# Useful Economic Theory and Mathematics

**Author:** Wenxiao Yang

**Institute:** Haas School of Business, University of California Berkeley

**Date:** 2023

*All models are wrong, but some are useful.*

## Contents

# Chapter 1 Stochastic Dominance

Based on

- MIT 14.123 S15 Stochastic Dominance Lecture Notes
- Princeton ECO317 Economics of Uncertainty Fall Term 2007 Notes for lectures 4. Stochastic Dominance
- Jensen, M. K. (2018). Distributional comparative statics. *The Review of Economic Studies*, 85(1), 581-610.

## 1.1 General Definitions

**Definition 1.1 (Jensen (2018), Defnition 1)**

Let $F$ and $G$ be two distributions on the same measurable space. Let $u$ be a function for which the following expression is well-defined,

$$\int u(x)dF \geq \int u(x)dG \tag{1.1}$$

Then:

- $F$ **first-order stochastically dominates** $G$ if 1.1 holds for any increasing function $u$.
- $F$ is a **mean-preserving spread** of $G$ if 1.1 holds for any convex function $u$.
- $F$ is a **mean-preserving contraction** of $G$ if 1.1 holds for any concave function $u$.
- $F$ **second-order stochastically dominates** $G$ if 1.1 holds for any concave and increasing function $u$.
- $F$ **dominates** $G$ **in the convex-increasing order** if 1.1 holds for any convex and increasing function $u$.

**Note** *$F$ is a **mean-preserving contraction** of $G \Leftrightarrow G$ is a **mean-preserving spread** of $F$.*

**Definition 1.2 (MPS and MPC)**

We define the following notations of sets.

- MPS$(f)$ is the set of all **mean-preserving spread** of $f$;
- MPC$(f)$ is the set of all **mean-preserving contraction** of $f$;

## 1.2 First-order Stochastic Dominance

### 1.2.1 Two Equivalent Definitions

**Definition 1.3 (First-order Stochastic Dominance)**

For any lotteries $F$ and $G$, $F$ **first-order stochastically dominates** $G$ if and only if the decision maker weakly prefers $F$ to $G$ under every weakly increasing utility function $u$, i.e.,

$$\int u(x)dF \geq \int u(x)dG$$

**Definition 1.4 (First-order Stochastic Dominance)**

For any lotteries $F$ and $G$, $F$ **first-order stochastically dominates** $G$ if and only if

$$F(x) \leq G(x), \forall x$$

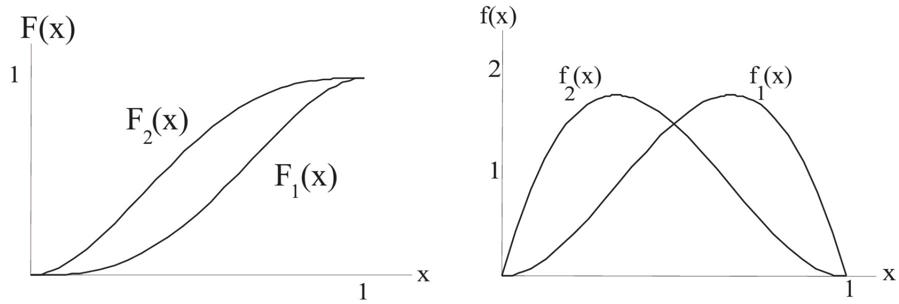


**Figure 1.1:** $F_1$ is FOSD over $F_2$: CDF and density comparison

## 1.3 Second-order Stochastic Dominance

### 1.3.1 Definition in terms of final goals

**Definition 1.5 (Second-order Stochastic Dominance)**

For any lotteries $F$ and $G$, $F$ **second-order stochastically dominates** $G$ if and only if the decision maker weakly prefers $F$ to $G$ under every weakly **increasing concave** utility function $u$, i.e.,

