Inventory Pooling to Deliver Differentiated Service*

Aydın Alptekinoğlu
SMU Cox School of Business

* Joint work with Anand Paul (U.Florida) and Nikhil Jain (Servigistics)

The Model

- A firm supplies a single product to N customers (indexed by i) from a centralized pool of inventory.
- Customer i has a random demand X_i for this product, and requires a minimum type-one service level guarantee: P{X_i is fully satisfied} $\geq \beta_i$
- Events unfold in the following sequence during a single period:
  - The firm orders S units of the product in advance so as to receive them at the beginning of the period.
  - Actual customer demands realize.
  - The firm allocates the available pool of inventory among the N customers and makes shipments accordingly at the end of the period.
- The firm wants to find the minimum S (along with an allocation policy) that satisfies every customer's service level.

The Pie

How large should it be?
How should it be cut?

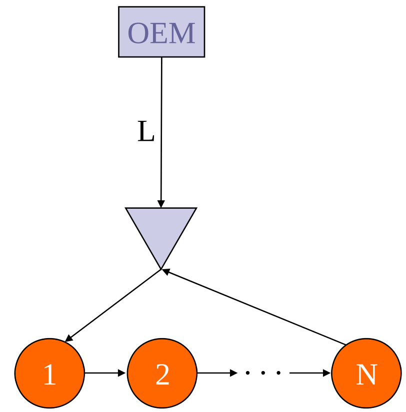
Theoretical Motivation

- Inventory pooling is at the root of many celebrated ideas in OM for 'managing' product variety
  - E.g. postponement (delayed differentiation), component commonality, resource flexibility
- Yet, our understanding of pooling has been largely shaped by cost models (rather than service level models)

Practical Motivation

- Service parts management
  - Gold and blue contracts
- Delayed differentiation for fashion goods
- Stock allocation in perishable goods retailing
  - Inventory management of fresh foods in grocery industry (Swaminathan and Srinivasan 1999)

Practical Motivation:
After-Sales Service in Automobile Industry



- A spare-parts warehouse regularly delivers parts to regional dealers
- "There is a distinct correlation between the quality of after-sales service and customer intent to re-purchase." Therefore, "customer-focused metrics" are essential.
- Decision Variables: System order-up-to level (S), and the allocation rule (x)
- Objective: Finding the optimal ordering and allocation policies so as to satisfy desired service levels

Source: Cohen et al. 2006 (HBR)

The Model

- A firm supplies a single product to N customers (indexed by i) from a centralized pool of inventory.
- Customer i has a random demand X_i for this product, and requires a minimum type-one service level guarantee: P{X_i is fully satisfied} $\geq \beta_i$
- Events unfold in the following sequence during a single period:
  - The firm orders S units of the product in advance so as to receive them at the beginning of the period.
  - Actual customer demands realize.
  - The firm allocates the available pool of inventory among the N customers and makes shipments accordingly at the end of the period.
- The firm wants to find the minimum S (along with an allocation policy) that satisfies every customer's service level.

Outline

- Model
- Motivation
- Classes of Allocation Policies
- Analytical Results
  - List Policies
  - Cyclic Policies
  - Dynamic Policies
- Summary

Classes of Allocation Policies

- Priority Policies
  - Deterministic
  - Randomized
    - Static ← Zhang (2003)
    - Dynamic ← Swaminathan & Srinivasan (1999)

Examples of Priority Policies

1. Service customers in decreasing order of service level
2. Service customers on the basis of a priority list generated randomly (before demands realize)
3. Service customers on the basis of a priority list generated after demand realizations are observed, e.g., serve customers in increasing order of demand realizations

The first policy is almost always suboptimal; the second policy is sometimes optimal; the third policy is potentially optimal.

Example of a Non-Priority Policy

- Two customers, A and B.
- Observe demand realizations.
- A's demand is filled from 80% of the on-hand stock. Then B's demand is filled from residual stock + a fixed reserve of 20% of the stock. Any stock left over is funneled back to A.

Service Level

- Type 1:

  Probability {X_i completely met from stock}

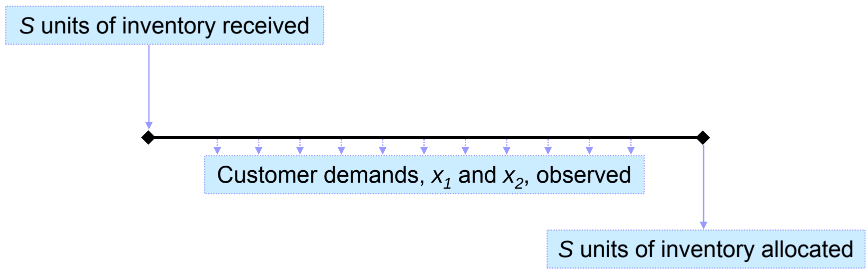
- Type 2 (Expected Fill Rate):

$$E\left[\frac{Customer\ i\ demand\ \ met\ \ from\ stock}{Customer\ i\ demand}\right]$$

