# Optimal Control of Partially Observed Inventory Systems

Alain Bensoussan, Metin Çakanyıldırım and Suresh P. Sethi

School of Management
University of Texas at Dallas

# Partially Observed Inventory

- Transaction errors: Unintentional mistakes happen
  - Eg: Cashiers ask for the name of vegetables at the checkout; Swap the same yogurt twice when you buy two different yogurts.
  - Axsäter (2001): …deployment of …technology is not always economically justifiable and does not eliminate …errors.
  - Raman et al. (2001): inventory records for 65% of sku's at a *publicly traded retailer* are inaccurate.
- Product quality, yield, and spoilage problems not immediately observed
  - Eg: Blood cells, chemicals, batteries.

# Partially Observed Inventory

- Misplaced inventory
  - Eg: Why cannot a bookstore clerk find a book on the shelf?
  - Raman et al. (2001): . . . significant losses – as much as 25% of profits at a leading retailer.
- Pilferage: small but continuous theft
  - Eg: Furnitures, frozen foods.
  - Axsäter (2001): Apart from the loss in value, thefts will also lead to inaccurate inventory records.
- Inventory is counted to eliminate uncertainty in the records; time between two counts in a row can be termed as *information delay*.
- Filtered demand: Unmet demand is not observed.
  - Not finding a product on the shelf, customers often *quitely* leave the store.
  - Lariviere and Porteus (1999), and Ding, Puterman and Bisi (2002).

# Some Literature

- Treharne and Sox (2002): Markovian modulated demand whose state is unobserved.
- Kök and Shang (2004): Inventory level is unobserved for *various* reasons; Inspection policy sought.

## Seemingly Similar Work

- Kaplan (1970): Random delivery lead times.

  The *information delay* shifts *demand observations* in time. The *lead time* shifts *order deliveries* in time.
- Chen (1999): In a multiechelon inventory context, *information lead time* at an echelon is the time it takes for an order placed by the echelon to reach its upstream echelon.
- Neither has partially observed inventories.

# Why a Few Studies?

- Practically: Retailers do not publicize their partial inventory information.
- Theoretically: Information must be inferred from surrogate measures.
  - Current inventory levels inferred from *inventory levels in earlier periods* (Information Delay),
  - Current inventory levels inferred from *stock-outs* (Zeo Balance Walk),
  - Demand inferred from *the demand when it is met* (Filtered Newsvendor).
- Inferences yield conditional probability distributions for the inventory level.
  - Current inventory conditioned on the last observed inventory (Information Delay),
  - Current inventory conditioned on the time of the last stockout (Zero Balance Walk),
  - Demand conditioned on the demand observation given that it is less than the inventory (Filtered Newsvendor),

# Why a Few Studies?

- The state of the system is a conditional distribution evolving in an infinite dimensional space.
- Check if there is a finite dimensional sufficient statistics?
  - Yes, the last inventory observation is a sufficient statistic for information delay case.
  - If yes, check the validity of the basestock and $(s, S)$ policies?
  - No, there is no sufficient statistic for zero balance walk or the filtered newsvendor case.
- In the sequel, we examine the models for
  1. Information delay,
  2. Filtered newsvendor,
  3. Zero balance walk.

# 1. Information Delay

- *Information delays* exist when the most recent inventory information available to the Inventory Manager (IM) is dated.
- Eg: When the sales are not reported as they happen.
- Does not modern technology facilitate the information collection?
- Yes, but it also makes decision making frequent.
- Eg: Dell orders for components about every two hours.
- Still, some decisions are made without full knowledge of the up-to-date events.

# 1. Some Notation

- $I_t, q_t, D_t$: Inventory level, order quantity, demand (in the order of appearance) in period $t$, $t = 1, 2, \ldots, T$.
- $\theta$: Information delay, measured in the number of periods.
- $z_t$: The signal (surrogate variable) observed in period $t$.

$$z_t = \left\{ \begin{array}{ll} I_1 & \text{if } t \leq \theta + 1 \\ I_{t-\theta} & \text{if } t \geq \theta + 2 \end{array} \right\}$$

  $z_t$ is the most recent inventory level observed by period $t$.
- IM knows the signal history $\{z_j : j \leq t\}$, which generates

$$\mathcal{Z}_t := G(\{z_j : j \leq t\})$$

- In classical models, $\mathcal{F}_t := G(\{I_j : j \leq t\})$ is known
- $\mathcal{Z}_t \subseteq \mathcal{F}_t \implies$ Inventory is partially observed.

# 1. Problem Definition

- Inventory is backordered

$$I_{t+1} = I_t + q_t - D_t.$$

- $c(I, q)$ is the one-period cost function.
- Objective function:

$$J(\mathbf{q}) := \mathbb{E} \sum_{t=1}^{T} c(I_t, q_t).$$

- Find the order quantity process adapted to the observed history to minimize $J(\mathbf{q})$.

