IE 516

Wenxiao Yang[*]

[*]Department of Mathematics, University of Illinois at Urbana-Champaign

2022

Contents

1 Lattice Programming

1.1 Lattice

Definition 1. *$(X, \geq)$ is a* ***lattice*** *if for any $x, y \in X$,*

$x \vee y = \inf\{z \in X | x \leq z, y \leq z\} \in X$

$x \wedge y = \sup\{z \in X | x \geq z, y \geq z\} \in X$

Definition 2. *$(X', \geq)$ is a* ***sublattice*** *of $(X, \geq)$: inherit $x \vee y$, $x \wedge y$ from X.*

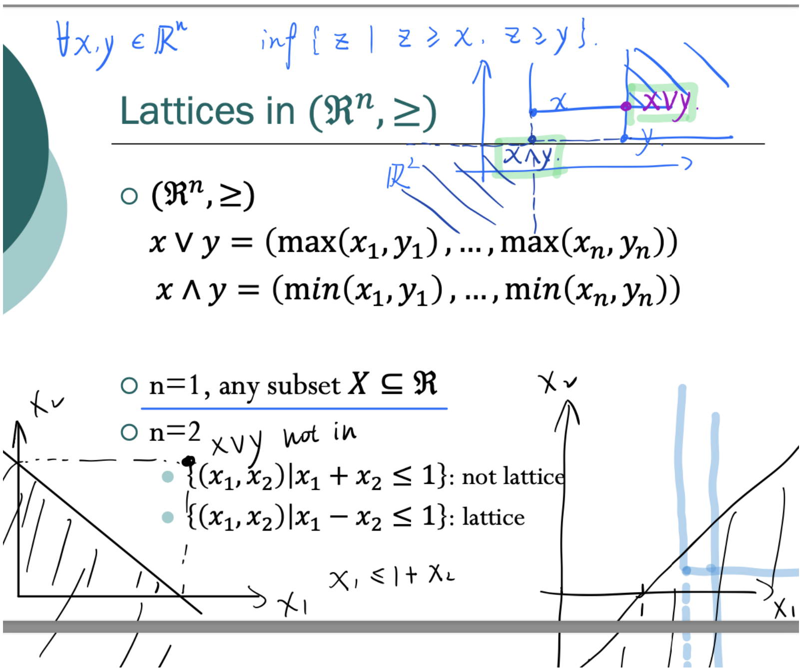


Figure 1:

Example 1. *Lattices:*

1. $\{0, 1\}^n$

2. Z^n

3. *a chain is a lattice (whose elements are ordered)*

4. *Intersection of two lattices*

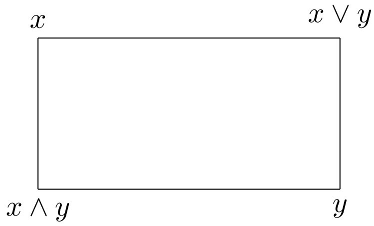
1.2 Supermodularity

1.2.1 Definition: Supermodular $g(x \vee y) + g(x \wedge y) \geq g(x) + g(y), \forall x, y \in X$

Definition 3. *A function $g : X \to \bar{\Re}(= \Re \cup \{+\infty\})$ is submodular if*

$$g(x \vee y) + g(x \wedge y) \leq g(x) + g(y), \forall x, y \in X$$

g is supermodular if $-g$ is submodular.



Claim 1. $\mathrm{dom}(g) = \{x \in X \mid g(x) < +\infty\}$ *is a lattice if g is submodular.*

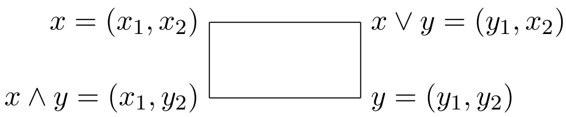
Proof. $\forall x, y \in \mathrm{dom}(g)$, prove $g(x \vee y) < +\infty$, $g(x \wedge y) < +\infty$. □

1.2.2 Lemma: Supermodular $\Leftrightarrow \frac{\partial^2 g(x)}{\partial x_i \partial x_j} \geq 0, \forall i \neq j$

Lemma 1. *Suppose g is twice partially differentiable in $\Re^n$. Then g is supermodular if and only if it has nonnegative cross partial derivatives, i.e.,*

$$\frac{\partial^2 g(x)}{\partial x_i \partial x_j} \geq 0, \forall i \neq j$$

Proof.



$$x_1 \leq y_1; \; y_2 \leq x_2$$

$$
\begin{aligned}
& g \text{ is supermodular} \\
\Leftrightarrow & g(x \vee y) - g(x) \geq g(y) - g(x \wedge y), \forall x, y \in X \\
& g(y_1, x_2) - g(x_1, x_2) \geq g(y_1, y_2) - g(x_1, y_2), \forall x, y \in X \\
& (\text{if } y_1 \to x_1,\ y_2 \text{ kept unchanged}) \\
& \frac{\partial g(x_1, x_2)}{\partial x_1} \geq \frac{\partial g(x_1, y_2)}{\partial x_1} \\
& (\text{if } y_2 \to x_2,\ y_2 \leq x_2) \\
& \frac{\partial^2 g(x)}{\partial x_i \partial x_j} \geq 0,\ \forall i \neq j
\end{aligned}
$$

□

Note: Supermodularity ≈ Economic Complementarity

g is the profit function of selling products x_1 and x_2, $\frac{\partial}{\partial x_2}(\frac{\partial g(x_1,x_2)}{\partial x_1}) \geq 0$

Example 2 (Examples of Supermodular Functions).

1. $f(x) = x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_n^{\alpha_n}$ $(\alpha_i \geq 0)$ *for* $x \geq 0$

2. $f(x, z) = \sum_{i=1}^{n} g_i\left(\alpha_i x_i - \beta_i z_i\right)$ *for any univariate concave function* $g_i : \Re \to \bar{\Re}$ $(\alpha_i \beta_i \geq 0)$

3. $f(x) = \sum_{i,j=1}^{n} a_{ij} x_i x_j = x^T A x$ *with* $a_{ij} = a_{ji}$ *is supermodular if and only if* $a_{ij} \geq 0\ \forall i \neq j$

1.2.3 Lemma: Preservation of Supermodularity

Lemma 2 (Preservation of Supermodularity).

a) If f_i *is supermodular, then* $\lim_{i \to \infty} f_i(x), \sum_i \alpha_i f_i$ $(\alpha_i \geq 0)$ *are supermodular*

b) If $f : \Re \to \Re$ *is convex and nondecreasing (nonincreasing) and* $g : \Re^n \to \Re$ *is increasing and supermodular (submodular), then* $f(g(x))$ *is supermodular*

c) Given $f : \Re^n \times \Re^m \to \Re$*, if* $f(\cdot, y)$ *is supermodular for all* y*, then* $E_\xi[f(x, \xi)]$ *is supermodular in* x

Lemma 3 (Supermodularity of composite functions).

If $X = \Pi_{i=1}^{n} X_i$ *and* $X_i \subseteq \Re, f_i\left(x_i\right) : X_i \to \Re$ *is increasing (decreasing) on* X_i *for* $i = 1, \ldots, n$*, and* $g\left(z_1, \ldots, z_n\right) : \Re^n \to \bar{\Re}$ *is supermodular in* $\left(z_1, \ldots, z_n\right)$*, then*

$$
g\left(f_1\left(x_1\right), \ldots, f_n\left(x_n\right)\right)
$$

is supermodular on X

Lemma 4 (Topkis 1998). *If X is a lattice, $f_i(x)$ is increasing and supermodular (submodular) on X for $i = 1, \ldots, k, Z_i$ is a convex subset of R^1 containing the range of $f_i(x)$ on X or $i = 1, \ldots, k$, and $g(z_1, \ldots, z_k, x)$ is supermodular in $(z_1, \ldots, z_k, x)$ and is increasing (decreasing) and convex in z_i for fixed z_{-i} and x, then $g(f_1(x), \ldots, f_k(x), x)$ is supermodular on X*

1.3 Parametric Optimization Problems

Definition 4.

$$
\begin{aligned}
f(s) = & \max g(s, a) \\
& s.t. \ a \in A(s)
\end{aligned}
$$

S: subset of $\Re^m$

$A(s)$: finite dimensional

$C := \{(s, a) \mid s \in S, a \in A(s)\}$ (the graph of the constraint operator)

$A^*(s)$, the optimal solution set, is nonempty for every $s \in S$

Definition 5. *A set $A(s)$ is* ***ascending on*** *S if for $s \leq s'$, $a \in A(s)$, $a' \in A(s')$, we have $a \wedge a' \in A(s)$ and $a \vee a' \in A(s')$.*

Example 3. *$A(s) = [s, +\infty)$ is ascending on S.*

1.3.1 Theorem: Maximizer of supermodular func is ascending, the maximum value is also supermodular

Theorem 1 (Ascending Optimal Solutions and Preservation).

If

1. *S: sublattice of $\Re^m$*
2. *$C := \{(s, a) \mid s \in S, a \in A(s)\}$ is a sublattice*
3. *g is supermodular on C*

Then

1. *$A^*(s)$ is* ***ascending*** *on S. Under some conditions, the largest/smallest element of $A^*(s)$ exists, and is increasing in* s.

2. $f(s)$ is supermodular.

Proof. Take $s \leq s'$, $a^* \in A^*(s),\ a'^* \in A^*(s')$, i.e.

$$g(s, a^*) = \max g(s, a) \text{ s.t. } a \in A(s)$$
$$g(s', a'^*) = \max g(s', a) \text{ s.t. } a \in A(s')$$

$$(s, a^*) \vee (s', a'^*) = (s', a^* \vee a'^*)$$
$$(s, a^*) \wedge (s', a'^*) = (s, a^* \wedge a'^*)$$

As we know C is a sublattice, we have

$$(s', a^* \vee a'^*) \in C \Rightarrow a^* \vee a'^* \in A(s')$$
$$(s, a^* \wedge a'^*) \in C \Rightarrow a^* \wedge a'^* \in A(s)$$

Hence,

$$g(s', a^* \vee a'^*) \leq g(s', a'^*);\ g(s, a^* \wedge a'^*) \leq g(s, a^*)$$

Since g is supermodular on C,

$$g(s', a^* \vee a'^*) + g(s, a^* \wedge a'^*) \geq g(s, a^*) + g(s', a'^*)$$
$$0 \geq g(s', a^* \vee a'^*) - g(s', a'^*) \geq g(s, a^*) - g(s, a^* \wedge a'^*) \leq 0$$

Hence,

$$g(s', a^* \vee a'^*) = g(s', a'^*);\ g(s, a^*) = g(s, a^* \wedge a'^*)$$

which means,

$$a^* \vee a'^* \in A^*(s'),\ a^* \wedge a'^* \in A^*(s)$$

Then, " $A^*(s)$ is ascending on S " is proved.

What's more, the largest elements of $A(s')$ and $A(s)$ are $a^* \vee a'^*$ and a^*, the smallest elements of $A(s')$ and $A(s)$ are a'^* and $a^* \wedge a'^*$, which are both increased as s increases to s'. □

Proof. $\forall s, s' \in S$, $a \in A^*(s), a' \in A^*(s')$.

$$f(s) + f(s') = g(s, a) + g(s, a')$$

(Since g is supermodular on C)

$$\leq g(s \wedge s', a \wedge a') + g(s \vee s', a \vee a')$$

$$\leq f(s \wedge s') + f(s \vee s')$$

" $f(s)$ is supermodular " is proved. □

Example 4. *Pricing:* $p^*(c) = \text{argmax}_{p \geq c'}(p-c)D(p)$, $(c' > c)$

1. $C = \{(p,c)|c < c', p \geq c'\}$ is a sublattice of $\mathbb{R}^2$.
2. $g(p,c) = (p-c)D(p)$, $\frac{\partial^2 g(p,c)}{\partial p \partial c} = -D'(p) \geq 0 \Rightarrow g$ is supermodular on C.

Hence, $p^*(c)$ is increasing in c.