$$\int u(x)dF \geq \int u(x)dG$$

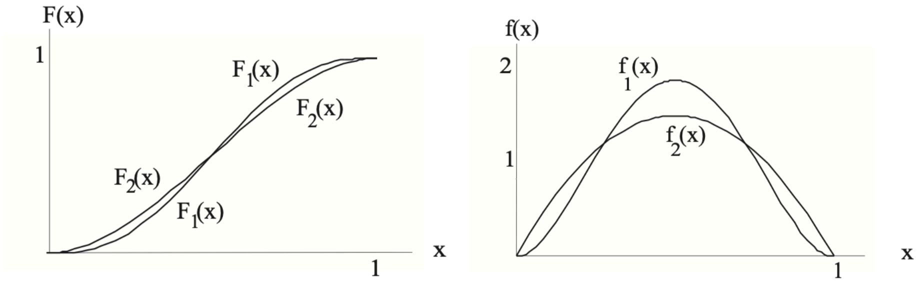


**Figure 1.2:** $F_1$ is SOSD over $F_2$: CDF and density comparison

### 1.3.2 Mean-Preserving Spread/Contraction

**Definition 1.6 (Mean-Preserving Spread)**

Let $x_F$ and $x_G$ be the random variables associated with lotteries $F$ and $G$. Then $G$ is a **mean-preserving spread** of $F$ if and only if

$$x_G \stackrel{d}{=} x_F + \varepsilon$$

for some random variable $\varepsilon$ such that $\mathbb{E}(\varepsilon \mid x_F) = 0 \ \forall x_F$.

The "$\stackrel{d}{=}$" means "is equal in distribution to" (that is, "has the same distribution as").

**Note** *Given $G$ is a **mean-preserving spread** of $F$, $G$ has larger variance than $F$.*

**Example 1.1**

$F(198) = \frac{1}{2}, F(202) = \frac{1}{2}$ and $G(100) = \frac{1}{100}, G(200) = \frac{98}{100}, G(300) = \frac{1}{100}$. Then

$$x_G \stackrel{d}{=} x_F + \varepsilon$$

where the distribution of $\varepsilon$ can be solved by

$$\begin{cases} G(300) = F(198)P(\varepsilon = 102|x_F = 198) + F(202)P(\varepsilon = 98|x_F = 202) \\ G(200) = F(198)P(\varepsilon = 2|x_F = 198) + F(202)P(\varepsilon = -2|x_F = 202) \\ G(100) = F(198)P(\varepsilon = -98|x_F = 198) + F(202)P(\varepsilon = -102|x_F = 202) \end{cases}$$

### 1.3.3 For Same Mean Distributions, Second-order Stochastic Dominance is equivalent to Mean-Preserving Spread

**Theorem 1.1 (Second-order Stochastic Dominance Equivalence)**

Given $\int x dF = \int x dG$ (same mean). The following are equivalent.

1. $F$ second-order stochastically dominates $G$: $\int u(x)dF \geq \int u(x)dG$ for every weakly increasing concave utility function $u$.

2. $F$ is a mean-preserving contraction of $G$ ($G$ is a mean-preserving spread of $F$).
3. For every $t \geq 0$, $\int_a^t G(x)dx \geq \int_a^t F(x)dx$.

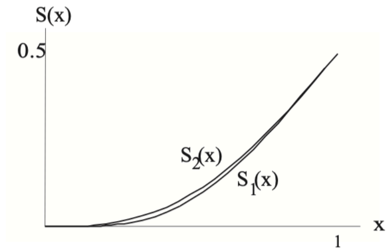


**Figure 1.3:** $F_1$ is SOSD over $F_2$, $S(t) : \int_a^t F_2(x)dx \geq \int_a^t F_1(x)dx$

**Corollary 1.1 (Euqivalent Definitions of MPC and MPS)**

$F$ is a mean-preserving contraction of $G$ (or $G$ is a mean-preserving spread of $F$) if and only if

(1). $\int xdF = \int xdG$

(2). $\int_a^t G(x)dx \geq \int_a^t F(x)dx, \forall t$

**Corollary 1.2 (MPC($f$) and MPS($f$) are convex and compact)**

MPC($f$) and MPS($f$) are **convex** and **compact**.

# Chapter 2 Tools for Comparative Statics

Consider the function $f : (0, 2\pi) \times \mathbb{R} \to \mathbb{R}$ s.t.

$$f(x, a) = \sin x + a$$

Let $X = (0, 2\pi)$ and let $f_a(x) = f(x, a) = \sin x + a$ denote the perturbed function for fixed $a$.