# 1. Reference Inventory Position

- $x_t$: The latest observed inventory plus all the orders since then. Suppose $t \geq \theta + 2$

$$x_t := I_{t-\theta} + \sum_{i=1}^{\theta} q_{t-i}.$$

- Obtain the inventory level from the reference inventory position

$$I_t = x_t - \underbrace{\sum_{i=1}^{\theta} D_{t-i}}_{=:D^\theta}.$$

- Then evolve $x_t$ instead of $I_t$

$$x_{t+1} = x_t + q_t - D_{t-\theta}.$$

# 1. Rewrite the Problem

- Single-period cost w.r.t. reference inventory position

$$c^i(x,q) := \mathbb{E}\, c(x - D^i, q).$$

- Objective function

$$J(\mathbf{q}) = \mathbb{E} \sum_{t=1}^{T} c_t(x_t, q_t)$$

where

$$c_t(x,q) = \left\{ \begin{array}{ll} c^{t-1}(x,q) & \text{if } t \leq \theta + 1 \\ c^{\theta}(x,q) & \text{if } t \geq \theta + 2 \end{array} \right\}$$

# 1. Dynamic Program

$$
\begin{aligned}
V_{T+1}(x;\theta) &= 0, \\
V_t(x;\theta) &= \begin{cases} \inf_{q\geq 0} c^{t-1}(x,q) + V_{t+1}(x+q;\theta) & \text{for } t \leq \theta, \\ \inf_{q\geq 0} c^{\theta}(x,q) + \mathbb{E}\, V_{t+1}(x+q-D;\theta) & \text{for } t \geq \theta+1. \end{cases}
\end{aligned}
$$

- State $x$ is the reference inventory position.
- Nonzero salvage value and discounting can be added.
- For policy structure, let $c^i(x,q) = cq + \mathbb{E}\, h(x - D^i)$.

# 1. Basestock Policy

Optimal order quantity $q_t^*(x)$ is

$$q_t^*(x) = (u_t^*(\theta) - x)^+, \quad \text{where the base stock } u_t^*(\theta) \text{ is}$$

$$u_t^*(\theta) := \begin{cases} \arg\min_u \; cu + V_{t+1}(u;\theta), & t \le \theta, \\ \arg\min_u \; cu + \mathbb{E}\, V_{t+1}(u-D;\theta), & t \ge \theta + 1. \end{cases}$$

- Delay $\theta$ dependent basestock $u_t^*(\theta)$ is computed against the reference inventory position $x$.

# 1. Monotonicity Properties

- Marginal benefit of inventory is more with longer delay; Marginal cost of inventory is less with a longer delay,

$$\frac{d}{dx}V_t(x;\theta) \le \frac{d}{dx}V_t(x;\theta+1)$$

  Either for $[t \ge \theta + 1]$ or for $[x \le u_t^*(\theta)$ and $t \le \theta]$.

- Basestocks are larger with a longer delay

$$u_t^*(\theta) \le u_t^*(\theta+1)$$

  Either for $[t \ge \theta + 2]$ or for $[t \le \theta]$.

- Comparison of $u_{\theta+1}^*(\theta)$ and $u_{\theta+1}^*(\theta+1)$ requires additional conditions on the cost.
- Costs are smaller with shorter delay:$V_1(x;\theta) \le V_1(x;\theta+1)$.

# 1. Random Information Delay

- Delay $\theta$ is random.
- Eg: Building / repairing an information system which would become available at the end of period $\theta+1$.
- Delay has probabilities $p_k := \mathrm{P}(\theta = k)$.
- $\beta_t$: The period to which the most recent inventory observation belongs,

$$\beta_t = \left\{ \begin{array}{ll} 1 & \text{if } t \leq \theta + 1 \\ t - \theta & \text{if } t \geq \theta + 2 \end{array} \right\}, \quad \beta_1 = 1.$$

  $\{\beta_t : t \geq 1\}$ is a stochastic process derived from $\theta$.

- Eg: If $[\theta = 2]$, then $\beta_1 = \beta_2 = \beta_3 = 1$ and $\beta_4 = 2$, $\beta_5 = 3$.

# 1. Dynamic Program

Evolution of the system state

$$(x_t, \beta_t) \mapsto (x_t + q_t - D^{\beta_{t+1}-\beta_t}, \beta_{t+1})$$

where $(\beta_{t+1} - \beta_t)$ demands are observed at the begining of period $t+1$.

$$W_t(x,\beta) = \inf_{q \geq 0} \underbrace{c^{t-\beta}(x,q)}_{1-\text{period cost computed with } D^{t-\beta}} + \mathbb{E}\left[ \underbrace{W_{t+1}(x+q-D^{\beta_{t+1}-\beta}, \beta_{t+1})}_{\text{Future cost with } \beta_{t+1}} \middle| \beta_t = \beta \right]$$

Note that $W_t(x,\beta) = V_t(x, t-\beta)$ for $\beta \geq 2$.