Example 5. *Newsvendor model:* $\min_{x \geq 0} f(x) = cx + h_+ E[(x-\xi)^+] + h_- E[(\xi - x)^+]$

2 $L^\natural$-Convexity

2.1 Discrete Midpoint Convexity

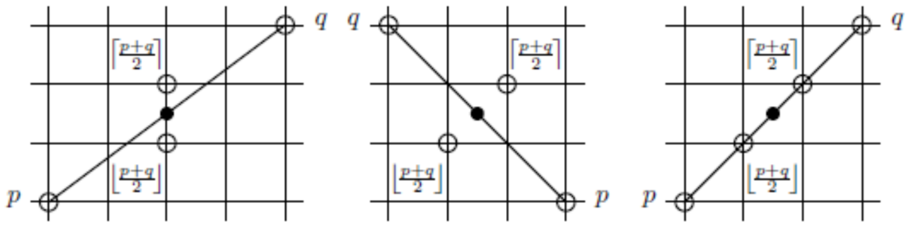


Figure 2: Discrete Midpoint Convexity

Definition 6. *A function f is **discrete midpoint convexity** if*

$$f(\lceil \frac{p+q}{2} \rceil) + f(\lfloor \frac{p+q}{2} \rfloor) \leq f(p) + f(q)$$

2.2 $L^\natural$-Convexity on $\mathbb{Z}^n$

Definition 7. *A function $f : Z^n \to \bar{\Re}$ is called $L^\natural$ convex if f satisfies the discrete midpoint convexity.*

An equivalent definition: A function $f : Z^n \to \bar{\Re}$ is $L^\natural$-convex if and only if

$$g(x,\alpha) := f(x - \alpha e) = f([x_1 - \alpha, x_2 - \alpha, ..., x_n - \alpha]^T)$$

is submodular in (x,α) on Z^{n+1}(e : all-ones vector).

2.3 $L^\natural$-Convexity on $\mathcal{F}^n(\mathcal{F} = \mathbb{Z}$ or $\Re)$

Definition 8 (Murota 2003).

A function $f : \mathcal{F}^{\mathrm{n}} \to \Re$ is $\underline{L^\natural\text{-convex if and only if } g(x,\xi) := f(x - \xi e) \text{ is submodular in } (x,\xi) \in \mathcal{F}^{\mathrm{n}} \times S}$, where e is a vector with all components equal to 1 and S is the intersection of $\mathcal{F}$ with any unbounded interval in $\Re$. (f is required to be convex if $\mathcal{F} = \Re$)

Definition 9. *A set V is $\underline{L^\natural\text{-convex if and only if its indicator function } \delta_V(x) \text{ is } L^\natural.}$*

$$\delta_V(x) = \begin{cases} +\infty & , x \notin V \\ 0 & , x \in V \end{cases}$$

$\Leftrightarrow$ *$g(x,\xi) = \delta_V(x - \xi e)$ is subnormal, i.e.*

$$g(x \vee y, \max\{\xi_x, \xi_y\}) + g(x \wedge y, \min\{\xi_x, \xi_y\}) \leq g(x, \xi_x) + g(y, \xi_y),\ \forall (x, \xi_x), (y, \xi_y)$$

If $x - \xi_x e, y - \xi_y e$ in V, $x \vee y - \max\{\xi_x, \xi_y\}e$ and $x \wedge y - \min\{\xi_x, \xi_y\}$ must in V.

Note: f is $L^\natural$-concave if $-f$ is $L^\natural$-convex.

2.4 Properties of $L^\natural$-Convexity

2.4.1 Proposition: $L^\natural$-convex $\Leftrightarrow a_{ij} \leq 0, \forall i \neq j, a_{ii} \geq 0, \sum_{j=1}^n a_{ij} \geq 0, \forall i$

Proposition 1. *A quadratic function $f(x) = \sum_{i,j=1}^n a_{ij} x_i x_j$ with $a_{ij} = a_{ji}$ is $\underline{L^\natural\text{-convex on } \mathcal{F} \text{ if and only if}}$ its Hessian is a diagonally dominated M-matrix*

$$a_{ij} \leq 0\ \forall i \neq j, \quad a_{ii} \geq 0, \quad \sum_{j=1}^n a_{ij} \geq 0\ \forall i$$

Proof.

$f(x)$ is $L^\natural$-convex $\Leftrightarrow g(x,\xi) = f(x - \xi e) = \sum_{i,j=1}^n a_{ij}(x_i - \xi)(x_j - \xi)$ is submodular in (x, ξ) i.e.

$$\begin{aligned} \frac{\partial^2 g}{\partial \xi \partial x_i} &= \frac{\partial}{\partial \xi}\Big(\sum_{j=1}^n a_{ij}(x_j - \xi) + \sum_{j=1}^n a_{ji}(x_j - \xi)\Big) = -2\sum_{j=1}^n a_{ij} \leq 0 \\ \frac{\partial^2 g}{\partial x_j \partial x_i} &= \frac{\partial}{\partial x_k}\Big(\sum_{k=1}^n a_{ik}(x_k - \xi) + \sum_{k=1}^n a_{ki}(x_k - \xi)\Big) = 2a_{ij} \leq 0 \end{aligned}$$

□

Proposition 2. *A twice continuous differentiable function $f : \Re^n \to \Re$ is $\underline{L^\natural\text{-convex if and only if}}$ its Hessian is a diagonally dominated M-matrix, that is*

$$a_{ij} \leq 0, \forall i \neq j, a_{ii} \geq 0, \sum_{j=1}^n a_{ij} \geq 0, \forall i$$

Proof.

$L^\natural$-convex $\Leftrightarrow g(x,\xi) = f(x - \xi e)$ is subnormal

(if twice differentiable)

$\Leftrightarrow \frac{\partial^2 g(x,\xi)}{\partial x_i \partial x_j} = \frac{\partial^2 f(x-\xi e)}{\partial x_i \partial x_j} \leq 0, i \neq j,\ \frac{\partial^2 g(x,\xi)}{\partial x_i \partial \xi} = -\sum_{j=1}^n \frac{\partial^2 f(x-\xi e)}{\partial x_i \partial x_j} \leq 0,\ \forall (x,\xi) \in \Re^{n+1}$

$\Leftrightarrow \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \leq 0, i \neq j;\ \sum_{j=1}^n \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \geq 0\ \forall x \in \Re^n$ ☐

2.4.2 Corollary: $L^\natural$-convex $\longrightarrow$ convex + submodular

Corollary 1. *If a twice differentiable function f is $L^\natural$-convex, then the function is convex and submodular.*

Proof.

$\underline{a_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \leq 0, i \neq j}$ means the cross partial derivatives are nonpositive, which equals to $\underline{f \text{ is submodular}}$.

$$\begin{aligned}
x^T \nabla^2 f(x) x &= \sum_{i,j=1}^n a_{ij} x_i x_j \\
&= \sum_k^n a_{kk} x_k^2 + \sum_{j=1}^n \sum_{i<j} a_{ij} 2 x_i x_j \\
&\geq \sum_k^n a_{kk} x_k^2 + \sum_{j=1}^n \sum_{i<j} a_{ij} (x_i^2 + x_j^2) \\
&\geq \sum_k^{n-1} a_{kk} x_k^2 + \sum_{j=1}^{n-1} \sum_{i<j} a_{ij} (x_i^2 + x_j^2) \\
&\cdots \\
&\geq 0, \quad \forall x \in \Re^n
\end{aligned}$$

Then f is convex. ☐

Example 6.

- Given any univariate (discrete) convex function $g_i : \mathcal{F} \to \bar{\Re}$ and $h_{ij} : \mathcal{F} \to \Re$, the function $f : \mathcal{F}^n \to \bar{\Re}$ defined by

$$f(x) := \sum_i g_i(x_i) + \sum_{i \neq j} h_{ij}(x_i - x_j)$$

is $L^\natural$-convex.

Example 7.

- A set with a representation

$$\{x \in \mathcal{F}^n : l \leq x \leq u, x_i - x_j \leq v_{ij}, i \neq j\}$$

is $L^\natural$-convex, where $l, u \in \mathcal{F}^n, v_{ij} \in \mathcal{F}$.

2.4.3 Theorem: Minimizer of $L^\natural$-convex func is nondecreasing with bounded sensitivity, the minimum value is also $L^\natural$-convex

Theorem 2. *Assume $g : \mathcal{F}^n \times \mathcal{F}^m \to \overline{\Re}$ and set $C \subset \mathcal{F}^n \times \mathcal{F}^m$ are $L^\natural$-convex, define*

$$f(s) = \inf_{a:(s,a)\in C} g(s, a)$$

Then,

1. *The optimal solution set $A^*(s)$ is nondecreasing in s with bounded sensitivity i.e.,*

   $$A^*(s + \omega e) \leq A^*(s) + \omega e, \ \forall \omega \in F_+$$

   (Zipkin 2008, Chen et al. 2018)

2. *f is $L^\natural - convex$. (Zipkin 2008)*

2.5 Relationship with Multimodularity

Definition 10. *A function $f(x_1, x_2, ..., x_n)$ is multimodular if $f(x_1 - x_0, x_2 - x_1, ..., x_n - x_{n-1})$ submodular in $(x_0, x_1, ..., x_n)$.*

Multimodularity and $L^\natural$-convexity are equivalent subject to an unimodular linear transformation.

3 Optimization with decisions truncated by random variables

$$\min_{u\in\mathcal{U}} E[f(u \wedge \xi)]$$

Question 1 (Supply uncertainty in SCM): u: ordering quantities; ξ: random capacities.

Question 2 (Demand uncertainty in RM): u: booking limits; ξ: random demands.

Difficulty: the object function is not convex (even if f is).

3.1 Unconstrained Problem

Consider

$$\tau^* = \min_{u\in\mathcal{F}^n} E[f(u\wedge\xi)]$$

$\mathcal{F}$ is either the real space or the set with all integers.

Random vector $\xi \in \mathcal{X} \subseteq \mathcal{F}^n$

3.1.1 Reformulation

Reformulation:

$$\begin{array}{lll} \min & E[f(v(\xi))] & \\ s.t. & v(\xi)= (v_1(\xi_1),...,v_n(\xi_n)) & ,\forall\xi\in\mathcal{X} \\ & v(\xi)= u\wedge\xi & ,\forall\xi\in\mathcal{X} \end{array}.$$

Turn finding u^* into finding $v^*()$

$v()$ is not convex.

Theorem 3 (Equivalent Transformation, Chen, Gao and Pang 2018)**.** *Suppose that (Assumption I)*

(a) the function f is lower semi-continuous with $f(u)\to+\infty$ for $|u|\to+\infty$;

(b) the function f is componentwise (discrete) convex;

(c) the random vector ξ has independent components.

Then τ^* is also the optimal objective value of the following optimization problem:

$$\begin{array}{lll} \min & E[f(v(\xi))] & \\ s.t. & v(\xi)\leq\xi & ,\forall\xi\in\mathcal{X} \\ & v(\xi)= (v_1(\xi_1),...,v_n(\xi_n)) & ,\forall\xi\in\mathcal{X} \end{array}.$$

3.1.2 $n=1$

$\hat{u}$: minimizer of $f(u)$

Need to show

$$\min_u E[f(u\wedge\xi)] = \min_{v(\xi)\leq\xi} E[f(v(\xi))]$$

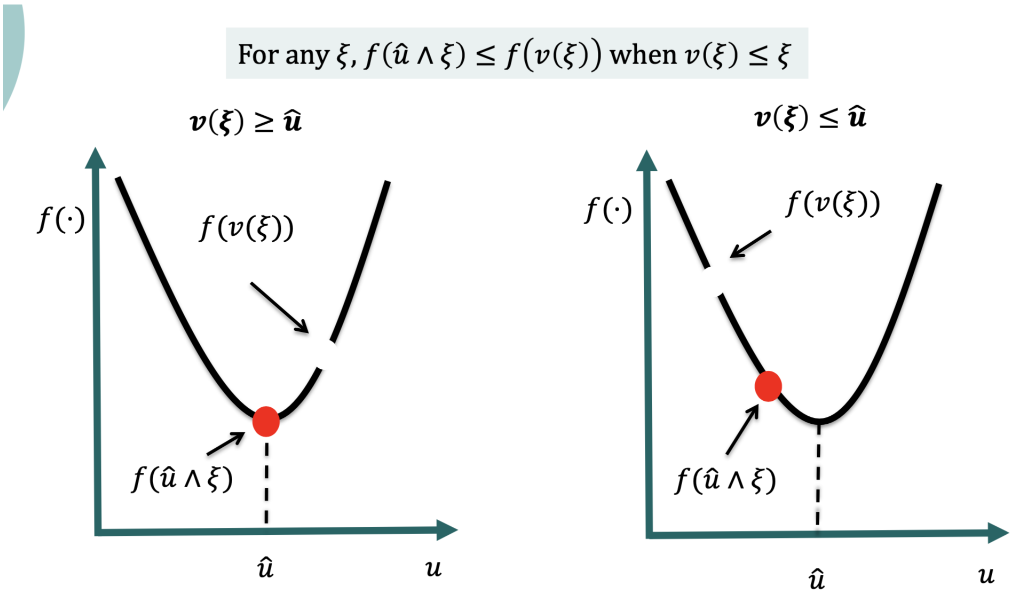


Figure 3: Easy to show $\forall \xi,\ f(\hat{u} \wedge \xi) \le f(v(\xi))$ when $v(\xi) \le \xi$

