A queueing approach to a multi class $M/G/1$ make-to-stock with backlog

Yoav Kerner

U. Toronto & Technion

Joint with Opher Baron

YEQT III, Eindhoven, 19.11.09

Objectives

1. The equivalence of inventory model and queueing problem
2. Improving existing policies

$M/G/1$ make to stock

Single class

- A single machine produces a single item
- Demand process according to a $PP(\lambda)$
- i.i.d production times (G)

$M/G/1$ make to stock

Single class

- A single machine produces a single item
- Demand process according to a $PP(\lambda)$
- i.i.d production times (G)
- Linear holding (h) and backlog (b) costs

$M/G/1$ make to stock

Single class

- A single machine produces a single item
- Demand process according to a $PP(\lambda)$
- i.i.d production times (G)
- Linear holding (h) and backlog (b) costs
- Decision: To produce or not. Taken at
  - production completion

$M/G/1$ make to stock

Single class

- A single machine produces a single item
- Demand process according to a $PP(\lambda)$
- i.i.d production times (G)
- Linear holding (h) and backlog (b) costs
- Decision: To produce or not. Taken at
  - production completion
  - arrival at idelness

$M/G/1$ make to stock

Stationary policy:
Produce $\Leftrightarrow$ the stock level is smaller than a *base stock level* S

$M/G/1$ make to stock

Stationary policy:
Produce $\Leftrightarrow$ the stock level is smaller than a *base stock level* S

Average long-run cost:

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T (hI_S(t) + bB_S(t))dt$$

$M/G/1$ make to stock

Stationary policy:
Produce $\Leftrightarrow$ the stock level is smaller than a *base stock level* S

Average long-run cost:

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T (hI_S(t) + bB_S(t))dt = h\mathrm{E}_S(I) + b\mathrm{E}_S(B)$$

$M/G/1$ make to stock

Stationary policy:
Produce $\Leftrightarrow$ the stock level is smaller than a *base stock level* S

Average long-run cost:

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T (hI_S(t) + bB_S(t))dt = h\mathrm{E}_S(I) + b\mathrm{E}_S(B)$$

Optimal policy is stationary (MDP)

$M/G/1$ make to stock

Stationary policy:
Produce $\Leftrightarrow$ the stock level is smaller than a *base stock level* S

Average long-run cost:

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T (hI_S(t) + bB_S(t))dt = h\mathrm{E}_S(I) + b\mathrm{E}_S(B)$$

Optimal policy is stationary (MDP)

What is the optimal S?

Equivalent queueing problem

An $M/G/1$ queue (λ, G)

Equivalent queueing problem

An $M/G/1$ queue (λ, G)

Optimization problem:

$$\min_{S \geq 0} h\mathrm{E}(QI_{\{Q<S\}}) + b\mathrm{E}(QI_{\{Q>S\}})$$

Equivalent queueing problem

An $M/G/1$ queue (λ, G)

Optimization problem:

$$\min_{S \geq 0} h\mathrm{E}(QI_{\{Q<S\}}) + b\mathrm{E}(QI_{\{Q>S\}})$$

The optimal solution S^* (Veatch and Wein 96):

$$\mathrm{P}(Q \leq S^*) = \frac{b}{b+h}$$

Equivalent queueing problem

An $M/G/1$ queue (λ, G)

Optimization problem:

$$\min_{S \geq 0} h\mathrm{E}(QI_{\{Q<S\}}) + b\mathrm{E}(QI_{\{Q>S\}})$$

The optimal solution S^* (Veatch and Wein 96):

$$\mathrm{P}(Q \leq S^*) = \frac{b}{b+h}$$

In this queue customers are served before their arrival!!

Multi class problem

- N classes of customers
- arrival rates λ_i
- backlog costs b_i ($b_i > b_{i+1}$)

Make to stock policies

- FCFS

Make to stock policies

- FCFS
- Strict Pririty (SP)

Make to stock policies

- FCFS
- Strict Pririty (SP)
  - Stock $\longrightarrow$ Give to any arrival when there is a

Make to stock policies

- FCFS
- Strict Pririty (SP)
  - Stock $\longrightarrow$ Give to any arrival when there is a
  - Shortage $\longrightarrow$ Give to the most expensive "$C\mu$ rule"

Make to stock policies

- FCFS
- Strict Pririty (SP)
  - Stock $\longrightarrow$ Give to any arrival when there is a
  - Shortage $\longrightarrow$ Give to the most expensive "$C\mu$ rule"
- Inventory Rationing (IR)

Make to stock policies

- FCFS
- Strict Pririty (SP)
  - Stock $\longrightarrow$ Give to any arrival when there is a
  - Shortage $\longrightarrow$ Give to the most expensive "$C\mu$ rule"
- Inventory Rationing (IR)
  - Constants $S = R_{n+1} \geq R_n \ldots \geq R_1 \geq 0$

Make to stock policies

- FCFS
- Strict Pririty (SP)
  - Stock $\longrightarrow$ Give to any arrival when there is a
  - Shortage $\longrightarrow$ Give to the most expensive "$C\mu$ rule"
- Inventory Rationing (IR)
  - Constants $S = R_{n+1} \geq R_n \ldots \geq R_1 \geq 0$
  - allocate to a class i customer $\Leftrightarrow$ stock level $> R_i$.