# 1. When is the delay observed?

If the information system is not built by $t$, it may not be built by $t+1$ either:

$$\begin{aligned} \mathrm{P}(\beta_{t+1}=1|\beta_t=1) &= \mathrm{P}(\theta \geq t|\theta \geq t-1) \\ &= \frac{1-\sum_{k=0}^{t-1} p_k}{1-\sum_{k=0}^{t-2} p_k}. \end{aligned}$$

Or it is built by the begining of period $t+1$

$$\mathrm{P}(\beta_{t+1}=2|\beta_t=1) = \frac{p_{t-1}}{1-\sum_{k=0}^{t-2} p_k} =: \lambda_{t-1}.$$

- What is $\mathrm{P}(\beta_{t+1}=3|\beta_t=1)$? Hint: What is $\beta_t$ if $\beta_{t+1}=3$? $\beta_{t+1}=3 \implies \theta=t-2$. Then $\beta_t=2$. Interestingly, $\beta_{t+1}-\beta_t \leq 1$.
- Note that $\lambda_t$ is the hazard rate function for $\theta$.

# 1. Dynamic Program : $\beta_t = 1$

DP recursion with yet to be observed delay becomes

$$\begin{aligned} W_t(x,1) &= -cx + \mathbb{E}\, h(x - D^{t-1}) \\ &+ \min_{w \ge x} \left[ cw + \lambda_{t-1} \mathbb{E}\, V_{t+1}(w - D, t-1) \right. \\ &\qquad \left. + (1-\lambda_{t-1}) W_{t+1}(w,1) \right]. \end{aligned}$$

The recursion implies a basestock policy with

$$\begin{aligned} w_t^* &:= \arg\min_w \left[ cw + \lambda_{t-1} \mathbb{E}\, V_{t+1}(w - D, t-1) \right. \\ &\qquad \left. + (1-\lambda_{t-1}) W_{t+1}(w,1) \right], \\ q_t^*(x,\beta) &= \left\{ \begin{array}{ll} (w_t^* - x)^+ & \text{if } \beta = 1 \\ (u_t^*(t-\beta) - x)^+ & \text{if } \beta \ge 2 \end{array} \right\}. \end{aligned}$$

# 1. Markovian delays: $\{\theta_t : t \geq 1\}$

- Delay process is a Markov Chain characterized by transition probabilities

$$p^t_{i,j} := \mathrm{P}(\theta_{t+1} = j | \theta_t = i) \quad \text{for } t \geq 1$$

  and $\theta_0 = 0$.
- $\{\beta_t : t \geq 1\}$ is not necessarily a Markov chain anymore.
- Construct a new process $\{\tau_t : t \geq 1\}$ by

$$\tau_t := \max_{1 \leq j \leq t} \{j : \beta_{j-1} < \beta_j\},$$

  $\tau_t$ is the calendar time of the latest inventory observation.
- $\{(\beta_t, \tau_t) : t \geq 1\}$ is a Markov chain because

$$\tau_{t+1} = \mathbb{1}_{\beta_{t+1} = \beta_t} \tau_t + \mathbb{1}_{\beta_{t+1} > \beta_t} (t+1).$$

# 1. Markovian delays: DP formulation

- Let $\sigma_t := t - \tau_t$ and $\eta_t := \theta_{\tau_t}$. They are respectively the age of the latest inventory observation and the delay inferred from that observation.
- $\{(\sigma_t, \eta_t)\}$ is a Markov chain on account of $\{(\beta_t, \tau_t)\}$. Let

$$\gamma^t_{\sigma,\eta} := \mathrm{P}(\theta_{t+1} \geq \eta + \sigma + 1 | \theta_{t-\sigma} = \eta, \theta_{t-\sigma+1} \geq \eta + 1, \ldots, \theta_t \geq \eta + \sigma)$$

$$\gamma^t_{k,\sigma,\eta} := \mathrm{P}(\theta_{t+1} = k | \theta_{t-\sigma} = \eta, \theta_{t-\sigma+1} \geq \eta + 1, \ldots, \theta_t \geq \eta + \sigma)$$

- Basetock policy still holds in view of the DP equation

$$V_t(x, \sigma, \eta) = \inf_{q \geq 0} \Big\{ c^{\sigma+\eta}(x, q) + \alpha \gamma^t_{\sigma,\eta} V_{t+1}(x + q, \sigma + 1, \eta) + \alpha \sum_{k=0}^{\sigma+\eta} \gamma^t_{k,\sigma,\eta} V^{\sigma+\eta+1-k}_{t+1}(x + q, 0, k) \Big\}.$$

# 1. Concluding Remarks

- Referenece inventory position is a sufficient statistic for the delay problem.
- Basestock policies are optimal with Constant delays.
- State $\beta$ dependent basestock policies are optimal with Random delays.
- $(s, S)$ policies are optimal if there also are fixed ordering costs.