Easy to show $\forall \xi,\ f(\hat{u} \wedge \xi) \le f(v(\xi))$ when $v(\xi) \le \xi$. Then

$$\text{argmin}\, E[f(u \wedge \xi)] = \hat{u} = \text{argmin}\, f(u)$$

$$\begin{aligned}
E[f(\hat{u} \wedge \xi)] &\ge \min_u E[f(u \wedge \xi)] \\
&\ge \min_{v(\xi) \le \xi} E[f(v(\xi))] \text{ (Consider } v^*(\xi) \ge u) \\
&\ge E[f(\hat{u} \wedge \xi)] \text{ (See the figure)} \\
\Rightarrow \quad & \min_u E[f(u \wedge \xi)] = \min_{v(\xi) \le \xi} E[f(v(\xi))]
\end{aligned}$$

3.1.3 $n \ge 2$

$$\text{argmin}\, E[f(u \wedge \xi)] \ne \hat{u}$$

Example 8.

$$f(u_1, u_2) = (u_1 + u_2 - 2)^2 + (u_1 - 1)^2 + (u_2 - 1)^2$$

ξ_1, ξ_2 *can take values* 0 *and* 2 *with equal probability.*

$\hat{u} = (1, 1)$

$\text{argmin}\, E[f(u \wedge \xi)] = (1.2, 1.2)$

3.2 Transformation for Constrained Problem

$$\min_{u \in \mathcal{U}} E[f(u \wedge \xi)]$$

$$\downarrow$$

$$\begin{array}{lll} \min & E[f(v(\xi))] & \\ s.t. & v(\xi) \leq \xi & , \forall \xi \in \mathcal{X} \\ & v(\xi) = (v_1(\xi_1), ..., v_n(\xi_n)) \in \mathcal{V} & , \forall \xi \in \mathcal{X} \\ & \mathcal{V} = \{u \wedge \xi \ | u \in \mathcal{U}, \xi \in \mathcal{X}\} & \end{array} .$$

Sufficient Conditions for the Transformation

(a) $\mathcal{U} = \{u | Au \leq b, u \geq l\}$, where $A \geq 0$

(b) $\mathcal{X}_j \subseteq [l_j, +\infty)$

(Example: some situations $l = (l_1, ..., l_n) = (0, ..., 0)$)

3.3 Generalization

$$\min_{u \in \mathcal{F}^n} l(u) + E[f(u \wedge \xi)]$$

- $l : \mathcal{F}^n \to \bar{\Re}$, $f : \mathcal{F}^n \to \bar{\Re}$
- $\xi \in \mathcal{X} \subseteq \mathcal{F}^n$
- ξ **dependent** (different from before !)

3.3.1 Positive Dependence

Let F_{ξ_i} be the joint CDF of $\xi_1, \dots, \xi_{i-1}, \xi_{i+1}, \dots, \xi_n$ conditioned on ξ_i

$\{\xi_1, \dots, \xi_{i-1}, \xi_{i+1}, \dots, \xi_n \mid \xi_i\}$ is <u>stochastically increasing</u> if $\int_S dF_{\xi_i}(w)$ is an increasing function of ξ_i for each increasing set S

$\{\xi_1, \dots, \xi_{i-1}, \xi_i, \xi_{i+1}, \dots, \xi_n\}$ has <u>positive dependence</u> if $\{\xi_1, \dots, \xi_{i-1}, \xi_{i+1}, \dots, \xi_n \mid \xi_i\}$ is stochastically increasing for all i

Proposition 3. *The collection of random variables generated by nonnegative linear combination of **independent log-concave** random variables has positive dependence.*

3.3.2 Transformation

Theorem 4 (Equivalent Transformation, Chen and Gao 2018). *Suppose that (Assumption II)*

(1) the function f is lower semi-continuous with $f(u) \to +\infty$ for $|u| \to +\infty$;

(2) the function f is componentwise (discrete) convex and supermodular;

(3) the random vector ξ is positive dependent;

(4) $l(u)$ is componentwise increasing.

Then problem $\min_{u \in \mathcal{F}} l(u) + E[f(u \wedge \xi)]$ has the ***same optimal objective value*** *of*

$$\begin{array}{lll} \min & l(u) + E[f(v(\xi))] & \\ s.t. & v(\xi) \leq \xi & , \forall \xi \in \mathcal{X} \\ & v(\xi) \leq u & , \forall \xi \in \mathcal{X} \\ & v(\xi) = (v_1(\xi_1), ..., v_n(\xi_n)) & , \forall \xi \in \mathcal{X} \\ & v_i(\xi_i) \text{ is increasing for all } i & \end{array}.$$

4 Single-Leg Capacity Allocation

(Seats reserved for future consumers)

4.1 Two-Class Model

Two periods: Period 1, random demand D_2 for price p_2; Period 2, random demand D_1 for price p_1.

$p_1 > p_2$

Provide y in period 1 and the remaining will be provided in period 2.

$$\begin{array}{ll} \max & p_1 E_{D_1,D_2}[D_1 \wedge (c - (c - y) \wedge D_2)] + p_2 E_{D_2}[(c - y) \wedge D_2] \\ s.t. & 0 \leq y \leq c, y \in \mathcal{F} \end{array}.$$

Where $\mathcal{F} = \mathbb{R}$ or $\mathbb{Z}$ and $a \wedge b = \min(a, b)$

4.1.1 Theorem: convex f, $\operatorname{argmin} E_D f(u \wedge D) = \operatorname{argmin} f(u)$

When D_2 is sufficiently high. Let $b = c - y$, and the question transferred to

$$\begin{array}{ll} \max & v(b) = p_1 E_{D_1}[D_1 \wedge (c - b)] + p_2 b \\ s.t. & 0 \leq b \leq c, b \in \mathcal{F} \end{array}.$$

$v(b)$ is a concave function.

Theorem 5. *Consider the following optimization problem*

$$\begin{aligned} \min \quad & E_D f(u \wedge D) \\ s.t. \quad & 0 \le u \le c, u \in \mathcal{F} \end{aligned} \quad .$$

Assume D is a nonnegative random variable.

*If f is **convex** and $\mathcal{F} = \mathbb{R}$ or f is **discrete convex** and $\mathcal{F} = \mathbb{Z}$, then any optimal solution of*

$$\begin{aligned} \min \quad & f(u) \\ s.t. \quad & 0 \le u \le c, u \in \mathcal{F} \end{aligned} \quad .$$

is also optimal for the former optimization problem.

(Actually, quasi-convexity suffices)

According to the $n = 1$ discussion of section 3, the theorem is easy to be proved. Then, the global-max in $v(b)$ is global-max for objective function.

Then we consider the equivalent minimum problem,

$$\begin{aligned} \max \quad & \phi(y) = p_2 y - p_1 E_{D_1}[D_1 \wedge y] \\ s.t. \quad & 0 \le y \le c, y \in \mathcal{F} \end{aligned} \quad .$$

We need to find the optimal y^* minimize the $\phi(y)$. To simplify the analysis, we find the y° which is the optimal y regardless constraints. $y^* = \begin{cases} 0 & \text{if } y^\circ < 0 \\ y^\circ & \text{if } y^\circ \in [0, c] \\ c & \text{if } y^\circ > c \end{cases}$

4.1.2 Discrete, $\mathcal{F} = \mathbb{Z}$

$$\phi(y) - \phi(y-1) = p_2 - p_1 P(D_1 \ge y)$$

Then, the y° is

$\overline{y} = \min\{y \in \mathbb{Z} : P(D_1 > y) < r\}$

$\underline{y} = \max\{y \in \mathbb{Z} : P(D_1 \ge y) > r\}$ **(Littlewood's rule)**

$y^\circ = [\underline{y}, \overline{y}] \cap \mathbb{Z}$

Where $r = \frac{p_2}{p_1}$, higher r causes lower y°.

Example 9. *Suppose that D_1 is a Poisson random variable with mean 80, the full fare is $p_1 = 100$ and the discounted fare is $p_2 = 60$*

$r = 60/100 = 0.6$, $y^* = \max\{y \in \mathbb{Z} : P(D_1 \geq y) > r\} = 78$

4.1.3 Continuous, $\mathcal{F} = \mathbb{R}$

y° is the y s.t. $1 - F_1(y) = r$, where $F_1(\cdot)$ is the CDF of D_1.

$$y^\circ = F_1^{-1}(1 - r)$$

Special Case: $D_1 \sim \mathcal{N}(\mu, \sigma^2)$

$$F_1(y) = \Phi(\frac{y - \mu}{\sigma})$$

$\Phi(\cdot)$ is the CDF of the standard normal $\mathcal{N}(0, 1)$. Then,

$$y^\circ = \mu + \sigma\Phi^{-1}(1 - r)$$

If $\frac{p_2}{p_1} = r < \frac{1}{2}$, y° increases as variance σ increases.

4.2 Multi-Class Model

- $p_1 > p_2 > \cdots > p_n$
- Lower class demand arrives earlier.
- Demand of different classes are independent.
- Control: demand to accept or reject.

4.2.1 Sequence of Events

At stage j with remaining capacity x,

1. Select booking limit b for class j, equivalently, protection level $y = x - b$ for classes l, $l < j$.
2. Demand D_j is realized.
3. Accept $b \wedge D_j$ of class j and collect revenue $p_j(b \wedge D_j)$.
4. Move on to stage $j - 1$ with remaining capability $x - b \wedge D_j$.

4.2.2 Dynamic Programming

Set $f_j(x,b) = p_j b + V_{j-1}(x-b)$, $V_0(x) = 0, V_j(0) = 0$, $x = 0, 1, ..., c$ (discrete), $x \in [0, c]$ (continuous)

$$V_j(x) = \max_{b \in [0,x], b \in \mathcal{F}} \mathbb{E}[f_j(x, b \wedge D_j)] = \mathbb{E}[p_j(b \wedge D_j)] + \mathbb{E}[V_{j-1}(x - b \wedge D_j)]$$

Proposition 4. *(1). $\forall j$, f_j is $L^\natural-$concave, V_j is (discrete) convex; (2). The optimal solution of the dynamic programming b_j^* is the same as*

$$\max_{b \in [0,x], b \in \mathcal{F}} f_j(x,b) = p_j b + V_{j-1}(x-b)$$

Define y_{j-1}^* be the optimal solution of

$$\max_{y \geq 0, y \in \mathcal{F}} -p_j y + V_{j-1}(y)$$

Then

$$b_j^* = (x - y_{j-1}^*)^+$$

$$\begin{aligned}
V_j(x) &= \mathbb{E}[f_j(x, (x - y_{j-1}^*)^+ \wedge D_j)] \\
&= \mathbb{E}[p_j(x - y_{j-1}^*)^+ \wedge D_j + V_{j-1}(x - (x - y_{j-1}^*)^+ \wedge D_j)] \\
&= \begin{cases} V_{j-1}(x) & \text{if } x \leq y_{j-1}^* \\ \mathbb{E}[p_j(x - y_{j-1}^*) \wedge D_j + V_{j-1}(x - (x - y_{j-1}^*) \wedge D_j)] & \text{if } x > y_{j-1}^* \end{cases}
\end{aligned}$$

4.3 Discrete Case

Define

$$\Delta V_j(x) = V_j(x) - V_j(x-1)$$

Lemma 5. *If $x > y_{j-1}^*$, $\Delta V_j(x) = \mathbb{E}[\min\{p_j, \Delta V_{j-1}(x - D_j)\}]$*

Proof.

$$
\begin{aligned}
\Delta V_j(x) &= p_j(\mathbb{E}[(x - y^*_{j-1}) \wedge D_j] - \mathbb{E}[(x-1-y^*_{j-1}) \wedge D_j]) \\
&\quad + \mathbb{E}[V_{j-1}(x - (x - y^*_{j-1}) \wedge D_j)] - \mathbb{E}[V_{j-1}(x - 1 - (x-1-y^*_{j-1}) \wedge D_j)] \\
&= \begin{cases} p_j & \text{if } x - y^*_{j-1} \leq D_j \\ \Delta V_{j-1}(x - D_j) & \text{if } x - y^*_{j-1} > D_j \end{cases} \\
&= \mathbb{E}[p_j \mathbb{I}(x - D_j \leq y^*_{j-1}) + \Delta V_{j-1}(x - D_j)\mathbb{I}(x - D_j > y^*_{j-1})] \\
&\quad (\text{ Since } y^*_{j-1} \text{ maximizes } -p_j y + V_{j-1}(y), \\
&\quad \Delta V_{j-1}(y) > p_j \text{ if } y \leq y^*_{j-1} \text{ and } \Delta V_{j-1}(y) \leq p_j \text{ if } y > y^*_{j-1}) \\
&= \mathbb{E}[\min\{p_j, \Delta V_{j-1}(x - D_j)\}]
\end{aligned}
$$

□

Proposition 5 (1.5 of GT 19)**.**

(i) $\Delta V_j(x+1) \leq \Delta V_j(x)$ *(proved by* V_j *is discrete concave)*

(ii) $\Delta V_{j+1}(x) \geq \Delta V_j(x)$

Proof.

