Leadtime-inventory trade-offs in assemble-to-order systems

by

Paul Glasserman and Yashan Wang

15.764 The Theory of Operations Management

March 02, 2004

Presentation by Nicolas Miegeville

This is a summary presentation based on: Wang, Y., and P. Glasserman. "Lead-time Inventory Trade-offs in Assemble-to-order Systems." ***Operations Research*** **43, no. 6 (1998).**

Objective of the paper

- Usual qualitative statement : inventory is the currency of service.
  - Operations Management books
  - Management reviews
  - Research papers

- May we find a quantitative measure of the marginal cost of a service improvement in units of inventory ?

Methodology and Results

- We focus on a particular class of models : assemble-to-order models with stochastic demands and production intervals
  - items are made to stock to supply variable demands for finished products
    - Multiple FP are ATO from the items (one product may contend several times the same item)
    - For each item : continuous-review base-stock policy : one demand for a unit triggers a replenishment order
  - Items are produced one at time on dedicated facilities
- We measure the service by the fill rate : proportion of orders filled before a target (delivery leadtime).
- We prove that there is a LINEAR trade-off between service and inventory, at high levels of service.

Overview

- Intuitive simplest model
- General Model
- Three theorems for three models
  - Single item model
  - Single product multi item model
  - Multi product multi item model
- Checking the Approximations given by the theorems
- Conclusion

Overview

- Intuitive simplest model
- General Model
- Three theorems for three models
  - Single item model
  - Single product multi item model
  - Multi product multi item model
- Checking the Approximations given by the theorems
- Conclusion

Simplest Model : intuitive result (1/2)

- Single one item-product model
  - Orders arrive in a Poisson stream (rate λ)
  - Time to produce one unit is exponentially distributed (mean $\frac{1}{\mu}$)
  - s denote the base-stock level
  - x denote the delivery time
  - R denote the steady-state (we assume $\lambda < \mu$) response time of an order
- Queuing Theory : M/M/1 results

$$P(R \le x) = 1 - P(R > x) = 1 - \left(\frac{\lambda}{\mu}\right)^{s} e^{-(\mu-\lambda)x}$$

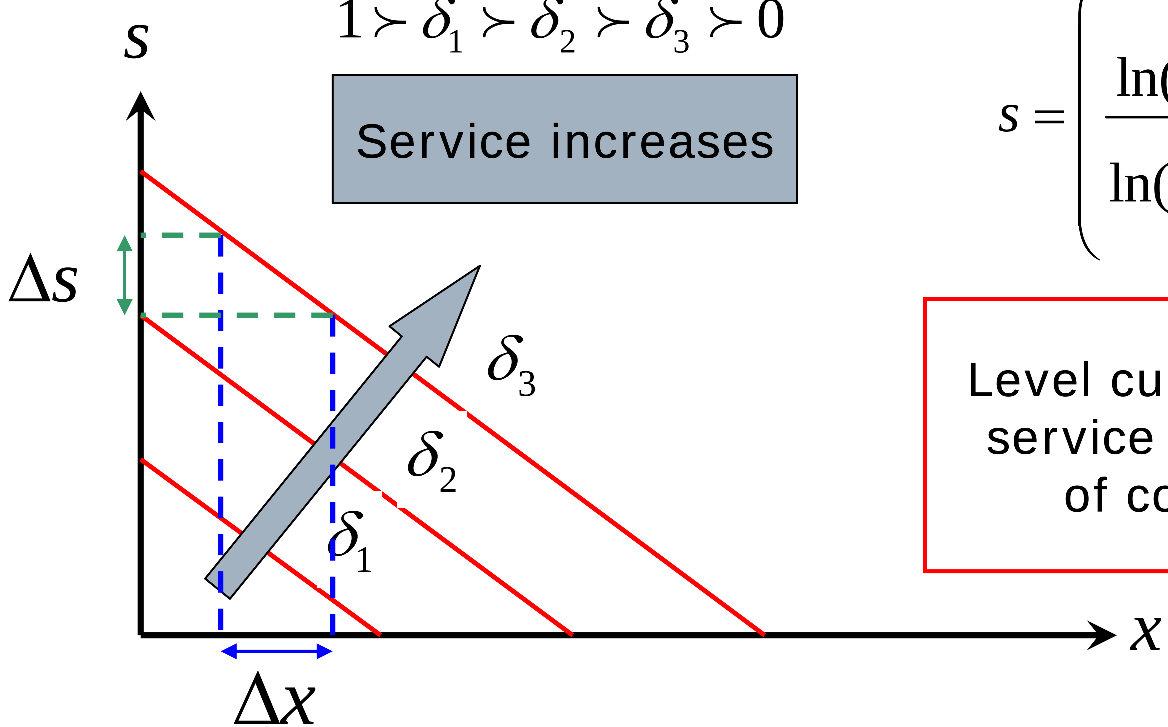
- At a fixed fill rate $(1-\delta) \in [0,1]$:

$$s = \left(\frac{\ln(\delta)}{\ln(\frac{\lambda}{\mu})}\right) - \left(\frac{\mu-\lambda}{\ln(\frac{\mu}{\lambda})}\right) x$$

