerned with the performance evaluation of inventory policies in this system. The performance measures of primary interest are

\[
F^K = \text{type-} K \text{ order fill rate} = \text{probability that all items in a type-} K \text{ order are filled immediately};
\]

\[
\bar{B}^K = \text{average number of backordered type-} K \text{ demands};
\]

for any demand type \( K \).

Let \( t \geq 0 \) be the continuous time variable, and for each \( t \) denote

\[
IN_i(t) = \text{net inventory of item } i \text{ at time } t;
\]

\[
IP_i(t) = \text{inventory position of item } i \text{ at time } t;
\]

\[
D_i(t) = \text{cumulative demand for item } i \text{ by time } t.
\]

Let \( D_i(t) \) stand for the steady-state limit of \( D_i(t) = D_i(t + L_i) − D_i(t) \), the leadtime demand of item \( i \). Then, \( D_i \) has the same distribution as \( D_i(L) \), a compound Poisson distribution.

To conduct the performance evaluation, it is essential to obtain the joint distribution of the steady-state net inventories \( (IN_1, \ldots, IN_J) \), which is the steady-state limit of \( (IN_1(t), \ldots, IN_J(t)) \). However, it is well known that

\[
IN_i(t + L_i) = IP_i(t) − D_i(t, t + L_i), \quad 1 \leq i \leq J.
\]

If we can show that \( \{IP(t) = (IP_1(t), \ldots, IP_J(t)), \ t \geq 0\} \) has a steady-state limit \( IP = (IP_1, \ldots, IP_J) \), then

\[
IN_i = IP_i − D_i, \quad 1 \leq i \leq J. \tag{1}
\]

Thus, the question now becomes whether there exists \( IP \) and if so how to determine its distribution.

### 2. Main Results

Due to the nature of the demand process, in which a demand typically requests a kit of items, the inventory positions of different items may change simultaneously at a demand epoch. This may cause one to suspect that the determination of the joint limiting distribution of inventory positions will be difficult. We show below, however, that under certain general conditions the item inventory positions have a joint uniform distribution in steady state (Theorem 1).
The Equilibrium Distribution

Observe that the future evolution of $\text{IP}(t)$ depends only on its current state and the future demand. In addition, the demand is independent of the inventory status. Therefore, the process $\{\text{IP}(t), \ t \geq 0\}$ is a Markov chain. Its state space is

$$S = S_1 \times S_2 \times \cdots \times S_p,$$

where $S_i$ is the integer set $[R_i + 1, \ldots, R_i + Q_i]$. Denote the cardinality of $S$ by $M$, then $M = \prod_{i=1}^{p} Q_i$. Throughout this note we make the following assumption and we will present sufficient conditions for it later.

**Assumption 1.** The Markov chain $\{\text{IP}(t), \ t \geq 0\}$ is irreducible and aperiodic.

It is well known that, under this assumption, $\{\text{IP}(t), \ t \geq 0\}$ has a unique stationary distribution that is also its limiting distribution (see Feller 1971, for example).

**Theorem 1.** Under Assumption 1, $\{\text{IP}(t), \ t \geq 0\}$ has a unique limit $\text{IP}$ that is uniformly distributed on $S$. Thus, (1) holds, and $\text{IP}$ is independent of $\text{D}$.

**Proof.** Assume $\text{IP}(0)$ is uniformly distributed on $S$. It suffices to show that $\text{IP}(t)$ is also uniformly distributed on $S$ for any $t$. For convenience, we shall show the same property for a transformed process $\{\text{IP}^\ast(t), \ t \geq 0\}$, where $\text{IP}^\ast(t) = R_i + Q_i - \text{IP}(t)$ for all $i$. The state space of the new process becomes $S^\ast = S_1^\ast \times S_2^\ast \times \cdots \times S_p^\ast$, where $S_i^\ast = \{0, \ldots, Q_i - 1\}$ for all $i$. Its cardinality is still $M = \prod_{i=1}^{p} Q_i$. Clearly, $\text{IP}(t)$ is uniformly distributed on $S$ if and only if $\text{IP}^\ast(t)$ is uniformly distributed on $S^\ast$. Now, by assumption, $\text{IP}^\ast(0)$ is uniformly distributed on $S^\ast$. Observe that

$$\text{IP}^\ast(t) = [\text{IP}^\ast(0) + \text{D}(t)] \mod Q_i$$

for all $i = 1, \ldots, J$ and all $t \geq 0$. Condition on $\text{D}(t) = [\text{D}(t)]$. Because $\text{IP}(0)$ is uniform on $S$, so is $[\text{IP}(t)]D(t)$, for any fixed $\text{D}(t)$. Now decondition on $\text{D}(t)$. The unconditional distribution of $\text{IP}^\ast(t)$ is also uniform on $S^\ast$. (The unconditional distribution is just a weighted average of the conditional distributions. However, those conditional distributions are all the same, i.e., uniform.)