If $x \leq y^*_{j-1}$,

$$\Delta V_j(x) = V_{j-1}(x) - V_{j-1}(x-1) = \Delta V_{j-1}(x)$$

If $x > y^*_{j-1}$ (i.e. $x - 1 \geq y^*_{j-1}$),

$$
\begin{aligned}
\Delta V_j(x) &= \mathbb{E}[\min\{p_j, \Delta V_{j-1}(x - D_j)\}] \\
&\quad (\ V_{j-1}(x) \text{ is discrete concave}) \\
&\geq \mathbb{E}[\min\{p_j, \Delta V_{j-1}(x)\}] \\
&\quad (\text{ Since } x > y^*_{j-1},\ V_{j-1}(x) < p_j) \\
&= \Delta V_{j-1}(x)
\end{aligned}
$$

□

Theorem 6 (part of 1.6 of GT 19)**.**

The optimal protection level at stage j *is*

$$y^*_{i-1} = \max\{y \in \mathbb{N}_+ : \Delta V_{j-1}(y) > p_j\}$$

Moreover, $y^*_{n-1} \geq y^*_{n-2} \geq \cdots \geq y^*_1 = y^*_0 = 0$

(Easy to prove: Since y_{j-1}^* maximizes $-p_j y + V_{j-1}(y)$, $\Delta V_{j-1}(y) > p_j$ if $y \le y_{j-1}^*$ and $\Delta V_{j-1}(y) \le p_j$ if $y > y_{j-1}^*$)

Note: Littlewood's rule is a special case for $n = 2$.

4.3.1 Discrete Case: Reformulation

$$\begin{aligned}
V_j(x) &= \mathbb{E}[p_j(x-y_{j-1}^*)^+ \wedge D_j + V_{j-1}(x-(x-y_{j-1}^*)^+ \wedge D_j)] \\
&= V_{j-1}(x) + \mathbb{E}[p_j(x-y_{j-1}^*)^+ \wedge D_j + (V_{j-1}(x-(x-y_{j-1}^*)^+ \wedge D_j) - V_{j-1}(x))] \\
&= V_{j-1}(x) + \mathbb{E}[p_j(x-y_{j-1}^*)^+ \wedge D_j - \sum_{z=1}^{(x-y_{j-1}^*)^+ \wedge D_j} \Delta V_{j-1}(x+1-z)] \\
&= V_{j-1}(x) + \mathbb{E}[\sum_{z=1}^{(x-y_{j-1}^*)^+ \wedge D_j} (p_j - \Delta V_{j-1}(x+1-z))] \\
V_j(x) &= V_{j-1}(x) + \mathbb{E}[\sum_{z=1}^{u^*}(p_j - \Delta V_{j-1}(x+1-z))] \\
&\quad u^* = \min\{(x-y_{j-1}^*)^+, D_j\} \\
&\quad y_{i-1}^* = \max\{y \in \mathbb{N}_+ : p_j < \Delta V_{j-1}(y)\}
\end{aligned}$$

- $y_1^* \le y_2^* \le \cdots \le y_n^*$
- The "nested" booking limit $b_j^* = C - y_{j-1}^*$, $j = 2, ..., n$

  (nested booking limit is the total amount can be booked in $j, j+1, ..., n$)

$$b_j^* = y_j$$

- The marginal utility at j of choosing to reserve one more item in the next stage $j-1$:

$$\pi_j(x) = \Delta V_{j-1}(x)$$

- The amount of selling at stage j

$$u^* = \begin{cases} 0 & \text{if } p_j < \pi_j(x) \\ \min\{\max\{z : p_j \ge \pi_j(x-z)\}, D_j\} & \text{if } p_j \ge \pi_j(x) \end{cases}$$

  $p_j < \pi_j(x)$ means the marginal utility of reserving is larger than selling it now.

We can further compute, if $x > y_{j-1}^*$,

$$\Delta V_j(x) = p_j Pr(D_j \ge x - y_{j-1}^*) + \sum_{k=0}^{x-y_{j-1}^*-1} \Delta V_{j-1}(x-k) Pr(D_j = k)$$

If $x \leq y^*_{j-1}$, $\Delta V_j(x) = \Delta V_{j-1}(x)$.

Which will simplify the computation.

4.3.2 Discrete Case: Computation

The policy is implemented as follows:

1. At stage n, we start with $x_n = c$ units of inventory and we protect $y_{n-1}(x_n) = \min\left\{y^*_{n-1}, x_n\right\}$ units of capacity for fares $n-1, n-2, \ldots, 1$.
2. Therefore, we allow up to $\left[x_n - y^*_{n-1}\right]^+$units of capacity to be sold to fare class n.
3. We sell $\min\left\{\left[x_n - y^*_{n-1}\right]^+, D_n\right\}$ units of capacity to fare class n and we have a remaining capacity of $x_{n-1} = x_n - \min\left\{\left[x_n - y^*_{n-1}\right]^+, D_n\right\}$ at stage $n-1$.
4. We protect $y_{n-2}(x_{n-1}) = \min\left\{y^*_{n-2}, x_{n-1}\right\}$ units of capacity for fares $n-2, n-1, \ldots, 1$.
5. Therefore, we allow up to $\left[x_{n-1} - y^*_{n-2}\right]^+$units of capacity to be sold to fare class $n-1$.
6. We continue until we reach stage 1 with x_1 units of capacity, allowing $(x_1 - y_0)^+ = (x_1 - 0)^+ = x_1$ to be sold to fare class 1.

$$V_j(x) = \mathbb{E}[p_j \min\{(x - y^*_{j-1})^+, D_j\} + V_{j-1}(x - \min\{(x - y^*_{j-1})^+, D_j\})]$$

$y^*_0 = 0, V_0(x) = 0$, then we can compute $y^*_1, V_1(x)$,...

Backward: Use

$$\Delta V_j(x) = p_j Pr(D_j \geq x - y^*_{j-1}) + \sum_{k=0}^{x - y^*_{j-1} - 1} \Delta V_{j-1}(x-k) Pr(D_j = k)$$

$$y^*_{j-1} = \max\{y \in \mathbb{N}_+ : p_j < \Delta V_{j-1}(y)\}$$

1. $V_1(x_1) = \mathbb{E}[p_1 \min\{x_1, D_1\}]$, then $\Delta V_1(x) = p_1 Pr(D_1 \geq x)$
2. $y^*_1 = \max\{y \in \mathbb{N}_+ : p_2 < \Delta V_1(y)\} = \max\{y : Pr(D_1 \geq y) > \frac{p_2}{p_1}\}$
$$\Delta V_2(x) = p_2 Pr(D_2 \geq x - y^*_1) + \sum_{k=0}^{x - y^*_1 - 1} p_1 Pr(D_1 \geq x - k) Pr(D_2 = k)$$
3. $y^*_2 = \max\{y \in \mathbb{N}_+ : p_3 < \Delta V_2(y)\} = \max\{y : Pr(\Delta V_1(y - D_2) > p_3)\}$
4. ...

The complexity is $O(nC^2)$

Example 10. *Suppose that there are five fare classes. The demand for all fare classes is a Poisson random variable. The fares and the expected demand for the five fare classes are given by* $(p_5, p_4, p_3, p_2, p_1) = (15, 35, 40, 60, 100)$ *and* $(\mathbb{E}D_5, \mathbb{E}D_4, \mathbb{E}D_3, \mathbb{E}D_2, \mathbb{E}D_1) = (120, 55, 50, 40, 15)$. *For this problem instance, the optimal protection levels are*

1. $V_1(x_1) = \mathbb{E}[100 \min\{x_1, D_1\}]$, then $\Delta V_1(x) = 100Pr(D_1 \geq x)$
2. $y_1^* = \max\{y : Pr(D_1 \geq y) > \frac{3}{5}\} = 14$

$$\Delta V_2(x) = 60Pr(D_2 \geq x - 14) + \sum_{k=0}^{x-15} 100Pr(D_1 \geq x - k)Pr(D_2 = k)$$

3. $y_2^* = \max\{y \in \mathbb{N}_+ : p_j <$

4.4 Continuous Case

Skip

4.5 Generalized Newsvendor Problem: High-before-low arrival pattern

Consider the problem of selecting c to maximize

$$\Pi_n(c) = V_n(c) - kc$$

Where $V_n(c)$ is the expected revenue to the multi-fare RM problem.

Assume high-before-low arrival pattern. Then

$$V_n(c) = \sum_{j=1}^{n} p_j \mathbb{E}[D_j \wedge (c - D_{1:j-1})^+]$$

and

$$\Delta V_n(c) = \sum_{j=1}^{n} (p_j - p_{j+1}) Pr(D_{1:j} > c)$$

Where $D_{1:j} = \sum_{l=1}^{j} D_l, p_{n+1} = 0$

4.6 Heuristics

When there are two classes, we find y^*: $\max\{y \in \mathbb{Z} : P(D_1 \geq y) > r\}$

We try to use this form to simplify our computation,

EMSR (expected marginal seat revenue)

- **EMSR - a**

$$y_k^{j+1} = \max\{y : P(D_k \geq y) > \frac{p_{j+1}}{p_k}\}, k = j, j-1, ..., 1$$

$$y_j = \sum_{k=1}^{j} y_k^{j+1}$$

- **EMSR - b**

$$\bar{p}_j = \frac{\sum_{k=1}^{j} p_k \mathbb{E}[D_k]}{\sum_{k=1}^{j} \mathbb{E}[D_k]}$$

$$y_j = \max\{y : P(\sum_{k=1}^{j} D_k \geq y) > \frac{p_{j+1}}{\bar{p}_j}\}$$

4.7 Bounds on Optimal Expected Revenue

4.7.1 Upper Bound

$$\overline{V}(c|D) := \max\{\sum_{j=1}^{n} p_j x_j | \sum_{j=1}^{n} x_j \leq c, 0 \leq x_j \leq D_j, j = 1, ..., n\}$$

$$V_n^U(c) := \mathbb{E}[\overline{V}(c|D)]$$

$$= \sum_{j=1}^{n} (p_j - p_{j+1}) \sum_{k=1}^{c} Pr(D_{1:j} \geq k), \quad \text{(Set } p_{n+1} = 0)$$

$$\mathbb{E}[\overline{V}(c|D)] \leq \overline{V}(c|D) = \sum_{j=1}^{n} (p_j - p_{j+1}) \min\{\overline{D}_{1:j}, c\}$$

4.7.2 Lower Bound

Using zero protection level

$$V_n^L(c) = \sum_{j=1}^{n} p_j \mathbb{E}[\min\{D_k, (c - D_{j+1:n})^+\}]$$

$$= \sum_{j=1}^{n} (p_j - p_{j-1}) \mathbb{E}[\min\{D_{j:n}, c\}], \quad \text{(Set } p_0 = 0)$$

4.8 Dynamical Models

- $p_1 \geq p_2 \geq \cdots \geq p_n$.
- T periods.
- At most one arrival each period.
- λ_{jt}: probability of an arrival of class j in period t.

- M_t: set of offered classes.

4.8.1 Discrete Time

$$\begin{aligned}
V_t(x) &= \sum_{j\in M_t} \lambda_{jt} \max\{p_j + V_{t-1}(x-1), V_{t-1}(x)\} + (1 - \sum_{j\in M_t} \lambda_{jt})V_{t-1}(x) \\
&= V_{t-1}(x) + \sum_{j\in M_t} \lambda_{jt}[p_j - \Delta V_{t-1}(x)]^+ \\
&= V_{t-1}(x) + R_t(\Delta V_{t-1}(x))
\end{aligned}$$

Where we set $R_t(z) = \sum_{j\in M_t} \lambda_{jt}[p_j - z]^+$, $V_t(0) = 0$, $V_0(x) = 0, \forall x \geq 0$

4.8.2 Continuous Time: Poisson arrival

$$\frac{\partial V_t(x)}{\partial t} = R_t(\Delta V_t(x))$$

4.8.3 Optimal Policy: discrete time

Let

$$a(t,x) = \max\{j : p_j \geq \Delta V(t-1,x)\}$$

Optimal to accept all fares in the active set

$$A(t,x) = \{j \in M_t : j \leq a(t,x)\}$$

and reject the remaining fare classes

4.8.4 Structural Properties

Theorem 7 (1.18 of GT)**.**

- *$V_t(x)$ is increasing in t, x.*

- *$\Delta V_t(x)$ is increasing in t and decreasing in x.*

- *$a(t,x), A(t,x)$ is increasing in x.*

If $\lambda_{jt} \equiv \lambda_j > 0$, $M_t \equiv M = \{1,...,n\}$, then

- *$V_t(x)$ is strictly increasing and concave in t.*

- *$a(t,x), A(t,x)$ is decreasing in t.*

4.8.5 Discrete Case: Computation

$$V_t(x) = V_{t-1}(x) + \sum_{j \in M_t} \lambda_{jt}[p_j - \Delta V_{t-1}(x)]^+$$

$V_0(x) = 0$, then $V_1(x)$, then $\Delta V_1(x)$.