Make to stock policies

- FCFS
- Strict Pririty (SP)
  - Stock $\longrightarrow$ Give to any arrival when there is a
  - Shortage $\longrightarrow$ Give to the most expensive "$C\mu$ rule"
- Inventory Rationing (IR)
  - Constants $S = R_{n+1} \geq R_n \ldots \geq R_1 \geq 0$
  - allocate to a class i customer $\Leftrightarrow$ stock level $> R_i$.

$$SP \subseteq IR \qquad (R_i = 0)$$

Some specific papers

- Ha ('97) Rationing $M/M/1$
- De Vericourt, Karasmen and Fallarey ('02) IR, FCFS, SP $M/M/1$
- J. P. Gayon, F. de Véricourt and F. Karaesmen ('07) IR $M/E_k/1$.
- Benjaafer, Elhafsi and Kim (05') FCFS $M/G/1$
- Benjaafer, Elhafsi and Kim (07') FCFS, IR $M/M/1$
- Abouee-Mehrizi, Balcioglu and Baron ('09) IR $M/G/1$

Markovian systems $\longrightarrow$ Dynamic programming

Does not work for $M/G/1$

Inventory Rationing

Let cheaper customers wait to avoid backlog costs of more expensive customers.

Inventory Rationing

Let cheaper customers wait to avoid backlog costs of more expensive customers.

Is IR an optimal control policy?

Inventory Rationing

Let cheaper customers wait to avoid backlog costs of more expensive customers.

Is IR an optimal control policy?

Optimal in $M/M/1$ (De Vericourt et al. '02)

Infer from low priority

The problem: IR ignores the number of low priority customers.

Infer from low priority

The problem: IR ignores the number of low priority customers.

Can we improve by a function of $Q_i, \ldots, Q_n$?

Infer from low priority

The problem: IR ignores the number of low priority customers.

Can we improve by a function of $Q_i, \ldots, Q_n$?

The idea: Take the "risk" of no high priority arrivals until production completion and save the backlog cost of class i customer until then.

Infer from low priority

The problem: IR ignores the number of low priority customers.

Can we improve by a function of $Q_i, \ldots, Q_n$?

The idea: Take the "risk" of no high priority arrivals until production completion and save the backlog cost of class i customer until then.

The distribution of the time until production completion depends on $Q_i, \ldots, Q_n$

Infer from low priority

The problem: IR ignores the number of low priority customers.

Can we improve by a function of $Q_i, \ldots, Q_n$?

The idea: Take the "risk" of no high priority arrivals until production completion and save the backlog cost of class i customer until then.

The distribution of the time until production completion depends on $Q_i + \ldots + Q_n$.

Extended Inventory Rationing

Pairs R_i, q_i.

Extended Inventory Rationing

Pairs R_i, q_i.

At arrival epochs, allocate to class i customer

$$\Updownarrow$$

1. Inventory $> R_i$ or
2. Inventory $= R_i$ and $\sum_{j=i}^{n} Q_j \geq q_i\}$

Extended Inventory Rationing

Pairs R_i, q_i.

At arrival epochs, allocate to class i customer

$$\Updownarrow$$

1. Inventory$>R_i$ or
2. Inventory$=R_i$ and $\sum_{j=i}^{n} Q_j \geq q_i\}$

$SP \subseteq IR \subseteq EIR \qquad (q_i = \infty)$

Analysis

Priority $M/G/1$ queue with state-dependent arrival rates.

$Q_1 = B_1 + S - I$

Embed at production completion epochs.

Balance equations....

Minimize

$$hE(Q_1 I_{\{Q_1<S\}}) + b_1 E(Q_1 I_{\{Q_1>S\}}) + \sum_{i=2}^{n} b_i E(Q_i)$$

Optimality

The idea behind EIR: Short remaining production time when $\sum Q_i$ is large.

Optimality

The idea behind EIR: Short remaining production time when $\sum Q_i$ is large.

This is the case if G is with Increasing Failure Rate.

Optimality

The idea behind EIR: Short remaining production time when $\sum Q_i$ is large.

This is the case if G is with Increasing Failure Rate.

IR should be extended according to the G.

Optimality

The idea behind EIR: Short remaining production time when $\sum Q_i$ is large.

This is the case if G is with Increasing Failure Rate.

IR should be extended according to the G.

For example: If G is DFR, the extension is:

Optimality

The idea behind EIR: Short remaining production time when $\sum Q_i$ is large.

This is the case if G is with Increasing Failure Rate.

IR should be extended according to the G.

For example: If G is DFR, the extension is:

If an arriving class i customer finds $R_i + 1$,

allocate iff $\sum_{j=i+1}^{n} Q_j$ is small.

QUESTIONS?