## 2.1 Regular and Critical Points and Values

### 2.1.1 Rank of Derivatives $\text{Rank} df_x = \text{Rank} Df(x)$

Suppose $X \subseteq \mathbb{R}^n$ is open. Suppose $f : X \to \mathbb{R}^m$ is differentiable at $x \in X$, and let $W = \{e_1, ..., e_n\}$ denote the standard basis of $\mathbb{R}^n$. Then $df_x \in L(\mathbb{R}^n, \mathbb{R}^m)$, and

$$\begin{aligned}
\text{Rank} df_x &= \dim \text{Im}(df_x) \\
&= \dim \text{span}\{df_x(e_1), ..., df_x(e_n)\} \\
&= \dim \text{span}\{Df(x)e_1, ..., Df(x)e_n\} \\
&= \dim \text{span}\{\text{column 1 of } Df(x), ..., \text{column n of } Df(x)\} \\
&= \text{Rank} Df(x)
\end{aligned}$$

Thus,

$$\text{Rank} df_x \leq \min\{m, n\}$$

$df_x$ has **full rank** if $\text{Rank} df_x = \min\{m, n\}$, that is, is $df_x$ has the maximum possible rank.

### 2.1.2 Regular and Critical Points and Values

**Definition 2.1 (Regular and Critical Points and Values)**

Suppose $X \subseteq \mathbb{R}^n$ is open. Suppose $f : X \to \mathbb{R}^m$ is differentiable at $x \in X$.

1. $x$ is a **regular point** of $f$ if $\text{Rank} df_x = \min\{m, n\}$.
2. $x$ is a **critical point** of $f$ if $\text{Rank} df_x < \min\{m, n\}$.
3. $y$ is a **critical value** of $f$ if there exists $x \in f^{-1}(y)$ such that $x$ is a critical point of $f$.
4. $y$ is a **regular value** of $f$ if $y$ is not a critical value of $f$.

**Note** *Notice that if $y \notin f(X)$, so $f^{-1}(y) = \emptyset$, then $y$ is automatically a regular value of $f$.*

**Example 2.1**

Suppose $f(x,y) = (\sin x, \cos y)$, $Df(x,y) = \begin{bmatrix} \cos x & 0 \\ 0 & -\sin y \end{bmatrix}$. Critical point: $\{(\frac{k\pi}{2}, \mathbb{R}) : k \in 2\mathbb{Z}+1\} \cup \{(\mathbb{R}, k\pi) : k \in \mathbb{Z}\}$; Critical values: $\{(x,y) : x = 1 \text{ or } x = -1 \text{ or } y = 1 \text{ or } y = -1\}$

## 2.2 Inverse and Implicit Function Theorem

### 2.2.1 Inverse Function Theorem

Using Taylor's theorem to approximate

$$f(x) = f(x_0) + Df(x_0)(x - x_0) + o(x - x_0)$$

The requirement of "regular point" is necessary for the $Df(x_0)$ being invertible.

**Theorem 2.1 (Inverse Function Theorem)**

Suppose $X \subseteq \mathbb{R}^n$ is open. Suppose $f : X \to \mathbb{R}^n$ is $C^1$ on $X$, and $x_0 \in X$. If $\det Df(x_0) \neq 0$ (i.e., $x_0$ is a regular point of $f$), then there are open neighborhoods $U$ of $x_0$ and $V$ of $f(x_0)$ s.t.

$$f : U \to V \text{ is bijective (on-to-on and onto)}$$

$$\exists\, f^{-1} : V \to U \text{ is } C^1$$

$$Df^{-1}(f(x_0)) = [Df(x_0)]^{-1}$$

$$(\text{In } \mathbb{R}, (f^{-1})'(f(x_0)) = (f'(x_0))^{-1})$$

If in addition $f \in C^k$, then $f^{-1} \in C^k$.

### 2.2.2 Implicit Function Theorem

Using Taylor's theorem to approximate

$$f(x,a) = f(x_0,a_0) + Df(x_0,a_0)(x - x_0) + Df(x_0,a_0)(a - a_0) + \text{remainder}$$

The requirement of "regular point" is necessary for the $Df(x_0, a_0)$ being invertible.

We want to know how the function $x^*(a)$ changes with keeping $f(x^*, a) = 0$.