The complexity is $O(nCT)$ ($T \approx O(C)$)

5 Network Revenue Management with Independent Demands

5.1 Settings

- m resources with initial capacities $c = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix} \in \mathbb{Z}_+^m$

- Time from $T, T-1, T-2, ...$ to 0.

- ODF kj: Itineraries $k = 1, ..., K$; Possible fares for itinerary k, p_{kj}, $j \in \{1, ..., n_k\}$. (Every itinerary may have n_k kinds of prices).

- Demand arrives as compound Poisson arrival process with rate λ_{tkj} at time t for ODF kj.

- Resources utilized by itinerary k: $A_k = \begin{pmatrix} a_{1k} \\ \vdots \\ a_{mk} \end{pmatrix}$, $a_{ik} \in \{0, 1\}$ with $a_{ik} = 1$ if resource i is consumed by itinerary k.

- $V(t, x)$: the maximum total expected revenue that can be extracted when the remaining capacities are $x \in \mathbb{Z}_+^m$ and the remaining time is $t \in \mathbb{R}_+$.

- Decision: $u = \{u_{kj} : j = 1, ..., n_k, k = 1, ..., K\}$, $u_{kj} = \begin{cases} 1 & \text{accept a request for ODF } k_j \\ 0 & \text{others} \end{cases}$

- Feasible set of decisions: $u(x) = \{u_{kj} \in \{0, 1\} : A_k u_{kj} \leq x, j = 1, ..., n_k, k = 1, ..., K\}$

5.2 HJB Equation

Assume now that the state is (t, x) and consider a time increment δt that is small enough so that we can approximate the probability of an arrival of a request for fare j of itinerary k by $\lambda_{tkj}\delta k$.

$$\begin{aligned} V(t,x) =& \sum_{k=1}^{K}\sum_{j=1}^{n_k}\lambda_{tkj}\delta t \max_{u_{kj}\in\{0,1\}}\left[p_{kj}u_{kj} + V\left(t-\delta t, x - A_k u_{kj}\right)\right] \\ &+ \left\{1 - \sum_{k=1}^{K}\sum_{j=1}^{n_k}\lambda_{tkj}\delta t\right\} V(t-\delta t, x) + o(\delta t) \end{aligned}$$

where $o(\delta t)$ is a quantity that goes to zero faster than δt. Subtracting $V(t-\delta t, x)$ from both side of the equation, dividing by δt, and using the notation $\Delta_k V(t,x) = V(t,x) - V\left(t, x - A_k\right)$, we obtain the Hamilton-Jacobi-Bellman (HJB) equation

$$\frac{\partial V(t,x)}{\partial t} = \sum_{k=1}^{K}\sum_{j=1}^{n_k}\lambda_{tkj}\left[p_{kj} - \Delta_k V(t,x)\right]^{+}$$

with boundary conditions $V(t,0) = V(0,x) = 0$ for all $t \geq 0$ and all $x \geq 0$. Notice that term $[p_{kj} - \Delta V_k(t,x)]^{+}$is equivalent to the maximum of $p_{kj}u_{kj} + V(t, x- \ A_k u_{kj}) - V(t-\delta t, x)$ over $u_{kj} \in \{0,1\}$.

For any vector $z \geq 0$, Define

$$R_t(u,z) := \sum_{k=1}^{K}\sum_{j=1}^{n_k}\lambda_{tkj}\left[p_{kj} - z_k\right]u_{kj}$$

and

$$\mathcal{R}_t(z) := \max_u R_t(u,z) = \sum_{k=1}^{K}\sum_{j=1}^{n_k}\lambda_{tkj}\max_{u_{jk}\in\{0,1\}}\left[p_{kj} - z_k\right]u_{kj} = \sum_{k=1}^{K}\sum_{j=1}^{n_k}\lambda_{tkj}\left[p_{kj} - z_k\right]^{+}$$

Then

$$\frac{\partial V(t,x)}{\partial t} = \mathcal{R}_t(\Delta V(t,x)), \quad \Delta V(t,x) = \begin{pmatrix} \Delta_1 V(t,x) \\ \Delta_2 V(t,x) \\ \vdots \\ \Delta_K V(t,x) \end{pmatrix}$$

1. Let's aggregate ODF's into a single index.

2. $n = \sum_{k=1}^{K} n_k$

3. HJB equation:

$$\frac{\partial V(t,x)}{\partial t} = \mathcal{R}_t(\Delta V(t,x)) = \sum_{j \in M_t} \lambda_{tj} \left[p_j - \Delta_j V(t,x)\right]^+$$

   - $V(t,0) = V(0,x) = 0, \quad \forall t \geq 0, x \geq 0$
   - $M_t \subset \{1, ..., n\}$: offered set of fares at t
   - $\Delta_j V(t,x) = V(t,x) - V(t, x - A_j)$

4. Optimal Control:

$$u_j^*(t,x) = \begin{cases} 1 & \text{if } j \in M_t,\ A_j \leq x \text{ and } \underline{p_j \geq \Delta_j V(t,x)} \\ 0 & \text{others} \end{cases}$$

Compute exact $\Delta_j V(t,x)$ can be expensive, we can use heuristics to approx it by $\Delta_j \widetilde{V}(t,x)$

5.3 Upgrades

Let u_j be the set of products that can be used to fulfill a request for product j.

Customers are willing to take any products $k \in u_j$ at the price of product p_j.

$$\frac{\partial V(t,x)}{\partial t} = \sum_{j \in M_t} \lambda_j \max_{k \in u_j} \left[p_j - \Delta_k V(t,x)\right]^+ = \sum_{j \in M_t} \lambda_{tj} \left[p_j - \hat{\Delta}_j V(t,x)\right]^+$$

where $\hat{\Delta}_j V(t,x) = \min_{k \in u_j} \Delta_k V(t,x)$ (Use the least valuable product to fulfill p_j's request.)

5.4 Upsells

Selling j instead of k to get higher revenue, but may be rejected by customers.

- γ_{jk}: revenue obtained from selling product j and fulfilling it with product $k \in u_j$.

- π_{jk}: probability a customer will accept the upgrade from product j to product k.

$$\frac{\partial V(t,x)}{\partial t} = \sum_{j \in M_t} \lambda_j \max_{k \in u_j} \left[\pi_{jk}(r_{jk} - \hat{\Delta}_k V(t,x)) + (1 - \pi_{jk})(p_j - \hat{\Delta}_j V(t,x))\right]$$

5.5 Linear programming-based upper bound

The discrete maximum problem is

$$V(t,x) = \max_{u \in U(x)} \sum_{j=1}^{n} \lambda_{tj}[p_j u_j + V(t-1, x - u_j A_j)]$$

<u>**Deterministic Linear Program**</u>

Let D_j be the aggregate demand for ODF j over $[0,T]$.

Then D_j is Poisson with parameter $\Lambda_j = \int_0^T \lambda_{sj} ds$.

Define

$$\begin{aligned} \bar{V}(T,c) := \max \quad & \textstyle\sum_{j \in N} p_j y_j \\ \text{s.t.} \quad & \textstyle\sum_{j \in N} a_{ij} y_j \leq c_i \quad \forall i \in M \\ & 0 \leq y_j \leq \Lambda_j \quad \forall j \in N. \end{aligned}$$

$$\begin{aligned} \bar{V}(T,c|D) := \max \quad & \textstyle\sum_{j \in N} p_j y_j \\ \text{s.t.} \quad & \textstyle\sum_{j \in N} a_{ij} y_j \leq c_i \quad \forall i \in M \\ & 0 \leq y_j \leq D_j \quad \forall j \in N. \end{aligned}$$

Theorem 8 (2.2 of GT)**.**

$$V(T,C) \leq \mathbb{E}[\bar{V}(T,c|D)] \leq \bar{V}(T,c)$$

$\bar{V}(T,c)$ is the revenue of expected demand, $\mathbb{E}[\bar{V}(T,c|D)]$ is probability combination that is concave in D, so $\mathbb{E}[\bar{V}(T,c|D)] \leq \bar{V}(T,c)$. And $V(T,C)$'s decision is feasible in $\mathbb{E}[\bar{V}(T,c|D)]$, so $V(T,C) \leq \mathbb{E}[\bar{V}(T,c|D)]$.

<u>Dual formulation</u> of $\bar{V}(T,c)$

$$\begin{aligned} \bar{V}(T,c) := \min \quad & \textstyle\sum_{i \in M} c_i z_i + \sum_{j \in N} \Lambda_j \beta_j \\ \text{s.t.} \quad & \textstyle\sum_{i \in M} a_{ij} z_i + \beta_j \geq p_j \quad \forall j \in N \\ & z_i, \beta_j \geq 0 \quad \forall i \in M, \forall j \in N. \end{aligned}$$

We can simplify the formulation. Since $\beta_j \geq p_j - \sum_{i \in M} a_{ij} z_i$, $\beta_j \geq 0$ and dual is a minimization problem, we can rewrite $\beta_j = [p_j - \sum_{i \in M} a_{ij} z_i]^+$. Then,

$$\sum_{j \in N} \Lambda_j \beta_j = \sum_{j \in N} \Lambda_j [p_j - \sum_{i \in M} a_{ij} z_i]^+ = \int_0^T \mathcal{R}_t(A^T z) dt$$

so,

$$\bar{V}(T,c) = \min_{z \geq 0} \int_0^T \mathcal{R}_t(A^T z) dt + c^T z$$

The optimal solution z_i^* gives an estimation of the marginal value of the i^{th} resource. The approximation of $\Delta_j V(T,c)$ is $\sum_{i \in M} a_{ij} z_i^*$

Bid-price Heuristic

Accept ODF_j if and only if

$$p_j \geq \sum_{i \in M} a_{ij} z_i^* \text{ and } A_j \leq x$$

Probabilistic Admission Control (PAC) Heuristic

Accept ODF_j with probability $\frac{y_j^*}{\Lambda_j}$ whenever $A_j \leq x$.

Bid-price heuristic is not in general asymptotically optimal.

PAC heuristic is asymptotically optimal.

Theorem 9. *Let $\Pi^b(T,c)$ be the total expected revenue from PAC heuristic and $V^b(T,c)$ be the optimal total expected revenue corresponding to circumstance $b \geq 1$ with capacity bc and $b\lambda_{jt}$. Then*

$$\lim_{b \to \infty} \frac{\Pi^b(T,c)}{V^b(T,c)} = 1$$

5.6 Dynamic Programming Decomposition (DPD)

In this section, we describe two possible approaches for approximating the value functions $V(t,\cdot)$ for the discrete-time formulation

$$V(t,x) = V(t-1,x) + \mathcal{R}_t(\Delta V(t-1,x))$$

Consider the aggregated single index formulation

$$V(t,x) = \max_{u \in U(x)} \sum_{j=1}^{n} \lambda_{tj}[p_j u_j + V(t-1, x - u_j A_j)]$$

with $V(t,0) = V(0,x) = 0$ and $\sum_{j=1}^{n} \lambda_{tj} = 1, \lambda_{tj} \geq 0$. (scale can be standardized)

5.6.1 Deterministic Linear Program

The former DLP we use

$$\begin{aligned} \bar{V}(T,c) := \max \quad & \textstyle\sum_{j \in N} p_j y_j \\ \text{s.t.} \quad & \textstyle\sum_{j \in N} a_{ij} y_j \leq c_i \quad \forall i \in M \\ & 0 \leq y_j \leq \Lambda_j \quad \forall j \in N. \end{aligned}$$

Its dual optimal value is $(z_1^*, z_2^*, ..., z_m^*)$. We choose an arbitrary resource i and relax the first set of constraints for all of the resources except for resource i by associating the dual multipliers $(z_1^*, z_2^*, ..., z_m^*)$ with them.

We relax the first constraints $\sum_{j\in N} a_{ij}y_j \le c_i \quad \forall i \in M$, which won't change the objective value,

$$\begin{aligned}\max \sum_{j\in N} p_j y_j &= \sum_{j\in N} p_j y_j - \sum_{k\neq i}[\sum_{j\in N} a_{kj}y_j - c_k]z_k \\ &= \sum_{j\in N}[p_j - \sum_{k\neq i} a_{kj}z_k^*]y_j + \sum_{k\neq i} z_k^* c_k\end{aligned}$$

The new DLP is

$$\begin{aligned}\bar{V}(T,c) := \max \quad & \textstyle\sum_{j\in N}[p_j - \sum_{k\neq i} a_{kj}z_k^*]y_j + \sum_{k\neq i} z_k^* c_k \\ \text{s.t.} \quad & \textstyle\sum_{j\in N} a_{ij}y_j \le c_i \quad \forall i \in M \\ & 0 \le y_j \le \Lambda_j \quad \forall j \in N.\end{aligned}$$

We can prove the optimal y^* and optimal objective values are the same.