# 2. Filtered Newsvendor

- Many inventory models assume a static and a known demand distribution
- This distribution in reality can change over time and so it is likely to be unknown.
- The demand is observed *only* when it is less than the inventory level, i.e., unmet demand is *censored*.
- Ding et al. (2002), demands are iid.
- Our model, demands come from a Markov process, $p(x|\xi) := \mathrm{P}(D_{t+1} = x | D_t = \xi)$.
- Inventory is decoupled; Excess inventory is salvaged at the end of each period.
- $q_t$ is the inventory available to satisfy the demand $D_t$.
- Not the demand but the sale $z_t$ is observed. $z_t := \min\{D_t, q_t\}$. $\mathcal{Z}_t := \sigma(\{z_1, \dots, z_t\})$

# 2. Problem Definition

- $L(D, q)$, one-period cost function:

$$L(D,q) = \left\{ \begin{array}{ll} cq - h(q-D) & \text{if } D \le q \\ cq + b(D-q) & \text{if } q \le D \end{array} \right\},$$

  $h, c$ and $b$ are the salvage value per unit, the ordering cost per unit, and the shortage cost per unit. $0 \le h < c < b$.

- With the discount factor $0 < \alpha < 1$ and with $q_t$ adapted to $\mathcal{Z}_{t-1}$, minimize

$$J(\mathbf{q}) := \sum_{t=1}^{\infty} \alpha^{t-1} \mathbb{E}\, L(D_t, q_t).$$

# 2. DP with Normalized Probabilities

- $\pi_t$: the conditional probability density function of $D_t$.

$$\pi_{t+1}(x) \;=\; \mathbb{1}_{z_t=q_t} \frac{\int_{q_t}^{\infty} \pi_t(\xi) p(x|\xi) d\xi}{\int_{q_t}^{\infty} \pi_t(\xi) d\xi} + \mathbb{1}_{z_t<q_t} p(x|z_t).$$

Future costs should have two terms depending on whether the demand is observed. When the demand is not observed, the future costs depend on the first term on the right-hand side. Otherwise, they depend on the second term.

- Value function is:

$$V(\pi) \;=\; \min_q \left\{ \int L(x,q)\pi(x)dx + \alpha V\left(\frac{\int_q^{\infty} p(\cdot|\xi)\pi(\xi)d\xi}{\int_q^{\infty} \pi(\xi)d\xi}\right) \times \int_q^{\infty} \pi(\xi)d\xi + \alpha \int_0^q V(p(\cdot|\xi))\pi(\xi)d\xi \right\}.$$

# 2. Unnormalized Probabilities

- Define unnormalized probability $\rho_t$ by

$$\rho_{t+1}(x) = \mathbb{1}_{z_t=q_t} \int_{q_t}^{\infty} p(x|\xi)\rho_t(\xi)d\xi + \mathbb{1}_{z_t<q_t} p(x|z_t)$$

  with $\rho_1(x) = \pi_1(x)$.
- Set

$$\lambda_t := \int \rho_t(x)dx.$$

- Then

$$\lambda_1 = 1, \quad \lambda_{t+1} = \mathbb{1}_{z_t=q_t} \int_{q_t}^{\infty} \rho_t(\xi)d\xi + \mathbb{1}_{z_t<q_t}$$

# 2. DP with Unnormalized Probabilities

- Value function $W$ for unnormalized probabilities:

$$
\begin{aligned}
W(\rho) \;=\; \min_q \Bigg\{ \int L(x,y)\rho(x)dx + \alpha W\left(\int_q^\infty p(\cdot|\xi)\rho(\xi)d\xi\right) \\
+\alpha \int_0^q W(p(\cdot|\xi))\rho(\xi)d\xi \Bigg\}.
\end{aligned}
$$

- $W(0) = 0$ and $W$ is homogenous of degree 1, i.e., $W(a\rho) = aW(\rho)$ for $a > 0$.
- The right-hand side in the DP has linear growth facilitating the derivation of the existence results.

# 2. Existence and Monotonicity Results

- Define the map $T(W)$ as

$$T(W)(\rho) \quad := \quad \min_{q} \left\{ \int L(x,y)\rho(x)dx + \alpha W\left(\int_q^\infty p(\cdot|\xi)\rho(\xi)d\xi\right) \right. \\ \left. + \alpha \int_0^q W(p(\cdot|\xi))\rho(\xi)d\xi \right\}.$$

- $T$ is a contraction mapping when $\alpha$ is sufficiently small. Thus, there exists one and only one solution $W(\rho)$ of the DP equation. Moreover, $W(\rho)$ is continuous at each $\rho$ and there exists an optimal feedback policy.
- $X$ (its c.d.f. $\Pi$) is less than $X'$ (its c.d.f. $\Pi'$) in hazard rate ordering if $(1-\Pi'(z))/(1-\Pi(z))$ is increasing in $z$.
- $V(\pi) \leq V(\pi')$ if $\pi$ is smaller than $\pi'$ in the hazard rate order and $p(\cdot|\cdot)$ preserves the hazard rate order.