**Theorem 2.2 (Implicit Function Theorem)**

Suppose $X \subseteq \mathbb{R}^n$ and $A \subseteq \mathbb{R}^p$ are open and $f : X \times A \to \mathbb{R}^n$ is $C^1$. Suppose $f(x_0, a_0) = 0$ and $\det(D_x f(x_0, a_0)) \neq 0$, i.e. $x_0$ is a regular point of $f(\cdot, a_0)$. Then there are open neighborhoods $U$ of $x_0$ ($U \subseteq X$) and $W$ of $a_0$ such that

$$\forall a \in W,\ \exists! x \in U \text{ s.t. } f(x,a) = 0$$

For each $a \in W$ let $g(a)$ be that unique $x$. Then $g : W \to U$ is $C^1$ and

$$Dg(a_0) = -[D_x f(x_0, a_0)]^{-1}[D_a f(x_0, a_0)]$$

If in addition $f \in C^k$, then $g \in C^k$.

### 2.2.3 Prove Implicit Function Theorem Given Inverse Function Theorem

**Proof 2.1**

1. Firstly, we prove "$g$ is differentiable": The "change of $a$" incurs the value change:

$$\begin{aligned} f(x_0, a_0 + h) &= f(x_0, a_0) + D_a f(x_0, a_0)h + o(h) \\ &= D_a f(x_0, a_0)h + o(h) \end{aligned}$$

   Find a $\Delta x$ such that the new $x$ can let the value go back to 0, i.e., $f(x_0 + \Delta x, a_0 + h) = 0$. That is,

$$g(a_0 + h) = x_0 + \Delta x$$

   To prove "$g$ is differentiable", we want to prove "$\exists T \in L(A, X)$ s.t. $\Delta x = T(h) + o(h)$"

$$\begin{aligned} 0 &= f(x_0 + \Delta x, a_0 + h) \\ &= f(x_0, a_0) + D_x f(x_0, a_0 + h)\Delta x + D_a f(x_0, a_0)h + o(\Delta x) + o(h) \\ &= D_x f(x_0, a_0 + h)\Delta x + D_a f(x_0, a_0)h + o(\Delta x) + o(h) \end{aligned}$$

$$D_x f(x_0, a_0 + h)\Delta x = -D_a f(x_0, a_0)h + o(\Delta x) + o(h)$$

   Because $f$ is $C^1$ and the determinant is a continuous function of the entries of the matrix, $\det D_x f(x_0, a_0 + h) \neq 0$ for $h$ sufficiently small, so

$$\Delta x = -[D_x f(x_0, a_0 + h)]^{-1} D_a f(x_0, a_0)h + o(\Delta x) + o(h)$$

$$\text{Since } f \in C^1, \Delta x = -[D_x f(x_0, a_0) + o(1)]^{-1} D_a f(x_0, a_0)h + o(\Delta x) + o(h)$$

$$\text{Since } f \in C^1, \Delta x = -[D_x f(x_0, a_0)]^{-1} D_a f(x_0, a_0)h + o(\Delta x) + o(h)$$

   Hence, "$g$ is differentiable" is proved and the derivative of $g$ is $Dg(a_0) = -[D_x f(x_0, a_0)]^{-1}[D_a f(x_0, a_0)]$.

2. Secondly, given the "$g$ is differentiable", we can also compute the derivative by

$$Df(g(a), a)(a_0) = 0$$

$$D_x f(x_0, a_0)Dg(a_0) + D_a f(x_0, a_0) = 0$$

$$Dg(a_0) = -[D_x f(x_0, a_0)]^{-1} D_a f(x_0, a_0)$$

**Example 2.2**

$f:\mathbb{R}^3\to\mathbb{R}^2$, $f((3,-1,2))=(0,0)$, $Df(3,-1,2)=\begin{bmatrix}1 & 2 & 1\\ 1 & -1 & 1\end{bmatrix}$. Then, let $(x_0,a_0)=(3,-1,2)$, where $x_0=3$ and $a_0=(-1,2)$. Or, we can let $(x_0,a_0)=(3,-1,2)$, where $x_0=(3,-1)$ and $a_0=2$.