Claim 2. *The optimal values y_j^* and optimal objective values of these two DLP are the same.*

(This claim can help prove the upperbound).

$$V(t,x) = \max_{u\in U(x)} \sum_{j=1}^{n} \lambda_{tj}[p_j u_j + V(t-1, x - u_j A_j)]$$

We consider the optimal total expected revenue in the single-resource revenue management problem for resource i, the corresponding price of ODF_j should be $p_j - \sum_{k\neq i} a_{kj}z_k^*$. Then the formulation is

$$v_i(t,x_i) = \max_{u\in U_i(x_i)} \sum_{j\in N} \lambda_{tj} \left\{ [p_j - \sum_{k\neq i} a_{kj}z_k^*]u_j + v_i(t-1, x_i - u_j a_{ij}) \right\}$$

We can prove that

- $v_i(T,c_i) \le \bar{V}(T,c) - \sum_{k\neq i} z_k^* c_k$
- Theorem 2.11 of GT

$$V(t,x) \le \min_{i\in M}\{v_i(t,x_i) + \sum_{k\neq i} z_k^* x_k\}$$

5.6.2 Lagrangian Relaxation

$$\begin{aligned}V(t,x) = \max \quad & \textstyle\sum_{j=1}^{n} \lambda_{tj}[p_j u_j + V(t-1, x - u_j A_j)] \\ \text{s.t.} \quad & u_j A_j \le x \\ & u_j \in \{0,1\} \quad \forall \in N\end{aligned}$$

To demonstrate the Lagrangian relaxation strategy, we use decision variables $\{w_{ij} : i \in M, j \in N\}$ in the dynamic programming formulation of the network revenue management problem, where $w_{ij} = 1$ if we make ODF_j available for purchase on flight leg i, otherwise $w_{ij} = 0$.

$$
\begin{aligned}
V(t,x) = \max \quad & \textstyle\sum_{j\in N} \lambda_{tj} \left\{ p_j w_{\psi j} + V\left(t-1, x - \sum_{i\in M} w_{ij} a_{ij} e_i\right)\right\} \\
\text{s.t.} \quad & a_{ij} w_{ij} \le x_i \\
& w_{ij} = w_{\psi j} \\
& w_{ij} \in \{0,1\}, w_{\psi j} \in \{0,1\} \quad \forall i \in M, j \in N
\end{aligned}
$$

We can relex the second set of constraints by adding Lagrange mutipliers $\{\alpha_{tij} : i \in M, j \in N\}$. Relaxed dynamic program:

$$
\begin{aligned}
V^\alpha(t,x) = \max \quad & \textstyle\sum_{j\in N} \lambda_{tj} \left\{ \sum_{i\in M} \alpha_{tij} w_{ij} + \left[p_j - \sum_{i\in M} \alpha_{tij}\right] w_{\psi j} + V^\alpha\left(t-1, x - \sum_{i\in M} w_{ij} a_{ij} e_i\right)\right\} \\
\text{s.t.} \quad & a_{ij} w_{ij} \le x_i \\
& w_{ij} \in \{0,1\}, w_{\psi j} \in \{0,1\} \quad \forall i \in, j \in N
\end{aligned}
$$

Theorem 10 (2.13 of GT). *Assume that the value functions $\{v_i^\alpha(t,\cdot) : t = 1,\ldots,T\}$ are computed through the dynamic program*

$$
v_i^\alpha(t, x_i) = \max_{w_i \in \mathcal{U}_i(x_i)} \left\{ \sum_{j\in N} \lambda_{tj} \left\{ \alpha_{tij} w_{ij} + v_i^\alpha(t-1, x_i - w_{ij} a_{ij}) \right\} \right\}
$$

Then

$$
V^\alpha(t,x) = \sum_{i\in M} v_i^\alpha(t, x_i) + \sum_{\tau=1}^{t} \sum_{j\in N} \lambda_{\tau j} \left[p_j - \sum_{i\in M} \alpha_{\tau ij} \right]^+
$$

Theorem 11 (2.14 of GT). *For any set of Lagrange multipliers α, we have*

$$
V(t,x) \le V^\alpha(t,x) \quad \forall x \in \mathbb{Z}_+^m, t = 1, ..., t
$$

The tightest possible upper bound, we can solve the problem

$$
\min_{\alpha \in \mathbb{R}^{Tmn}} V^\alpha(T,c)
$$

Lemma 6 (2.15 of GT). *$V^\alpha(t,x)$ is a convex function of α for any $t = 1, ..., T$ and $x \in \mathbb{Z}_+^m$.*

Then compute $\min V^\alpha(T,c)$ can be easier.

6 Overbooking

7 Choice Modeling

Dependent Demand Models

With Choice Models, we can make the demand of offered products to the consumers dependent on the menu of the options.

7.1 Basics

- $N = \{1, \dots, n\}$: the potential products,
- $S \subseteq N$: a subset of products that is offered (referred to as assortment).
- $\pi_j(S)$: probability that a consumer will purchase product j when S is offered.
- $\Pi(S) = \sum_{j \in S} \pi_j(S)$: probability of a sale when S is offered.
- $\pi_0(S) = 1 - \Pi(S)$: probability that the consumer selects outside alternative.

7.1.1 Assortment Optimization Problem (AOP)

Let

- p_j : The price of product $j \in N$.
- z_j : The unit cost of product $j \in N$.
- $p := (p_1, p_2, \dots, p_n)$: The price vector
- $z := (z_1, z_2, \dots, z_n)$: The unit cost vector.

If we offer S, the expected profit from offering subset S is given by:

$$R(S, z) := \sum_{j \in S} (p_j - z_j)\, \pi_j(S)$$

AOP aims at finding an assortment S which maximizes $R(S, z)$:

$$\mathcal{R}(z) := \max_{S \subseteq N} R(S, z)$$

7.1.2 Joint Assortment and Pricing Optimization Problem (JAPOP)

If the Price is also a Decision Variable, we should solve the following:

$$\mathcal{R}(z) := \max_{p \geq 0, S \subseteq N} \sum_{j \in S} (p_j - z_j)\, \pi_j(S, p)$$

For many choice models, if $p_j = \infty$ then, $\pi_j(S, p) = 0$. Thus optimizing over prices could implicitly select an assortment.

7.2 Maximum Utility Models (MUM)

Suppose a customer has a Full Ordering of the preferences among the products.

- Assigns cardinal utilities, say $\{u_i : i \in N\}$, to products and ranks them based on utilities.
- The available product with the highest utility is selected with probability 1.

To introduce randomness into the MUM:

1. **Random Utility Models (RUM):** Add a random noise component to the utilities of the products:

   $$U_i = u_i + \varepsilon_i, i \in N.$$

   where the ε_i's are mean zero, possibly dependent random variables.

2. Assume a distribution of consumer types, each with a certain preference ordering. So, product $i \in S$ is selected by types of consumers for whom i is the highest ranked product in S.

7.3 Basic Attraction Model (BAM)

7.3.1 Settings

- In BAM each product has an attraction value $v_j > 0$, capturing attractiveness of product j. $v_0 > 0$ represents the attractiveness of the no-purchase alternative.
- The probability of purchasing each item is proportional to the attraction value:

$$\pi_j(S) = \frac{v_j}{v_0 + \sum_{i \in S} v_i} \quad \forall j \in S$$

Denote, for any $S \subseteq T$:

- $\pi_S(T) := \sum_{j\in S}\pi_j(T) = \frac{\sum_{j\in S} v_j}{v_0+\sum_{i\in T} v_i} = \frac{\sum_{j\in S} v_j}{\sum_{i\in T^+} v_i}$, the probability that a consumer selects a product in S when the set T is offered.

- $\pi_{S^+}(T) := \pi_S(T) + \pi_0(T) = \frac{v_0+\sum_{j\in S} v_j}{v_0+\sum_{i\in T} v_i} = \frac{\sum_{j\in S^+} v_j}{\sum_{i\in T^+} v_i}$, including the no-purchase alternative.

7.3.2 Luce Axioms for BAM

A discrete choice model satisfies the axioms iff it is of the BAM form:

- Axiom 1 :If $\pi_i(\{i\}) \in (0,1)$ for all $i \in T$, then for any $Q \subseteq S_+$, $S \subseteq T$:

$$\pi_Q(T) = \pi_Q(S)\pi_{S_+}(T)$$

Proof. $\frac{\sum_{i\in Q} v_i}{\sum_{i\in T^+} v_i} = \frac{\sum_{i\in Q} v_i}{\sum_{i\in S^+} v_i}\frac{\sum_{i\in S^+} v_i}{\sum_{i\in T^+} v_i}$ □

- Axiom 2: If $v_i = 0$ ($\Leftrightarrow \pi_i(\{i\}) = 0$) for some $i \in T$, then for any $S \in T$ such that $i \in S$:

$$\pi_S(T) = \pi_{S\setminus\{i\}}(T\setminus\{i\})$$

7.3.3 Mutinomial Logit Choice Model (MNL): BAM with RUM

MNL is a special case of BAM, and also has a RUM based justification as well. Specifically:

- If $U_j = u_j + \epsilon_j$ for $j \in N_+$denotes the random utility of product j,

- ϵ_j has a Gumbel distribution with zero mean and same scale parameters ϕ for all j.

- The MNL's probability structure:

$$\pi_j(S) = \frac{e^{\phi u_j}}{1+\sum_{k\in S} e^{\phi u_k}} \forall j \in S$$

<u>**Gumbel Random Variable:**</u> Gumbel with location and scale parameters ν and ϕ

- CDF: $F(x : \nu, \phi) = \exp(-\exp(-\phi(x-\nu)))$

- Model: v

- Median: $v - \ln(\ln 2)/\phi$

- Mean: $v + \gamma/\phi$, where γ is the Euler constant

- Variance: $\pi^2/6\phi^2$

7.3.4 Mixture of BAMs: Different type consumers

Denote:

- G be the set of consumer types,
- v_j^g : the attraction value of product j for consumer of type g.
- α^g : The probability that a consumer belongs to type g.

The Mixture of BAMs suggests the following Probability Structure:

$$\pi_j(S) = \sum_{g \in G} \alpha^g \frac{v_j^g}{v_0^g + \sum_{i \in S} v_i^g}$$

Any RUM-based choice model can be approximated to any degree of accuracy by a mixture of BAM's.

7.4 Generalized Attraction Model (GAM)

7.4.1 Motivation

The attraction value v_0 of the no-purchase option does not depend on the subset of offered products in BAM.

The probability of leaving the store is higher than the one suggested by BAM when eliminating an item.

The BAM ignores the possibility that the consumer may look for the products that are not offered in elsewhere or at a later time.

7.4.2 Model

Assume the **attraction** values $\{v_j : j \in N\}$ for each product $j \in N$

Shadow attraction values $\{w_j : j \in N\}$ with $w_j \in [0, v_j]$ for all $j \in N$, which changes the attractiveness of the no-purchase alternative:

$$v_0 + \sum_{k \in S^c} w_k$$

Note that <u>in GAM the attractiveness of the no-purchase alternative is dependent on the offered set S.</u>

The probability structure according to the GAM:

$$\pi_j(S) = \frac{v_j}{v_0 + \sum_{k \in S^c} w_k + \sum_{k \in S} v_k}$$

7.4.3 Parsimonious GAM: $w_j = \theta v_j$

Parsimonious GAM: $w_j = \theta v_j$ for all j for some $\theta \in [0, 1]$.

- Independent Demand Model, IDM: If $\theta = 1$, the choice probability is independent of S :

$$\pi_j(S) = v_j$$

  where $v_0 + \sum_{k \in N} v_k = 1$ is normalized.

- BAM: $\theta = 0$

7.4.4 Independence of Irrelevant Alternatives (IIA)

BAM and GAM satisfy the IIA property: adding a new product to an offered subset decrease the purchase probability of all offered products by the same relative amount, i.e., $\forall S \subseteq N$ and $i, j \in S, k \notin S$,

$$\frac{\pi_i(S)}{\pi_i(S \cup \{k\})} = \frac{\pi_j(S)}{\pi_j(S \cup \{k\})}$$

7.5 Nested Logit Model (to avoid IIA)

Some products are dependent, we shouldn't count the products in the same nest more than once.
Nested Logit (NL) model, is proposed to avoid IIA. In this model:
The products are organized into nests such that the products in the **same nest are regarded as closer** substitutes of each other relative to the products in different nests. (Buses in the same nest)
Under the NL model, the selection process of a consumer proceeds in **two stages**:

- First, the consumer selects either **one of the nests** or decides to leave without making a purchase.
- Second, if the consumer selects one of the nests, then the consumer chooses one of the products offered **in this nest**.