# 2. Example: A Two State Markov Chain for the Demand

- Demand takes only two values $x_1$ and $x_2$, $x_1 < x_2$.
- The unnormalized demand distribution $\rho = \theta_1 \delta_{x_1} + \theta_2 \delta_{x_2}$.
- The transition probabilities:

$$\begin{aligned} p(\cdot|x_1) &= \beta_1 \delta_{x_1} &+(1-\beta_1)\delta_{x_2} \\ p(\cdot|x_2) &= (1-\beta_2)\delta_{x_1} &+\beta_2 \delta_{x_2}. \end{aligned}$$

- DP equation for the cost $Z([\theta_1, \theta_2])$:

$$\begin{aligned} Z([\theta_1,\theta_2]) = \; & \min_y \{\theta_1 L_1(y) + \theta_2 L_2(y) \\ & + \alpha \mathbf{1}_{y \le x_1} Z([\theta_1\beta_1 + \theta_2(1-\beta_2), \theta_1(1-\beta_1) + \theta_2\beta_2]) \\ & + \alpha \mathbf{1}_{x_1 < y} [\theta_1 Z(\beta_1, 1-\beta_1) + \theta_2 Z([1-\beta_2, \beta_2])]\}. \end{aligned}$$

- $Z(a[\theta_1, \theta_2]) = aZ([\theta_1, \theta_2])$ for every scalar $a > 0$. In particular, $Z([\theta_1, \theta_2]) = \theta_2 Z([\theta_1/\theta_2, 1])$.

# 2. Example: A Two State Markov Chain for the Demand

- Define the unnormalized cost $\Phi(\theta) := Z([\theta, 1])$.
- DP equation for unnormalized cost $\Phi(\theta)$:

$$\begin{aligned}\Phi(\theta) &= \min_y \Big\{ \theta L_1(y) + L_2(y) \\ &\quad + \alpha \mathbb{1}_{y \le x_1} (\theta(1-\beta_1) + \beta_2) \Phi\left( \frac{\theta\beta_1 + 1 - \beta_2}{\theta(1-\beta_1)+\beta_2} \right) \\ &\quad + \alpha \mathbb{1}_{x_1 < y} \left[ \theta(1-\beta_1)\Phi\left(\frac{\beta_1}{1-\beta_1}\right) + \beta_2 \Phi\left(\frac{1-\beta_2}{\beta_2}\right) \right] \Big\} .\end{aligned}$$

where $\theta := \theta_1/\theta_2$.

# 2. Example: An Approximation to Break the Circle

- Computation of $\Phi(\theta)$ can be done by value iteration.
- Instead let $\beta = \beta_1 = \beta_2$, assume $\beta > 1/2$ and let $\sigma = \beta/(1-\beta)$
- $\beta > 1/2$ implies that the transition matrix preserves hazard rate order.
- Approximation:

$$\Phi\left(\frac{2\sigma}{\sigma^2+1}\right) \approxeq \Phi(1) \approxeq \Phi\left(\frac{\sigma^2+1}{2\sigma}\right).$$

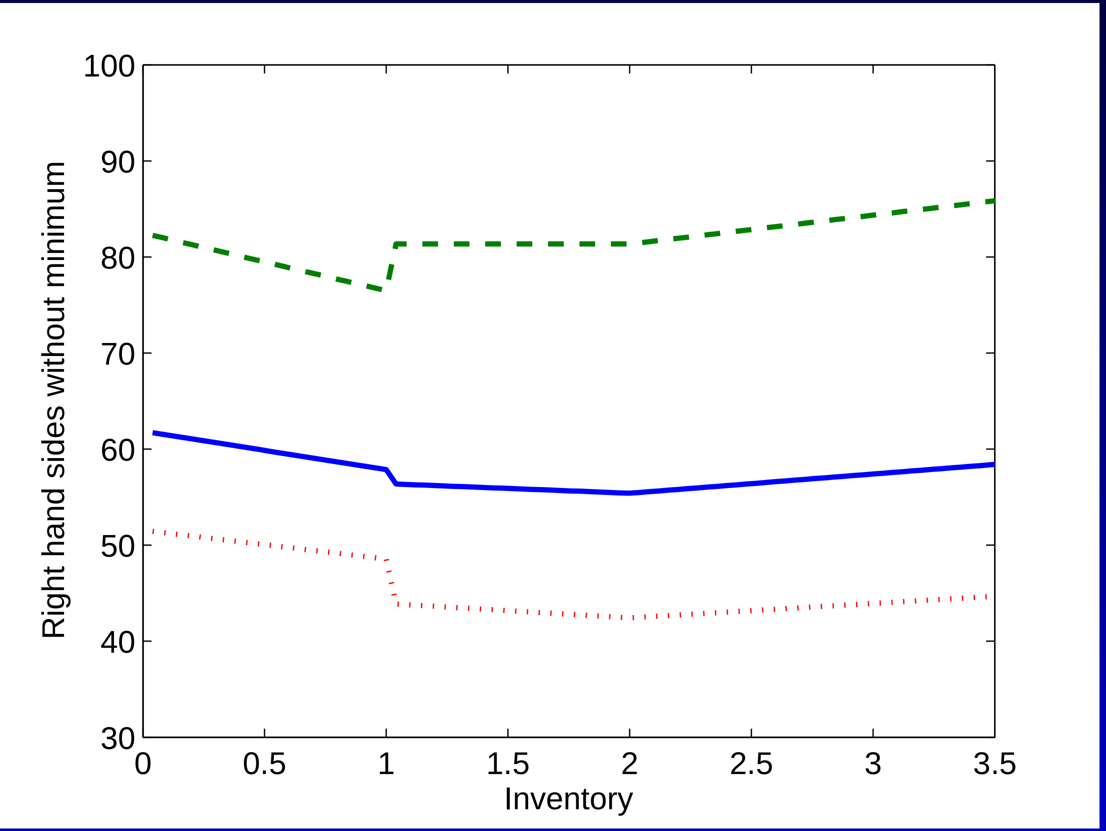
- Discretize $\Phi$ at three points $\{1/\sigma, 1, \sigma\}$. Solve 3 equations in 3 unknowns.