Examples

$X_1 \sim U(0,1) \quad \beta_1 = 98\%$

$X_2 \sim U(0,1) \quad \beta_2 = 92\%$



Examples

$S = 1.6$

$X_1 \sim U(0,1) \quad \beta_1 = 98\%$

$X_2 \sim U(0,1) \quad \beta_2 = 92\%$

1. A deterministic policy: Priority list = [1,2] $\Rightarrow$ Service levels = 100%, 92%
2. A randomized-static policy: Priority list = [1,2] with probability 0.75, or [2,1] with probability 0.25 $\Rightarrow$ Service levels = 98%, 94%
3. A randomized-dynamic policy: Priority list = [1,2] if $x_1 \leq x_2$, or [2,1] if $x_2 < x_1$ $\Rightarrow$ Service levels = 96%, 96% (infeasible!)

Classes of Allocation Policies

- Priority Policies
  - Deterministic « *List* »
  - Randomized
    - Static « *Cyclic* »
    - Dynamic « *Linear Knapsack* »

Outline

- Model
- Motivation
- Classes of Allocation Policies
- Analytical Results
  - List Policies
  - Cyclic Policies
  - Dynamic Policies
- Summary

List Policies

- Pre-determined priority list of customers, e.g., Lexus and Toyota customers at a car dealership
- The optimal List Policy: prioritize customers in decreasing order of their desired service levels
- The optimal inventory level: $\max_i \{G_{t_i}^{-1}(\beta_{t_i})\}$ where
  - t_i is the customer with *i*-th highest desired service level and
  - $G_{t_i}(\cdot)$ is the *cdf* of $X_{t_1} + \cdots + X_{t_i}$

Cyclic Policies

THEOREM 4. *Suppose the demands are iid random variables.*

(a) The optimal cyclic policy can be found by solving the following problem: Min S subject to $\mathbf{W} \cdot \mathbf{C}(S) \geq \mathbf{B}$.

(b) The unique solution S_E of the equation $\sum_{i=1}^{N} G_i(S) = \sum_{i=1}^{N} \beta_i$ is a lower bound for the optimal stock.

(c) All the service levels are exactly satisfied if and only if $\mathbf{C}(S_E)$ majorizes $\mathbf{B}$. The optimal inventory in this case is precisely S_E.

Example

- X_1, X_2, X_3 *iid* Normal(100,20)
- Service levels 0.9, 0.8, 0.7 respectively
- $G_1(S)+G_2(S)+G_3(S)=2.4$ when S=292, which is optimal by the preceding theorem
- Optimal cyclic policy: apply priority list
  - 2-1-3 with probability 1/2
  - 2-3-1 with probability 1/6
  - 3-1-2 with probability 1/3
- Optimal list policy: apply priority list 1-2-3 with S=318 units

Linear Knapsack Policy

- Suppose the inventory level is *S* and the demand realizations are $x_1, \ldots, x_N$.
- Linear Knapsack (LK) Policy, defined by two *N*-vectors $(k_1, \ldots, k_N)$ and $(t_1, \ldots, t_N)$, is the following procedure for allocating inventory among the *N* customers:
  - Apply the linear transformation $y_i = k_i x_i + t_i$ to each of the demand realizations ($i = 1, \ldots, N$)
  - Prioritize (or rank-order) customers in increasing order of y_i and allocate *S* accordingly

Linear Knapsack Policy

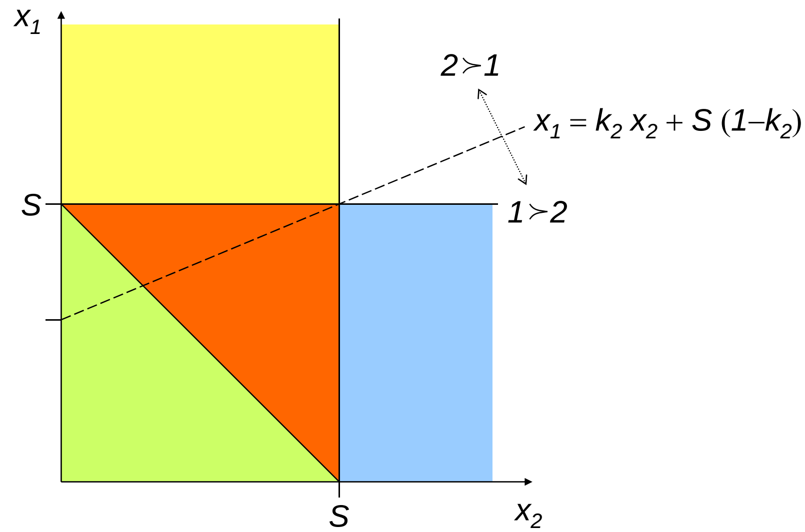
set $k_1 = 1$, $t_1 = 0$, and $t_i = S(1 - k_i)$