7.5.1 Notation

Denote:

- $M = \{1, \ldots, m\}$: the set of nests.
- N : set of all items.

- $S_i \subseteq N$: the offered set in nest i.

- $(S_1, \ldots, S_m)$: set of products over all nests.

- v_{ij} : attraction value for product j in nest i.

- $V_i(S_i) := \sum_{j \in S_i} v_{ij}$.

- γ_i : measure how easily the products in nest i substitute for each other

7.5.2 Model

- Given that a consumer has chosen nest i, the probability of selecting $j \in S_i$ is a BAM:

$$q_{j|i}(S_i) := \frac{v_{ij}}{V_i(S_i)}$$

- *If the items of each nest are more dissimilar, then the probability of choosing nest should increase as it absorbs more kind of consumers.* By this a consumer chooses nest i with probability:

$$Q_i(S_1, \ldots, S_m) := \frac{V_i(S_i)^{\gamma_i}}{v_{i0} + \sum_{\ell \in M} V_\ell(S_\ell)^{\gamma_\ell}}$$

- The selection probability of product j in nest i is given by:

$$Q_i(S_1, \ldots, S_m)\, q_{j|i}(S_i) := \frac{V_i(S_i)^{\gamma_i}}{v_0 + \sum_{\ell \in M} V_\ell(S_\ell)^{\gamma_\ell}} \frac{v_{ij}}{V_i(S_i)}$$

7.5.3 RUM-based justification

- $U_{ij} = u_{ij} + \epsilon_{ij}$

- $\epsilon = \{\epsilon_{ij} : i \in M, j \in N\}$: generalized extreme value distribution

$$F(x; \gamma) = \exp\left(-\sum_{i \in M}\left(\sum_{j \in N} \exp\left(-x_{ij}/\gamma_i\right)\right)^{\gamma_i}\right)$$

- The marginal distribution of ϵ_{ij} is Gumbel

- For distinct $i, \ell \in M, \epsilon_{ij}$ and $\epsilon_{\ell k}$ are independent

- For a given nest i, ϵ_{ij} and ϵ_{ik} are positively correlated, and $1 - \gamma_i$ measures the degree of correlation

The above assumptions lead to NL model with

$$v_{ij} = \exp\left(u_{ij}/\gamma_i\right)$$

7.6 Exponomial Choice (EC) Model

Gumbel distribution makes sense in many cases, particularly those where consumers have limited information regarding a product's value.

There are many cases, consumers are well-informed about products and prices. Example: Get information regarding particular cars before purchasing from a dealer.

One might expect the distribution of Willingness to pay (WTP) for a given product across consumers to be negatively skewed.

The utility that a random consumer has for choice i in the EC model is the linear function.

$$U_i = u_i - z_i$$

u_i is the ideal utility for choice i.

z_i is independent exponential random variables with λ rate to capture consumer heterogeneity.

The only difference between EC and MUM is the distribution of z_i.

The probability that a consumer prefers choice i is

$$\begin{aligned} Q(i) &= \operatorname{Prob}\left\{u_i - z_i \geq u_j - z_j \forall j, j \neq i\right\} \\ &= \operatorname{Prob}\left\{z_j \geq u_j - u_i + z_i \forall j, j \neq i\right\} \\ &= \int_0^{\infty} \prod_{j \neq i} \left[1 - F\left(u_j - u_i + z\right)\right] f(z) \mathrm{d}z \end{aligned}$$

where $f(z) = \lambda e^{-\lambda z}$ and $F(z) = 1 - e^{-\lambda z}$ for $z \geq 0$

Theorem 12 (Alptekinoğlu and Semple, 2016)**.** *Out of m choices with utilities $U(i) = u_i - z_i$, where $u_1 \leq \cdots \leq u_m$ and z_i follow independent exponential distributions with rate λ, the probability that the consumer prefers choice $i \in \{1, \ldots, m\}$, i.e., considers it as utility-maximizing, is*

$$Q(i) = \frac{\exp\left[-\lambda \sum_{j=i}^{m}\left(u_j - u_i\right)\right]}{m - i + 1} - \sum_{k=1}^{i-1} \frac{\exp\left[-\lambda \sum_{j=k}^{m}\left(u_j - u_k\right)\right]}{(m-k)(m-k+1)}$$

Remark: The parameters of the exponomial model are easy to estimate.

7.7 Random Consideration Set Model (RCS)

- An ordering for the preferences of the products:

$$0 \prec 1 \prec 2 \prec \cdots \prec n$$

- The consumer forms a consideration set $C(S)$, by independently including each product with probability λ_i.

- For a consumer to purchase product i :

1. Product **i** should be in her consideration set,

2. all offered products that are preferred to product i should NOT be in her consideration set.

Thus:

$$\pi_i(S) = \lambda_i \prod_{i \prec j, j \in S} (1 - \lambda_j) \quad \forall i \in S$$

- MUM can be recovered when we set $\lambda_\mathbf{j} = 1, \forall j \in N$. - Assume consumers pay no attention to the products below the no-purchase alternative and $\lambda_0 = 1$. Then

$$\pi_0(S) = \prod_{j \in S} (1 - \lambda_j)$$

- The RCS model is a special case of the Markov Chain choice model.

7.8 Markov Chain (MC) Choice Model

First, note that this is an approximation for the underlying choice models. Thus, the justification of the model is based on the fact that it can approximate other models that are Random Utility Based.

In general: substitution behavior is modeled as a sequence of state transitions of a Markov chain.

In particular:

(1) There is a state for each product including the no-purchase alternative,

(2) A consumer arrives in the state corresponding to his most preferable product,

- If that product is **available**, she purchases that product and exits.
- If that product is **not available**, she transitions to other product states according to the transition probabilities of the Markov chain.

- Markov chain choice model provides a good approximation to any random utility discrete choice models under mild assumptions.

- If the choice probabilities used to compute the Markov chain model parameters is generated from an **underlying GAM**, then the choice probability computed by the Markov chain model coincides with the probability given by the GAM model for all products and all assortments. Thus: **Exact Approximation**.

- Since **MNL** is a special case of GAM, Markov Chain gives an **Exact Approximation** of this model as well.

7.8.1 Notation

- $\mathcal{N} = \{1, 2, \ldots, n\}$: Universe of products,

- 0: No-Purchase Product,

- $S \subseteq \mathcal{N}$: offered set,

- $S_+ := S \cup \{0\}, \mathcal{N}_+ = \mathcal{N} \cup \{0\}$,

- $\pi(j, S)$: the choice probability of item $j \in S_+$,

- The set of states: $\mathcal{N}_+$,

7.8.2 How it works

(1) For any $i \in \mathcal{N}_+$, a consumer arrives in state i with probability $\lambda_i = \pi(i, \mathcal{N})$.

(2) Selects i if it is available.

(3) o.w. It transitions to another state $j \neq i$, with probability ρ_{ij}.

(4) After transitioning to state j, the consumer behaves exactly like a consumer whose most preferable product is j.

Note that: Substitution arising from a preference list by a Markovian transition model where transitions out of state i do not depend on the previous transitions.

The model is completely specified by: 1 Initial Arrival Probabilities λ_i for all states $i \in \mathcal{N}_+$,

2 The Transition Probabilities: ρ_{ij}, for all $i \in \mathcal{N}$ and $j \in \mathcal{N}_+$.

Note that one of the states is the no-purchase alternative i.e. $0 \in \mathcal{N}_+$.

7.8.3 Compute Parameters

Given the choice probabilities for the following $(n+1)$ assortments:

$$\mathcal{L} = \{\mathcal{N}, \mathcal{N} \backslash \{i\} \mid i = 1, \ldots, n\}$$

Then the parameters are calculated as follows:

$$\lambda_i = \pi(i, \mathcal{N})$$

$$\rho_{ij} = \begin{cases} 1, & \text{if } i = 0, j = 0 \\ \frac{\pi(j, \mathcal{N}\setminus\{i\}) - \pi(j, \mathcal{N})}{\pi(i, \mathcal{N})} & \text{if } i \in \mathcal{N}, j \in \mathcal{N}_+, i \neq j \\ 0, & \text{otherwise} \end{cases}$$

The above are pretty intuitive.

- Denote by $\Phi_j(S)$ the probability that a consumer considers product $j \notin S$ during the course of her choice process but does not purchase because $j \notin S$.

- By definition, $\Phi_j(S) = 0$ for all $j \in S$.

- The quantities $\pi_j(S)$ for $j \in S$ and $P_{ij}(S)$ for $j \in \bar{S}$, where $\bar{S} = N \setminus S$ are related by the system of equations

$$\pi_j(S) = \lambda_j + \sum_{i \in \bar{S}} \Phi_i(S) \rho_{ij} \forall j \in S$$

$$\Phi_j(S) = \lambda_j + \sum_{i \in \bar{S}} \Phi_i(S) \rho_{ij} \forall j \in \bar{S}$$

7.9 Contextual MNL Model

Definition 11 (Context Effects)**.** *Context effects are referred to the changes in the perception about preferability of a given alternative that depend on the presence or absence of other options beside the given alternative. (Note NOT talking about demand.)*

- **The Attraction Effect-Decoy Effect:** An asymmetrically dominated item is the one which is dominated by just one of the items of the choice set and not the other.

	Option A	Option B
Price	\$400	\$300
Storage	30 GB	20 GB

Table: MP3 Player Choice Set 1.

	Option A (Target)	Option B	Option C-Decoy
Price	\$400	\$300	\$450
Storage	30 GB	20 GB	25 GB

Table: MP3 Player Choice Set 2.

Figure 4: Attraction Effect-Decoy Effect

- **The compromise Effect:** Middle option gets more share. In particular, in presence of this effect, when adding an extreme option (very high-level or basic product), the share of the middle-level options will increase.

  For example, a car-shopper who is given three options:

  1. the low-priced basic model with no extras,
  2. a high-priced fully loaded model with all the extras,
  3. and a mid-priced model with some extras,

  will most likely choose the middle option.

- **The Similarity Effect:** Adding an item will hurt the market share and perceived preference of the options which are similar to it more than the dissimilar options.

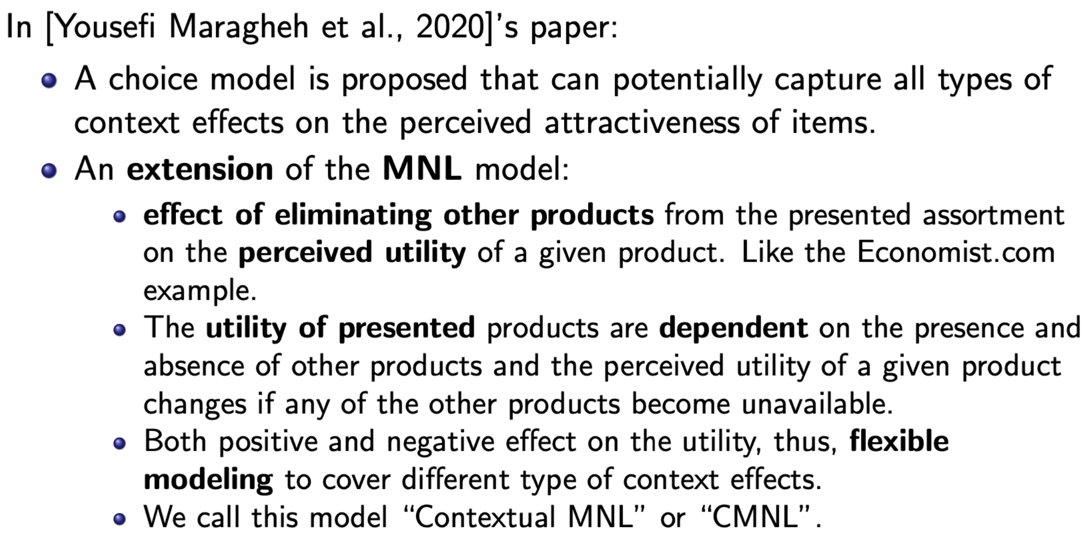


Figure 5: CMNL

7.10 Rank List-based Choice Model

7.10.1 Motivation

- Using historical sales data to predict the revenues or sales from offering a particular assortment of products to consumers.

- Fitting the ”right” parametric choice model to data are prone to overfitting and underfitting.
- The risk of model misspecification and leading to inaccuracies in decision making.
- Parametric models are prone to overfitting and underfitting.
  (a) Too simple model may make practically unreasonable assumptions.
  (b) Too complex model can lead to worse performance.
- Need to make a data-driven, nonparametric choice model.