# 2. Example: 3 Equation and Unknown

$$
\begin{aligned}
\Phi(1) &= \min_y \Big\{ L_1(y) + L_2(y) + \alpha 1\!\!1_{y \le x_1} \Phi(1) \\
&\quad + \alpha 1\!\!1_{x_1 < y} \left( \frac{1}{1+\sigma} \Phi(\sigma) + \frac{\sigma}{1+\sigma} \Phi(1/\sigma) \right) \Big\}, \\
\Phi(\sigma) &= \min_y \Big\{ \sigma L_1(y) + L_2(y) + \alpha 1\!\!1_{y \le x_1} \frac{2\sigma}{1+\sigma} \Phi(1) \\
&\quad + \alpha 1\!\!1_{x_1 < y} \frac{\sigma}{1+\sigma} (\Phi(\sigma) + \Phi(1/\sigma)) \Big\}, \\
\Phi(1/\sigma) &= \min_y \Big\{ \frac{1}{\sigma} L_1(y) + L_2(y) + \alpha 1\!\!1_{y \le x_1} \frac{1+\sigma^2}{\sigma(1+\sigma)} \Phi(1) \\
&\quad + \alpha 1\!\!1_{x_1 < y} \left( \frac{1}{\sigma(1+\sigma)} \Phi(\sigma) + \frac{\sigma}{1+\sigma} \Phi(1/\sigma) \right) \Big\}.
\end{aligned}
$$

# 2. Example: 3 Equation and Unknown

Set $x_1 = 1$, $x_2 = 2$, $\sigma = 2$, $h = 1$, $c = 2$, $b = 4$ and $\alpha = 0.9$.




The terms inside $\{\cdot\}$ on the right-hand sides of eqs for $\Phi(1)$, $\Phi(\sigma)$, $\Phi(1/\sigma)$: Solid line "—" for $\Phi(1)$; Dashed line "– –" for $\Phi(\sigma)$; Dotted line "···" for $\Phi(1/\sigma)$.

# 2. Example: Results

- The right-hand sides for $\Phi(1)$, $\Phi(\sigma)$, $\Phi(1/\sigma)$ are minimized respectively at $y=2$, $y=1$ and $y=2$.
- $\sigma=1$ indicates that demands $x_1$ and $x_2$ are equally likely, while a larger $\sigma$ implies $x_1=1$ is more likely.
- When $\sigma=2$, the right hand side of $\Phi(\sigma)$ is minimized at $y=1$, i.e., when we believe that the demand is smaller, we order less.
- Approximate value functions for expected discounted cost:

$$
\begin{aligned}
V\left(\frac{1/2}{1+1/2}\delta_{x_1}+\frac{1}{1+1/2}\delta_{x_2}\right) &= \frac{42.43}{1+1/2} = 28.28,\\
V\left(\frac{1}{1+1}\delta_{x_1}+\frac{1}{1+1}\delta_{x_2}\right) &= \frac{55.41}{1+1} = 27.70,\\
V\left(\frac{2}{1+2}\delta_{x_1}+\frac{1}{1+2}\delta_{x_2}\right) &= \frac{76.49}{1+2} = 25.50.
\end{aligned}
$$

# 2. Concluding Remarks

- Partially observed demand leads to a DP in infinite dimensional spaces.
- The DP is higly nonlinear.
- Unnormalized probabilities linearize the state transition equations.
- This linearization facilitates the proofs of existence.
- Other application areas of these ideas:
  – Parameterized $\pi$, where the parameter is updated.
  – Zero Balance Walk model.

# 3. Zero Balance Walk Model

- When the inventory level is uncertain, inventory can be counted.
- Counting is costly. Counting is free of charge when inventory is zero. This is because a quick glance at the shelves can indicate that inventory is zero.
- In zero balance walk,
  - The inventory level is observed only when there is no physical inventory.
  - When there is inventory, only the event that the inventory is positive is observed.
- Signal $z_t := \mathbb{1}_{I_t=0}$ for $t \geq 0$.
- This observation process mimics what is known as “zero-balance walk" (Fisher et al., 2000 and Raman et al., 2001) at some companies where employees walk around the shelves to identify the stocked-out items.