Simplest Model : intuitive result (2/2)

$$1 \succ \delta_1 \succ \delta_2 \succ \delta_3 \succ 0$$

$$s = \left(\frac{\ln(\delta)}{\ln(\frac{\lambda}{\mu})} \right) - \left(\frac{\mu - \lambda}{\ln(\frac{\mu}{\lambda})} \right) x$$



Level curves of constant service are stright lines of constant slope

For a fixed rate, increase of base-stock decreases the delivery leadtime

Overview

- Intuitive simplest model
- General Model
- Three theorems for three models
  - Single item model
  - Single product multi item model
  - Multi product multi item model
- Checking the Approximations given by the theorems
- Conclusion

General Model : Objective and notations

- Multiple items are produced on dedicated facilities and kept in inventories
- A product is a collection of a possible RANDOM number of items of each type (~components)
- The assembly operation is uncapacitated

- Notations :
  - A is the order interarrival time (products).
  - B is the unit production interval (items)
  - D is the batch order size (items per product)
  - R is the response time
  - s is the base-stock level (items)
  - x is the delivery time

General Model – Assumptions

i=1…d items
j=1…m products

(See Figure 1, page 859 of the Glasserman and Wang paper.)

We assume the production intervals (B), the interarrival times (A) and the batch size vectors (D) are ALL INDEPENDENTS of each other.

Overview

- Intuitive simplest model
- General Model
- Three theorems for three models
  - Single item model
  - Single product multi item model
  - Multi product multi item model
- Checking the Approximations given by the theorems
- Conclusion

General single item Model

- For any r.v. Y, we note the cumulant generating function :

$$\psi_Y(\theta) = \ln(E[e^{\theta Y}]) \qquad \psi_Y^{(1)}(0) = E[Y] \qquad \psi_Y^{(2)}(0) = Var[Y]$$

- Let's introduce the r.v. : $X = \sum_{j=1}^{D} B_j - A$
  - We have :

$$E[X] = E[B].E[D] - E[A]$$

$$Var[X] = E[D].Var[B] + Var[D].(E[B])^2 + Var[A]$$

$$\psi_X(\theta) = \psi_D(\psi_B(\theta)) + \psi_A(-\theta)$$

We require E[X]<0 s.t. the steady-state response time exists

FIRST THEOREM :

If it exists, it is unique
(f Convex and f(0)=0)

- If there is a $\gamma > 0$ at which $\psi_X(\gamma) = 0$, then with $\beta = \psi_B(\gamma)$

$$lim_{s+x\to\infty} e^{\gamma x+\beta s} P(R(s) > x) = C$$

with C constant >0

- For a level curve of constant service, we have approximately :

$$s = -\frac{\gamma}{\beta} x + \frac{1}{\beta} \ln(\frac{C}{\delta})$$

  - Proof of theorem uses the concept of associated queue to the response time (Lemma 1), in which we can show (if the system is stable) by using the Theorem of Gut (1988) a convergence in distribution of the waiting time. Then an exponential twisting gives the result (Wald's likelihood ratio identity).
- With an assembly time Un (random delay iid and bounded), the result is still true

FIRST THEOREM : interpretation

- For a level curve of constant service, one unit increase in s buys a reduction of $\dfrac{\beta}{\gamma}$ in x.

- Particular case, with a tractable constant C :

  (See Proposition 1 on page 860 of the Glasserman and Wang paper)

Single product (multi item) Model

- We consider a system with d items, but with a single product, and thus just one arrival stream A.
- We make a new assumption :
  - Each item i has a base stock level s^i
  - We note $s = \sum_{i=1}^{d} s^i$
  - We assume that for each item $k^i = \dfrac{s^i}{s}$ is constant when s increases.