### 2.2.4 Prove Inverse Function Theorem Given Implicit Function Theorem

**Proof 2.2 (Prove Inverse Function Theorem Given Implicit Function Theorem)**

Define $F:X\times\mathbb{R}^n$ s.t. $F(x,y)=y-f(x)$. Let $y_0=f(x_0)$.

$$D_xF(x,y)=-Df(x),\ D_yF(x,y)=I_{n\times n}$$

According to the implicit function theorem, there are open sets $U\subseteq X$ and $V\subseteq\mathbb{R}^n$ such that $x_0\in U$, $y_0\in V$ and a function $g:V\to U$ differentiable at $y_0$ such that $F(g(y),y)=0$ for all $y\in V$. So, $0=F(g(y),y)=y-f(g(y))$, we have $f(g(y))=y$, that is $g=f^{-1}$. $f:U\to V$ is bijective because it has inverse $g:V\to U$.

By the implicit function theorem, $g(y)$ is differentiable and

$$Df^{-1}(y_0)=Dg(y_0)=-[D_xF(x_0,y_0)]^{-1}[D_yF(x_0,y_0)]=[Df(x_0)]^{-1}$$

where $y_0=f(x_0)$.

By the implicit function theorem, the $g=f^{-1}$ is $C^k$ if $f$ is $C^k$.

All in all, the inverse function theorem is proved.

### 2.2.5 Example: Using Implicit Function Theorem

$x^2+y^2=c$. Define $g(x,y)=x^2+y^2-c$. The optimal solution of $y$ given $x$ is represented by $y^*(x)$. By the implicit function theorem,

$$\frac{\partial y^*}{\partial x}=-\frac{\frac{\partial g}{\partial x}\big|_{x,y^*}}{\frac{\partial g}{\partial y}\big|_{x,y^*}}$$

**Example 2.3**

Let us consider a firm that produces a good $y$; it uses two inputs $x_1$ and $x_2$. The firm sells the output and acquires the inputs in competitive markets: The market price of $y$ is $p$, and the cost of each unit of $x_1$ and $x_2$ are $w_1$ and $w_2$ respectively. Its technology is given by $f:\mathbb{R}^2_+\to\mathbb{R}_+$, where $f(x_1,x_2)=x_1^a x_2^b$, $a+b<1$. Its profits take the form

$$\pi(x_1,x_2;p,w_1,w_2)=px_1^a x_2^b-w_1x_1-w_2x_2$$

The firm selects $x_1$ and $x_2$ in order to maximize profits. **We aim to know how its choice of $x_1$ and $x_2$ is affected by a change in $w_1$.**

Assuming an interior solution, the first-order conditions of this optimization problem are

$$\frac{\partial \pi}{\partial x_1}(x_1^*, x_2^*; p, w_1, w_2) = pa(x_1^*)^{a-1}(x_2^*)^b - w_1 = 0$$
$$\frac{\partial \pi}{\partial x_2}(x_1^*, x_2^*; p, w_1, w_2) = pb(x_1^*)^{a}(x_2^*)^{b-1} - w_2 = 0$$

for some $(x_1, x_2) = (x_1^*, x_2^*)$.

Let us define

$$F(x_1^*, x_2^*; p, w_1, w_2) = \begin{bmatrix} pa(x_1^*)^{a-1}(x_2^*)^b - w_1 \\ pb(x_1^*)^{a}(x_2^*)^{b-1} - w_1 \end{bmatrix}$$

Jacobian matrices are

$$D_{(x_1,x_2)}F(x_1^*, x_2^*; p, w_1, w_2) = \begin{bmatrix} pa(a-1)(x_1^*)^{a-2}(x_2^*)^b & pab(x_1^*)^{a-1}(x_2^*)^{b-1} \\ pab(x_1^*)^{a-1}(x_2^*)^{b-1} & pb(b-1)(x_1^*)^{a}(x_2^*)^{b-2} \end{bmatrix}$$

$$D_{w_1}F(x_1^*, x_2^*; p, w_1, w_2) = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

By the implicit function theorem, we can get

$$\begin{aligned} \begin{bmatrix} \frac{\partial x_1^*}{\partial w_1} \\ \frac{\partial x_2^*}{\partial w_1} \end{bmatrix} &= -[D_{(x_1,x_2)}F(x_1^*, x_2^*; p, w_1, w_2)]^{-1}[D_{w_1}F(x_1^*, x_2^*; p, w_1, w_2)] \\ &= [D_{(x_1,x_2)}F(x_1^*, x_2^*; p, w_1, w_2)]^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \end{aligned}$$