# 3. Zero Balance Walk Model

- Unmet inventory is lost: $I_{t+1} = (I_t + q_t - D_t)^+$ for $t \geq 1$.
- Demands are i.i.d. A generic demand is denoted by $D$, which is i.i.d. with each $D_t$. Let $f$ denote the density and $F$ denote the cumulative distribution of $D$. Let $\bar{F} = 1 - F$.
- The order $q_t$ is adapted to $\mathcal{Z}_t := \sigma(\{z_j : 1 \leq j \leq t\})$. Clearly $\mathcal{Z}_t \subset \mathcal{F}_t := \sigma(\{I_j : 1 \leq j \leq t\})$.
- Given a stationary cost function $c(I_t, q_t)$ with $\mathbf{q}$ defining the admissible order quantities, the total discounted cost is

$$J(\zeta, \pi, \mathbf{q}) := \sum_{t=1}^{\infty} \alpha^{t-1} \mathbb{E}\, c(I_t, q_t).$$

- The initial conditions are a pair $(\zeta, \pi(x))$, where $\zeta$ is 1 or 0. If $\zeta$ is 1, then $I_1 = 0$. If $\zeta$ is 0, then $I_1 > 0$ and $\pi(\cdot)$ is the probability distribution of $I_1$.
- Find $q_t$ to minimize $J(\zeta, \pi, \mathbf{q})$.

# 3. Towards the Evolution of $\pi_t$

- Let $\pi_t(.)$ be the conditional probability density of $I_t$ given $\mathcal{Z}_{t-1}$ and $I_t > 0$.

- In order to obtain a recursive expression for $\pi_t$ in terms of $\pi_{t-1}$, we begin with expressing $\mathbb{E}(\phi(I_t)|\mathcal{Z}_t)$ in terms of conditional expectations with respect to $\mathcal{Z}_{t-1}$.

$$\begin{aligned}
\mathbb{E}(\phi(I_t)|\mathcal{Z}_t) &= \mathbb{1}_{I_t=0}\phi(0) + \mathbb{1}_{I_t>0}\frac{\mathbb{E}(\phi(I_t)\mathbb{1}_{I_t>0}|\mathcal{Z}_{t-1})}{\mathrm{P}(I_t > 0|\mathcal{Z}_{t-1})} \\
&= \mathbb{1}_{I_t=0}\phi(0) + \mathbb{1}_{I_t>0}\mathbb{E}(\phi(I_t)|\mathcal{Z}_{t-1}, I_t > 0).
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}(\phi(I_t)|\mathcal{Z}_t)\mathbb{1}_{I_t>0} &= \mathbb{1}_{I_{t-1}=0}\frac{\int_0^\infty \phi(z) f(q_{t-1} - z)\mathbb{1}_{q_{t-1}\geq z} dz}{F(q_{t-1})} \\
&+ \mathbb{1}_{I_{t-1}>0}\frac{\int_0^\infty \phi(z)\int_{(z-q_{t-1})^+}^\infty f(y + q_{t-1} - z)\pi_{t-1}(y)dydz}{\int_0^\infty F(y+q_{t-1})\pi_{t-1}(y)dy}.
\end{aligned}$$

# 3. Evolution of $\pi_t$

- Let $\pi_t(.)$ be the conditional probability density of $I_t$ given $\mathcal{Z}_{t-1}$ and $I_t > 0$.

$$\pi_t(x) \quad = \quad z_{t-1}\frac{f(q_{t-1}-x)\mathbb{1}_{x<q_{t-1}}}{F(q_{t-1})} \\ +(1-z_{t-1})\frac{\int_{(x-q_{t-1})^+}^{\infty} f(y+q_{t-1}-x)\pi_{t-1}(y)dy}{\int_0^{\infty} F(q_{t-1}+y)\pi_{t-1}(y)dy}$$

- This corresponds to the Kushner equation in our inventory context.
- The conditional probability evolves according to a highly nonlinear equation.

# 3. Unnormalized Probabilities

- Define unnormalized probability $\rho_t$ by

$$\begin{aligned} \rho_t(x) &= z_{t-1} f(q_{t-1} - x) \mathbb{1}_{x < q_{t-1}} \\ &\quad + (1 - z_{t-1}) \int_{(x-q_{t-1})^+}^{\infty} f(y + q_{t-1} - x) \rho_{t-1}(y) dy, \\ \rho_1(x) &= \pi(x). \end{aligned}$$

- Normalized and unnormalized probabilities can be related by

$$\rho_t(x) := \lambda_t \pi_t(x).$$

- Normalization constant $\lambda_t$ is given by

$$\lambda_t = z_{t-1} F(q_{t-1}) + (1 - z_{t-1}) \lambda_{t-1} \int F(q_{t-1} + y) \pi_{t-1}(y) dy.$$