> The proportion of total inventory held in each item remains constant

- We note for each item i : $X^i = \sum_{j=1}^{D^i} B_j^i - A$

$$\psi_X(\theta) = \psi_D(\psi_B(\theta)) + \psi_A(-\theta)$$

Single product (multi item) Model

- We apply the Theorem 1 to each item i separately :

$$\psi_i(\theta) = \psi_{X^i}(\theta) = \psi_{D^i}(\psi_{B^i}(\theta)) + \psi_A(-\theta)$$

$$\gamma_i > 0 \qquad \psi_i(\gamma_i) = 0 \qquad \alpha_i = k_i \beta_i \qquad \beta_i = \psi_{B^i}(\gamma_i)$$

$$lim_{s+x \to \infty} e^{\gamma_i x + \alpha_i s} P(R^i(s) > x) = C_i > 0$$

- The response time for the full order is the maximum of the response times for the individual items required (→interactions among the multiple items).

Single product (multi item) Model

- We note the set of :
  - leadtime-critical items : $\gamma = \min_i \gamma_i$ and $\mathcal{I}_x = \{i : \gamma_i = \gamma\}$

  → Their individual fill rates increase most slowly as x increases : it constraints the fill rate at long delivery intervals

  - inventory-critical items : $\alpha = \min_i \alpha_i$ and $\mathcal{I}_s = \{i : \alpha_i = \alpha\}$

  → Their individual fill rates increase most slowly as s increases : it constraints the fill rate at high base-stock levels

- These sets of items determine the fill rate when w or s becomes large

SECOND THEOREM :

(See Theorem 2, including equations 6 and 7, on page 861 of the Glasserman and Wang paper)

General Model with Poisson orders

λ_1

$\vdots$

λ_m

(For the remainder of this diagram, see Figure 1, page 859 of the Glasserman and Wang paper.)

i=1...d items
j=1...m products

$$\vartheta_j = \{ i : P(D_j^i > 0) > 0 \}$$

Is the set of items required by product j

$$\mathrm{P}^i = \{ j : P(D_j^i > 0) > 0 \}$$

Is the set of products requiring item i

In this part we require that arrivals of orders for the various products follow independent (compound) Poisson processes

Solution

- For each item i :
  - The demand is the superposition of independent (compound) Poisson processes with :

$$\lambda^i = \sum_{j \in P^i} \lambda_j$$

  - The batch size D^i is distributed as a mixture of $\{D_j^i\}$:
    - With probability $\frac{\lambda_j}{\lambda^i}$, D^i is distributed as D_j^i for $j \in P^i$
- We can apply the Theorem 1 to R^i, the steady state item-i response time !
- As in the second theorem, we find Rj the steady state product-j response time

$$R_j = \max_{i \in \vartheta_j} R^i$$

THIRD THEOREM

(See Theorem 3 on page 861 of the Glasserman and Wang paper)

Overview

- Intuitive simplest model
- General Model
- Three theorems for three models
  - Single item model
  - Single product multi item model
  - Multi product multi item model
- Checking the Approximations given by the theorems
- Conclusion

Single-Item Systems approximation

- Objective : test how the linear approximation works through several examples.
- Method in each example :
  - We calculate α and β according to Theorem 1
  - We study the systems from x=0 and choose some s>0 s.t. fill rate is high
  - We calculate pairs of (x, s) according to the trade-off :

$$\Delta s = -\frac{\gamma}{\beta} \Delta x$$

  - At each pair, the actual fill rate is estimate by MC simulation.
  - We graph the results

Single-Item Systems approximation

- Main results :

(See Figure 2, page 862 of the Glasserman and Wang paper.)

By ignoring the rounding effect, the linear limiting trade off seems to perfectly represent the behavior of the system :
- for higher fill rates (> 90%)
- from a global point of view
- Regardless of the distribution and utilization level

If we have only means and variances of A, B and D, the two-moment approximation give excellent results :

$$\psi_X(\theta) \approx E[X]\theta + \left(\frac{1}{2}\right)\mathrm{var}[X]\theta^2$$

Multiple-Item Systems approximation

- Theorem 2 gives us two limiting regimes, one on x and one on s.
- For each item we compute :

  (see the last paragraph on page 863 of the Glasserman and Wang paper)

- And we use the result of theorem 1 :

  (see equation 13 on page 863 of the Glasserman and Wang paper)

Multiple-Item Systems approximation

- Main results :

(See Figure 15, page 865 of the Glasserman and Wang paper.)

By ignoring the rounding effect, the linear limiting trade off works well :
- for higher fill rates (> 90%)
- Regardless of the distribution and utilization level

Overview

- Intuitive simplest model
- General Model
- Three theorems for three models
  - Single item model
  - Single product multi item model
  - Multi product multi item model
- Checking the Approximations given by the theorems
- Conclusion

Conclusion

- Quantify the trade-off between :
  - Longer leadtimes
  - Higher inventory levels
- In achieving a target fill rate
- Is possible both theoretically and numerically
- In a general class of production-inventory models

- The simple results on single-item systems give a way of analyzing the most constraining items in multiple-item systems.