7.10.2 Notations

- Consider N products, $\mathcal{N} = \{0, 1, 2, \ldots, N-1\}$.
- Assume that the $0th$ product is no-purchase.
- A consumer is associated with a permutation (or ranking) σ a of the product in $\mathcal{N}$.
- The consumer σ prefers product i to product j if and only if $\sigma(i) < \sigma(j)$.
- In the practice, a consumer σ purchases

$$\arg\min_{i \in N} \sigma(i).$$

- We then have $N!$ consumer types.

Figure 6: Notations

7.10.3 Model

- The set of all possible consumer types is S_N.
- The probability of consumer type σ arrival is $\lambda(\sigma)$.
- The set of consumer who preferred product j when set $\mathcal{M}$ has been offered is

$$\mathcal{L}_j(\mathcal{M}) = \{\sigma \in S_N : \sigma(j) < \sigma(i), \forall i \in \mathcal{M}, i \neq j\}.$$

- The probability of selecting product j from set $\mathcal{M}$ is

$$\mathbb{P}(j|\mathcal{M}) = \sum_{\sigma \in \mathcal{L}_j(\mathcal{M})} \lambda(\sigma) \triangleq \lambda^j(\mathcal{M}).$$

Figure 7: Model

- Form of observed sales transactions from a set of displayed assortment.
- If observed assortments are $\mathcal{M}_1, \ldots, \mathcal{M}_L$.
    - y_{il} is a fraction of customers purchased product i in assortment l.
    - Let $A \in \{0,1\}^{NL \times N!}$ and $A(\sigma)_{il} = 1$ iff $i \in \mathcal{M}_l$ and $\sigma(i) < \sigma(j)$ $\forall j \in \mathcal{M}_l \cup \{0\}$.
- If we have access data that provide us with the information about the fraction of customers that prefer product i to product j, for all pairs i and j (**Comparison data**).
    - y_{ij} is a fraction of customers purchased product i to product j.
    - Let $A \in \{0,1\}^{N(N-1) \times N!}$ and $A(\sigma)_{ij} = 1$ iff $\sigma(i) < \sigma(j)$.
- We then have $y = A\lambda$.

Figure 8: Data

- The goal is to use just these data to make predictions about the revenue rate (i.e., the expected revenues garnered from a random consumer) for some given assortment, say $\mathcal{M}$, that has never been encountered in past.
- Let p_j be the price of product j. Then,

$$
\begin{aligned}
&\min_{\lambda} \sum_{j\in\mathcal{M}} p_j \lambda^j(\mathcal{M}) \\
&\text{s.t. } A\lambda = y, \\
&\qquad \mathbf{1}^T\lambda = 1, \\
&\qquad \lambda \geq 0.
\end{aligned}
\tag{10}
$$

- Problem (10) offers the worst-case expected revenue possible for the unseen assortment.

Figure 9: Robust Approach

- Demonstrate that the robust approach can capture various underlying parametric structures and produce good revenue prediction.
- steps:
  1. Pick a structural model like the MNL model. This may be a model derived from real-world data or a purely synthetic model
  2. Use this structural model to simulate sales for a set of test assortments.
  3. Use this transaction data to estimate marginal information y, and use y to implement the robust approach.
  4. Use the implemented robust approach to predict revenues for a distinct set of assortments, and compare the predictions to the true revenues computed using the ground-truth structural model chosen for benchmarking in step 1.

Figure 10: Testing the Performance of Robust Approach

- Fit Amazon.com DVD sales data to the MNL model.
- The CNL (a type of the NL model) and MMNL models were obtained through perturbations of the MNL model.
- Generate comparison data for each parametric model.

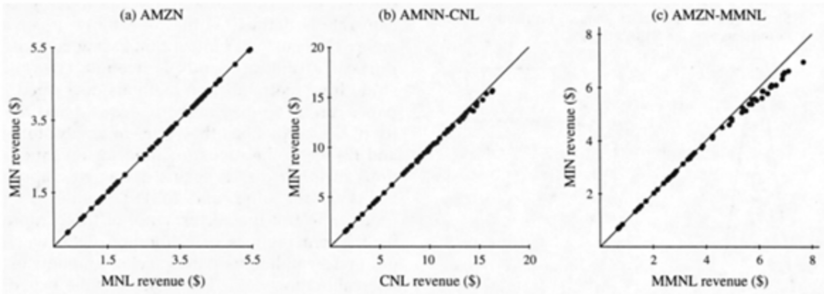


Figure 11: Computational Study

7.10.4 Testing the Performance of Robust Approach

7.10.5 Limitations

- $N!$ customer types.
- How to collect the best features from data that rationalize the rank lists.
- The objective is to get the worst-case revenue prediction for a given assortment, not maximizing the revenue.
- Disregarding the other input like customers' features and price.
- In general, the model they provided is for the prediction. How to use it to make decisions?

Figure 12: Limitations

7.11 Threshold Utility Model

buy multiple products

8 (Joint Pricing and) Assortment Optimization

8.1 Assortment Optimization Problem

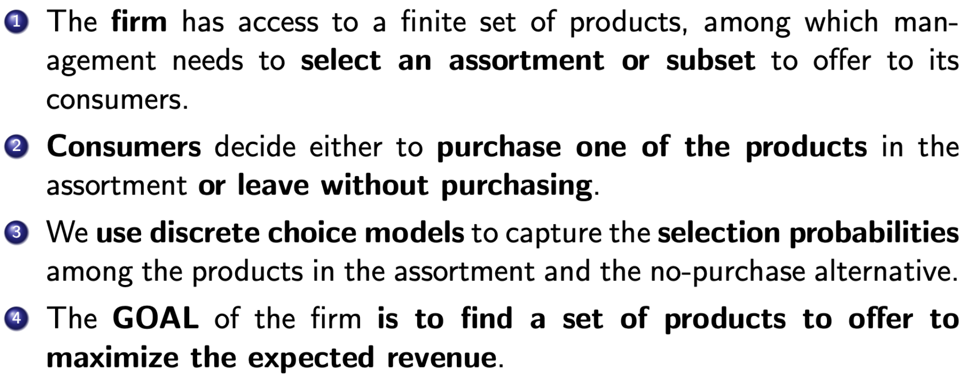


Figure 13: What is Assortment?

The fundamental Trade off in the Assortment optimization problem (AOP):

- Broad assortments: demand cannibalization and spoilage,
- Narrow assortments: disappointed consumers that may walk away without purchasing.

8.1.1 Notations

If we offer S, the expected profit from offering subset S is given by:

$$R\left(S, A'z\right) := \sum_{j \in S} \left(p_j - A'_j z\right) \pi_j(S)$$

AOP is finding the S which maximizes $R\left(S, A'z\right)$:

$$\mathcal{R}\left(A'z\right) := \max_{S \subseteq N} R\left(S, A'z\right)$$

For brevity of notation, we will make the transformation $p_j \leftarrow p_j - A'_j z$ for all $j \in N$. Also:

- $N = \{1, \ldots, n\}$: the **potential** products,
- $M = \{1, \ldots, m\}$: the set of **resources** utilized by products.
- A: **resource matrix**. $A_{ij} = 1$ if product j uses one unit of resource i and $A_{ij} = 0$ otherwise.
- $z \in \mathcal{R}_+^m$: a vector of **marginal cost** of the resources.
- $A'z$: the **marginal cost of the products**.
- p_j: **Price** for product j
- $S \subseteq N$: a **subset** of products that is offered (**assortment**)
- $\pi_j(S)$: purchase probability of product j when S is offered.
- $\Pi(S) = \sum_{j \in S} \pi_j(S)$: purchase probability
- $\pi_0(S) = 1 - \Pi(S)$: no-purchase probability

Figure 14: Notations

- $R(S)$ as a short-hand for $R(S, 0)$,
- $\mathcal{R}^*$ as a short-hand for $\mathcal{R}(0)$.

8.2 Pricing and Assortment Planning under the MNL model and Variants

8.2.1 MNL Model

Given an assortment $S \subseteq N$, the random utility U_i of product $i \in S$ is

$$U_i = \underbrace{u_i\left(p_i\right)}_{\text{deterministic part}} + \underbrace{\xi_i}_{\text{random part}}.$$

Define the attraction factor of product $i \in S$ as

$$v_i\left(p_i\right) = e^{u_i\left(p_i\right)}.$$

When $\xi_1, \ldots, \xi_n$ are i.i.d. random variables with a Gumbel distribution, through the utility maximization, a consumer chooses product $i \in S$ w.p.

$$q_i = \frac{v_i\left(p_i\right)}{1 + \sum_{j \in S} v_j\left(p_i\right)}, \quad \forall i \in S.$$

and the no-purchase probability is

$$q_0 = \frac{1}{1 + \sum_{j \in S} v_j\left(p_i\right)}$$

8.2.2 Joint Pricing and Assortment Planning under MNL Model

The profit of a given assortment S and given prices $\mathbf{p} = \left(p_1, \ldots, p_{|S|}\right)$ is

$$R(\mathbf{p}, S) = \sum_{i \in S} \left(p_i - c_i\right) q_i = \frac{\sum_{i \in S} \left(p_i - c_i\right) v_i \left(p_i\right)}{1 + \sum_{j \in S} v_j \left(p_j\right)}$$

(Costs $\mathbf{c} = (c_1, \ldots, c_n)$ are given parameters.)

Step 1: For any given assortment $S \subseteq N$, the optimal prices are

$$\mathbf{p}^*(S) \in \arg\max_{\mathbf{p} \geq \mathbf{c}} R(\mathbf{p}, S).$$

Step 2: Decide the optimal assortment S^* by $\mathbf{p}^*(S)$.

8.2.3 The MNL Model with Linear Deterministic Utilities

Assume $u_i\left(p_i\right) = a_i - b_i p_i, \forall i \in S$, where the price sensitivity $b_i > 0$.

For any given assortment $S \subseteq N$, the pricing problem is

$$\max_{\mathbf{p} \geq \mathbf{c}} R(\mathbf{p}, S) = \sum_{i \in S} \left(p_i - c_i\right) q_i = \frac{\sum_{i \in S} \left(p_i - c_i\right) v_i \left(p_i\right)}{1 + \sum_{j \in S} v_j \left(p_j\right)}$$

[Hanson and Martin, 1996] have shown that this formulation is not concave in prices.

A transformation:

$$\frac{q_i}{1 - \sum_{j \in S} q_j} = v_i\left(p_i\right) = e^{a_i - b_i p_i}$$

$$\Rightarrow p_i(\mathbf{q}) = \frac{a_i}{b_i} + \frac{1}{b_i}\left[\log\left(1 - \sum_{j \in S} q_j\right) - \log\left(q_i\right)\right]$$

where $\mathbf{q} = \left(q_1, \ldots, q_{|S|}\right)$ is the vector of market shares.

The profit function becomes

$$R(\mathbf{q}, S) = \sum_{i \in S} \left(p_i(\mathbf{q}) - c_i\right) q_i$$

The pricing problem for a given assortment S becomes

$$\max_{\mathbf{q}} R(\mathbf{q}, S)$$

Reference: [Li and Huh, 2011]

Theorem 13. *For the MNL model with linear deterministic utilities, for any given $S \subseteq N$, the profit function $R(\mathbf{q}, S)$ is* <u>*jointly concave*</u> *in each element $q_i, \forall i \in S$.*

Proof. Check the Hessian matrix of $R(\mathbf{q}, S)$ and use the preservation of concavity. $\square$

Remark: Theorem 1 is also true under the Nested Logit Model with linear deterministic utilities when the nest dissimilarity parameters $\gamma_k \leq 1$ and the products in the same nest has a common price sensitivity.

Reference: [Li and Huh, 2011]

Define a cost-adjusted quality for all products in set N as

$$\hat{v}_i := \exp\left(a_i - b_i c_i - 1\right), \quad \forall i \in N.$$

When assuming $b_1 = \cdots = b_{|S|} = b$, the optimal profit of given S is

$$\rho^*(S) = W\left(\sum_{j \in S} \hat{v}_j\right) / b$$

where $W(\cdot)$ is the Lambert W function which is the inverse of $f(x) = xe^x = z$, i.e., $W(z) = f^{-1}(z)$.

Lemma 7. *$W(z)$ is positive, increasing, strictly concave, and strictly log-concave.*

When assuming $b_1 = \cdots = b_{|S|} = b$, the optimal prices of given S are

$$p_i^*(S) - c_i = \rho^*(S) + \frac{1}{b} = \frac{W\left(\sum_{j \in S} \hat{v}_j\right) + 1}{b},$$

which is the so-called "equal markup pricing".

The normalized selling quantities (market shares) are

$$q_i^* = \frac{\hat{v}_i \exp\left(-b\rho^*(S)\right)}{1 + \sum_{j \in S} \hat{v}_j \exp\left(-b\rho^*(S)\right)} = \frac{\hat{v}_i}{\exp\left(W\left(\sum_{j \in S} \hat{v}_i\right)\right) + \sum_{j \in S} \hat{v}_i}$$

References: [Li and Huh, 2011]