# 3. Some Definitions

- Define the value function.

$$V(\zeta, \pi) := \inf_{\mathbf{q}} J(\zeta, \pi, \mathbf{q}).$$

- Define the innner product in appropriate functional spaces

$$\langle \rho, \phi \rangle = \int_0^\infty \rho(x)\phi(x)dx$$

- For any scalar $q > 0$, define the linear operator $\Phi$ as

$$\Phi(q, \rho)(x) = \int_{(x-q)^+}^\infty f(y + q - x)\rho(y)dy.$$

- Define the nonlinear operator $\Psi$ as

$$\Psi(q, \rho) = \frac{\Phi(q, \rho)}{\langle \Phi(q, \rho), 1 \rangle}.$$

# 3. DP with Normalized Probabilities

- We can write the state evolution with operators:

$$
\begin{aligned}
\pi_t &= z_{t-1}\Psi(q_t,\delta) \;+\; (1-z_{t-1})\Psi(q_t,\pi_{t-1}) \\
\rho_t &= z_{t-1}\Phi(q_t,\delta) \;+\; (1-z_{t-1})\Phi(q_t,\rho_{t-1})
\end{aligned}
$$

- Let $v := V(1,\pi)$ and $V(\pi) := V(0,\pi)$, then

$$
\begin{aligned}
V(\pi) &= \inf_q \Big\{ \langle c(.,q),\pi(.)\rangle \;+\alpha v \int \bar{F}(y+q)\pi(y)dy \\
&\quad +\alpha V(\Psi(q,\pi))F(y+q)\pi(y)dy \Big\}, \\
v &= \inf_q \left\{ c(0,q) + \alpha v\bar{F}(q) + \alpha V(\Psi(q,\delta))F(q) \right\}.
\end{aligned}
$$

# 3. DP with Unormalized Probabilities

- The study of the DP simplifies considerably when working with the unnormalized probability $\rho$.
- New value function $Z(.)$ where $Z(\rho) := V(\rho/\lambda)\,\lambda$.
- New system of DP equations:

$$
\begin{aligned}
Z(\rho) &= \inf_q \Big\{ \langle c(.,q), \rho(.) \rangle \; + \alpha v \int \bar{F}(y+q)\rho(y)dy \\
&+ \alpha Z(\Phi(q,\rho)) \Big\}, \\
v &= \inf_q \left\{ c(0,q) + \alpha v \bar{F}(q) + \alpha Z(\Phi(q,\delta)) \right\}.
\end{aligned}
$$

- We have $Z(0) = 0$ and $Z$ is homogenous of degree 1.

# 3. Existence of a Solution

- Bound the single period cost: Suppose that positive constants $c, c_0, c_1,$ and $h$ are such that $cq < c(x,q) \leq c_0 + c_1 q + hx$ for $x \geq 0$.
- Define the function $K(q,\rho;v,Z) :=$ $\langle c(.,q), \rho(.)\rangle + \alpha v \int \bar{F}(y+q)\rho(y)dy + \alpha Z(\Phi(q,\rho))$.
- Define the map $T(v;\ Z(\rho)) := (\inf_q K(q,\delta;v,Z);\ \inf_q K(q,\rho;v,Z))$.
- A value iteration scheme can be developed which converges to $(\bar{v}, \bar{Z})$.
- $(\bar{v}, \bar{Z})$ is the maximal solution satisfying $(v;Z) = T(v;Z)$. Also $\bar{Z}(\pi) = \inf_{\mathbf{q}} J(0,\pi,\mathbf{q})$ and $\bar{v} = \inf_{\mathbf{q}} J(1,\pi,\mathbf{q})$.
- $\bar{Z}(\pi)$ and $\bar{v}$ are interpreted as the infima of the costs. This, however, does not imply the existence of a feedback policy unless the order quantity $q$ is bounded.

# 3. Bounded Order Quantities

- Suppose that the orders are bounded by $m$:

$$
\begin{aligned}
Z^m(\rho) &= \inf_{q \le m} \Big\{ \langle c(.,q), \rho(.) \rangle \\
&\quad + \alpha v^m \int \bar{F}(y+q)\rho(y)dy + \alpha Z^m(\Phi(q,\rho)) \Big\} \\
v^m &= \inf_{q \le m} \left\{ c(0,q) + \alpha v^m \bar{F}(q) + \alpha Z^m(\Phi(q,\delta)) \right\}.
\end{aligned}
$$

- $Z^m$ is Lipschitz continuous.
- This additional smoothness property allows us to establish the uniqueness of a solution to the system above in the absence of a contraction property on $T$.

# 3. Concluding Remarks

- Partially observed inventory leads to a DP in infinite dimensional spaces.
- The DP is higly nonlinear.
- Unnormalized probabilities linearize the state transition equations.
- This linearization facilitates the proofs of existence.

**Unnormalized Probabilities are very useful in establishing the existence of solutions for partially observed systems.**