### 2.2.6 Corollary: $a \to \{x \in X : f(x,a) = 0\}$ is lhc

**Corollary 2.1**

Suppose $X \subseteq \mathbb{R}^n$ and $A \subseteq \mathbb{R}^p$ are open and $f : X \times A \to \mathbb{R}^n$ is $C^1$. If 0 is a regular value of $f(\cdot, a_0)$, then the correspondence

$$a \to \{x \in X : f(x,a) = 0\}$$

is **lower hemicontinuous** at $a_0$.

## 2.3 Transversality and Genericity

### 2.3.1 Lebesgue Measure Zero

**Definition 2.2 (Lebesgue Measure Zero)**

Suppose $A \subseteq \mathbb{R}^n$. $A$ has **Lebesgue measure zero** if for every $\varepsilon > 0$ there is a countable collection of rectangles $I_1, I_2, \ldots$ such that

$$\sum_{k=1}^{\infty} \text{Vol}(I_k) < \varepsilon \text{ and } A \subseteq \cup_{k=1}^{\infty} I_k$$

Here by a rectangle we mean $I_k = \times_{j=1}^{n}(a_j^k, b_j^k) = \{x \in \mathbb{R}^n : x_j \in (a_j^k, b_j^k), \forall j\}$ for some $a_j^k < b_j^k \in \mathbb{R}$, and

$$\text{Vol}(I_k) = \prod_{j=1}^{n} |b_j^k - a_j^k|$$

**Example 2.4**

1. "Lower-dimensional" sets have Lebesgue measure zero. For example, $A = \{x \in \mathbb{R}^2 : x_2 = 0\}$
2. Any **finite** set has Lebesgue measure zero in $\mathbb{R}^n$.
3. **Finite Union** of sets that have Lebesgue measure zero has Lebesgue measure zero: If $A_n$ has Lebesgue measure zero $\forall n$ then $\cup_{n \in N} A_n$ has Lebesgue measure zero.
4. Every **countable** set (e.g. $\mathbb{Q}$) has Lebesgue measure zero.
5. No open set in $\mathbb{R}^n$ has Lebesgue measure zero.

### 2.3.2 Sard's Theorem

**Theorem 2.3 (Sard's Theorem)**

Let $X \subseteq \mathbb{R}^n$ be open, and $f : X \to \mathbb{R}^m$ be $C^r$ with $r \geq 1 + max\{0, n - m\}$. Then the set of all critical values of $f$ has Lebesgue measure zero.

### 2.3.3 Transversality Theorem

**Theorem 2.4 (Transversality Theorem)**

Let $X \subseteq \mathbb{R}^n$ and $A \subseteq \mathbb{R}^p$ be open, and $f : X \times A \to \mathbb{R}^m$ be $C^r$ with $r \geq 1 + max\{0, n - m\}$. Suppose that 0 is a regular value of $f$ (that is all $(x, a)$ such that $f(x, a) = 0$ are regular points). Then,

1. $\exists A_0 \subseteq A$ such that $A \backslash A_0$ has Lebesgue measure zero.
2. $\forall a \in A_0$, 0 is a regular value of $f_a = f(\cdot, a)$.

**Example 2.5**

$f : \mathbb{R}^4 \to \mathbb{R}^3$ s.t. $f(x, y, z, w) = (g(x) + y, z^3 + 1, w + x + y^2)$

## 2.4 Envelope Theorem

**Theorem 2.5 (Envelope Theorem)**

Suppose that $f(x, \cdot)$ is absolutely continuous for all $x \in X$. Suppose there exists an integrable function $b : [0, 1] \to \mathbb{R}_+$ such that $|f_t(x, t)| \leq b(t)$ for all $x \in X$ and almost all $t \in [0, 1]$. Then $V(t) = \sup_{x \in X} f(x, t)$ is absolutely continuous.