$$y_i \equiv k_i x_i + S(1 - k_i)$$

Customer i has priority over customer j if

$$x_i < \left(\frac{k_j}{k_i}\right) x_j + S\left(1 - \frac{k_j}{k_i}\right)$$

Two-Customer Case



Two-Customer Case

Let S_0 be implicitly defined by

$$P\{X_1 + X_2 \le S_0\} = \beta_1 + \beta_2 - 1 + P\{X_1 > S_0,\ X_2 > S_0\}$$

Two-Customer Case

Define

$$\begin{array}{ll} \textit{Case 1}: & \beta_1 \ge \alpha_1 \text{ and } \beta_2 \ge \alpha_2 \\ \textit{Case 2}: & \beta_1 > \alpha_1 \text{ and } \beta_2 < \alpha_2 \\ \textit{Case 3}: & \beta_1 < \alpha_1 \text{ and } \beta_2 > \alpha_2 \end{array}$$

where

$$\alpha_1 \equiv P\{X_1 + X_2 \le S_0\} + P\{X_1 \le S_0,\ X_2 > S_0\}$$

$$\alpha_2 \equiv P\{X_1 + X_2 \le S_0\} + P\{X_1 > S_0,\ X_2 \le S_0\}$$

Two-Customer Case

THEOREM 8. *With two customers, the optimal inventory level is S^*, and the linear knapsack policy with $k_1 = 1$ and $k_2 = k^*$ is an optimal allocation policy. The optimal policy parameters are:*

	S^*	k^*
Case 1: $\beta_1 \geq \alpha_1$ *and* $\beta_2 \geq \alpha_2$	S_0	k_0
Case 2: $\beta_1 > \alpha_1$ *and* $\beta_2 < \alpha_2$	$F_1^{-1}(\beta_1)$	∞
Case 3: $\beta_1 < \alpha_1$ *and* $\beta_2 > \alpha_2$	$F_2^{-1}(\beta_2)$	0

with S_0 and k_0 uniquely determined by two implicit equations:

$$P\{X_1 + X_2 \leq S_0\} = \beta_1 + \beta_2 - 1 + P\{X_1 > S_0,\ X_2 > S_0\} \quad (1)$$

$$P\{\omega(S_0),\ X_1 < k_0 X_2 + S_0(1 - k_0)\} = \beta_1 - P\{X_1 + X_2 \leq S_0\} - P\{X_1 \leq S_0,\ X_2 > S_0\}$$

Example - revisited

$X_1 \sim U(0,1) \quad \beta_1 = 98\%$

$X_2 \sim U(0,1) \quad \beta_2 = 92\%$

The optimal inventory level: S* = 1.55

The optimal allocation policy: LK policy with k_1=1.00 and k_2=0.80; Priority list = [1,2] if $x_1 \leq 0.80\, x_2 + 0.31$, or [2,1] otherwise; Service levels = 98%, 92%

Dynamic Policies: Special Case

- Assume *iid* demands and identical service levels
- Optimal allocation policy is to serve the customers in ascending order of their demand realizations
- Optimal inventory level is the unique S* that satisfies

$$\sum_{i=1}^{N} H_i(S^*) = N\beta$$

where $H_i(\cdot)$ is the *cdf* of the sum of *i* smallest demands

Comparison:

Distribution	β	S_{list}	S_{cyclic}	$S_{dynamic}$
Normal (10,2)	75 %	33	28	28
	80 %	34	30	30
	85 %	34	31	31
	90 %	35	32	32
	95 %	36	34	34
Normal (10,3)	75 %	34	28	27
	80 %	35	30	29
	85 %	36	31	31
	90 %	37	33	33
	95 %	39	36	36
Lognormal (10,5)	75 %	35	28	25
	80 %	37	30	28
	85 %	39	32	31
	90 %	42	35	34
	95 %	47	40	39
Lognormal (10,10)	75 %	38	27	22
	80 %	41	31	25
	85 %	46	35	30
	90 %	52	41	36
	95 %	64	52	46
Lognormal (10,15)	75 %	38	26	19
	80 %	43	31	23
	85 %	49	36	28
	90 %	58	45	35
	95 %	78	62	50

Dynamic Policies: A Bound

- The unique solution S_{LB} of the equation

$$\sum_{i=1}^{N} H_i(S_0) = \sum_{i=1}^{N} \beta_i$$

is a lower bound on the optimal inventory level.

Summary / Contributions

- Identified three classes of allocation policies in varying degrees of ease of implementation
- Characterized the optimal ordering and allocation policies within these classes for any number of customers under several special cases & devised a general algorithmic solution
- Established a general lower bound on optimal inventory
- Developed a closed-form distribution-free solution for the optimal ordering and allocation policies in the case of two customers

Q&A

For more information:

Aydın Alptekinoğlu <aalp@cox.smu.edu>

Anand Paul <paulaa@ufl.edu>

"This ensures that we don't overanalyze."