For single-item $(R, nQ)$ systems (i.e., $J = 1$), this result was established by Simon (1968). (See also Hadley and Whitin (1961) for periodic-review models, and Sivazlian (1974) and Richards (1975) for continuous-review $(s, S)$ models.) Our contribution here is to show that this well-known single-item theory can be generalized considerably to the multi-item setting. Moreover, the multidimensional result appears to be even more remarkable: Although the item demands may be correlated, the item inventory positions are independent in steady state. Our proof generalizes the approach used in Zipkin (2000, Problem 6.5) for the single-item system. See also Feller 1971, p. 64). The fact that this generalization is possible even when correlation presents is largely due to the right nature of the uniform distribution and the mod function.

A similar finding was obtained by Caplin (1985) in a different context. He studied the macroeconomic effect of $(s, S)$ policies in a multiretailer setting. Each retailer followed an $(s, S)$ policy, and the demands for different retailers may be correlated. He used a discrete-time model, but assumed that replenishment decisions can be made continuously within each period. His model implicitly assumes unit demand at any demand point. In this setting, under some regularity conditions, Caplin showed that the inventory positions of the retailers are uniformly distributed. His approach is different from ours, however.

**Sufficient Conditions**

Assumption 1 requires that all the states in $S$ (or equivalently $S^\ast$) communicate in an aperiodic fashion. We now identify certain sufficient conditions for this. The most obvious one is the following:

(a) For any item $i$, there is a demand type that requires only this item and unit demand is possible. That is, $[i] \in \mathcal{K}$ and $P(z_i^{[i]} = 1) > 0$ for all $i$.

Condition (a) holds easily for most distribution systems, including the mail-order systems, because a single-item demand is not uncommon here. In the manufacturing setting, this condition is satisfied if individual components are requested by downstream productions or by part dealers.

It may be that the original system does not possess this property, but an equivalent revised system does. For example, suppose that item $i$ is always demanded in integral multiples of $u_i$, for some positive integer $u_i$.\[\text{SONG}\]
together in the same amounts and \( Q_i = Q_j \). Then, define a new system in which items \( i \) and \( j \) are replaced by a single item. (In the original system, it is optimal to keep equal amounts of items \( i \) and \( j \) in inventory, since extras of one can never be used to fill demands. Suppose \( L_i > L_j \). It is optimal always to order \( i \) first, then wait time \( L_i - L_j \), then order \( j \), so that the orders arrive at the same time. In the new system, therefore, the new item should have the longer leadtime, \( L_j \). See Zhang (1995).) In this way, a pure assembly system, in which all items have the same lot-sizes and are demanded together in the same amounts, can be reduced to a single-item system.

The inventory position vector in a pure assembly system is in general not uniformly distributed on \( S \), because some of the states in \( S \) may be never visited. Each time a demand occurs, all \( IP_i \) decrease by 1 mod\( (Q_i) \). Thus, \( IP \) will demonstrate certain cyclical behavior, but among which states it cycles depends on the initial state. However, once the initial condition is given, the cyclical behavior can be easily identified. Thus, the system can still be analyzed by randomizing the initial condition. Even better, in this case it may be worth negotiating with the suppliers to agree on the same lot-sizes, since we know how to operate such systems optimally, as mentioned in the last paragraph.

For more general ATO systems in which there are more than one demand types that share common components, the following propositions give conditions that are weaker than condition (a) but still guarantee the communication in \( S \) in an aperiodic fashion.

**Proposition 1.** Assumption 1 holds if there exists \( K_o \in \mathcal{K} \) so that the conditions (b)–(d) below are satisfied:

(b) for any \( k \in K_o \), either \( \{k\} \in \mathcal{K} \) or \( K_o \backslash \{k\} \in \mathcal{K} \);

(c) for any \( j \notin K_o \), \( K_o \cup \{j\} \in \mathcal{K} \);

(d) unit demand is possible for all the demand types mentioned above.

Conditions (b) and (c) reduce to condition (a) if \( K_o = \emptyset \). These conditions in effect require some breadth in the size of the demand types. (The size of the demand type \( K \) is \(|K|\), i.e., the number of components requested by type \( K \).) We now provide some examples. First, consider the two-item system \((J = 2)\). Here, there are three possible demand types: demands that require only a single component, i.e., demand type-{1} or type-{2}, and demands that require both items (the assembly case), i.e., demand type-{1, 2}. Except the trivial case in which \( \mathcal{K} \) contains all possible demand types, conditions (b) and (c) are also satisfied if

\[
\mathcal{K} = \{\{1, 2\}, \{1\}\} \quad \text{or} \quad \mathcal{K} = \{\{1, 2\}, \{2\}\}
\]

(In the first case, \{1, 2\} might represent a whole product and \{1\} a spare part.) Thus, any two-item system that has at least two demand types satisfies the conditions.

Now, let us consider the three-item system \((J = 3)\). In this case, there are seven possible demand types: demands that require only a single component, i.e., demand types-{i} for all \( i \), pair demands, i.e., demand type-{i, j} for all \( i < j \), and the demand type-{1, 2, 3}. Examples in which conditions (b) and (c) hold include

\[
\mathcal{K} = \{\{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\} \quad \text{with} \quad K_o = \{1, 2, 3\}
\]

and

\[
\mathcal{K} = \{\{1\}, \{1, 2\}, \{1, 3\}\} \quad \text{with} \quad K_o = \{1\}.
\]

(In the first example, \( \mathcal{K} \) includes all the demand types but the singleton ones. Here, any component is a common component that is shared by two or more products. In the second example, only component 1 is a common component, which is also requested as a spare part.) On the contrary, conditions (b) and (c) do not hold in a system that has only the pair demands, i.e., \( \mathcal{K} = \{\{1, 2\}, \{1, 3\}, \{2, 3\}\} \). In this case, Assumption 1 is not satisfied, since it is possible that some of the states in \( S \) can never be reached. For example, suppose all \( Q_i \) are even. If \( \Sigma \), \( IP_i(0) \) is even, then \( \Sigma \), \( IP_i(t) \) is even for all \( t \). However, if in addition there exists a singleton demand or the demand that requires all the items (like the example in (2)), then the conditions would be satisfied. From these examples we can see that, roughly, the sufficient conditions require a combination of even-
and odd-sized demand types so that the differences of some of these subsets contain only one item.

The following proposition weakens condition (d). Let \( e_j \) be the \( j \)th unit vector.

**Proposition 2.** Assumption 1 holds if there exist \( K_0 \in \mathcal{K} \) and a possible batch size \( z^{K_0} \) (i.e., \( P(Z^{K_0} = z^{K_0}) > 0 \)) such that

1. for any \( k \in K_0 \), either \( z^{K_0} + e_i \) or \( z^{K_0} - e_i \) is possible, and
2. for any \( j \notin K_0, K_0 \cup \{ j \} \in \mathcal{K} \) and the batch size \( z^{K_0} + e_j \) is possible.

Note that condition (c) is a subset of condition (f).

The proofs of the last two propositions are similar to that of Proposition 3.3 in Caplin (1985) because, for any \( x, k \in S_i^j \),

\[
[k + D_i(t)] \mod Q_i = x \Leftrightarrow D_i(t) \mod Q_i = \begin{cases} x - k, & \text{if } x \geq k; \\ Q_i + x - k, & \text{otherwise.} \end{cases}
\]

We omit the details here.

**Performance Evaluation**

Based on the results in Theorem 1, the performance measures of the system can be expressed as the averages of their counterparts in the base-stock systems. (This is achieved by conditioning on \( IP \) in (1).)

For example,

\[
F^K(R, nQ) = \frac{1}{\prod_{j \in K} Q_{e_i}} \sum_{x \in \bigcup_{i \in K} S_i} F^i(s),
\]

\[
\bar{B}^K(R, nQ) = \frac{1}{\prod_{j \in K} Q_{e_i}} \sum_{x \in \bigcup_{i \in K} S_i} \bar{B}^i(s).
\]

Here, \( F^i(s) \) and \( F^K(R, nQ) \) are the type-\( K \) order fill rates under a base-stock policy \( s = (s_1, \ldots, s_j) \) and an \((R, nQ)\) policy with parameters \((R_i, nQ_i)\) for item \( i, i = 1, \ldots, J \), respectively. \( \bar{B}^i(s) \) and \( \bar{B}^K(R, nQ) \) are the counterparts for the expected type-\( K \) backorders. These results simplify the performance evaluation of \((R, nQ)\) systems considerably. They resemble the corresponding single-item results; see, for example, Hadley and Whitin (1961) and Zipkin (2000). The evaluation of the order fill rates and the order-based back orders in the base-stock systems can be found in Song (1997, 1998). Following the same logic, all the qualitative results developed for the base-stock systems (e.g., easy-to-compute performance bounds) are preserved to the \((R, nQ)\) system considered here.

\footnote{The author is grateful to Paul Zipkin for his helpful comments on earlier drafts. She would also like to thank Morton Klein, the Associate Editor, and the referees for their suggestions to improve the exposition of the paper. The research was supported in part by NSF grants DMI-9700037 and DMI-9896339.}

**References**


Simon, R. M. 1968. The uniform distribution of inventory position for continuous review \((s, Q)\) policies. RAND Corporation Report No. P-3938, Santa Monica, CA.


Zhang, A. 1995. Determining jointly optimal inventory policies in a one-stage assembly system is as simple as solving a newsboy problem. Working paper, Department of Information and Operations Management, School of Business Administration, University of Southern California, Los Angeles, CA.


Note: Accepted by Awi Federgruen; received April 2, 1998. This paper was with the author 2 months for 2 revisions.