Suppose, in addition, that $f(x, \cdot)$ is differentiable for all $x \in X$, and that $X^*(t) = \{x \in X : f(x, t) = V(t)\} \neq \emptyset$ almost everywhere on $[0, 1]$. Then, for any selection $x^*(t) \in X^*(t)$,

$$V(t) = V(0) + \int_0^t f_t(x^*(s), s)ds$$

# Chapter 3 Fixed Point Theorem

## 3.1 Contraction Mapping Theorem (@ Lec 05 of ECON 204)

### 3.1.1 Contraction: Lipschitz continuous with constant $< 1$

**Definition 3.1**

Let $(X, d)$ be a nonempty complete metric space. An operator is a function $T : X \to X$. An operator $T$ is a **contraction of modulus** $\beta$ if $\beta < 1$ and

$$d(T(x), T(y)) \leq \beta d(x, y), \forall x, y \in X$$

A contraction shrinks distances by a *uniform* factor $\beta < 1$.

### 3.1.2 Theorem: Contraction $\Rightarrow$ Uniformly Continuous

**Theorem 3.1 (Contraction $\Rightarrow$ Uniformly Continuous)**

Every contraction is uniformly continuous.

**Proof 3.1**

Let $\delta = \frac{\varepsilon}{\beta}$.

### 3.1.3 Blackwell's Sufficient Conditions for Contraction

Let $X$ be a set, and let $B(X)$ be the set of all bounded functions from $X$ to $\mathbb{R}$. Then $(B(X), \|\cdot\|_\infty)$ is a normed vector space.

(Notice that below we use shorthand notation that identifies a constant function with its constant value in $\mathbb{R}$, that is, we write interchangeably $a \in \mathbb{R}$ and $a : X \to \mathbb{R}$ to denote the function such that $a(x) = a, \forall x \in X$.)

**Theorem 3.2 (Blackwell's Sufficient Conditions)**

Consider $B(X)$ with the sup norm $\|\cdot\|_\infty$. Let $T : B(X) \to B(X)$ be an operator satisfying

1. (monotonicity) $f(x) \leq g(x), \forall x \in X \Rightarrow (Tf)(x) \leq (Tg)(x), \forall x \in X$
2. (discounting) $\exists \beta \in (0, 1)$ such that for every $a \geq 0$ and $x \in X$,

$$(T(f + a))(x) \leq (Tf)(x) + \beta a$$

Then $T$ is a contraction with modulus $\beta$.

**Proof 3.2**

Fix $f, g \in B(X)$. By the definition of the sup norm,

$$f(x) \leq g(x) + \|f - g\|_\infty \forall x \in X$$

Then

$$(Tf)(x) \leq (T(g + \|f - g\|_\infty))(x) \leq (Tg)(x) + \beta\|f - g\|_\infty \quad \forall x \in X$$

where the first inequality above follows from monotonicity, and the second from discounting. Thus

$$(Tf)(x) - (Tg)(x) \leq \beta\|f - g\|_\infty \quad \forall x \in X$$

Reversing the roles of $f$ and $g$ above gives

$$(Tg)(x) - (Tf)(x) \leq \beta\|f - g\|_\infty \quad \forall x \in X$$

Thus

$$\|T(f) - T(g)\|_\infty \leq \beta\|f - g\|_\infty$$

Thus $T$ is a contraction with modulus $\beta$

## 3.2 Fixed Point Theorem (@ Lec 05 of ECON 204)

### 3.2.1 Fixed Point

**Definition 3.2 (Fixed Point)**

A **fixed point** of an operator $T$ is element $x^* \in X$ such that $T(x^*) = x^*$.

**Definition 3.3 (Fixed Point of Function)**

Let $X$ be a nonempty set and $f : X \to X$. A point $x^* \in X$ is a **fixed point** of $f$ if $f(x^*) = x^*$.

**Example 3.1**

Let $X = \mathbb{R}$ and $f : \mathbb{R} \to \mathbb{R}$

1. $f(x) = 2x$ has fixed point: $x = 0$.
2. $f(x) = x$ has fixed points: $x \in \mathbb{R}$.
3. $f(x) = x + 1$ doesn't have fixed points.

### 3.2.2 ★ Contraction Mapping Theorem: contraction ⇒ exist unique fixed point

**Theorem 3.3 (Contraction Mapping Theorem)**

Let $(X, d)$ be a nonempty complete metric space and $T : X \to X$ a contraction with modulus $\beta < 1$. Then

1. $T$ has a unique fixed point $x^*$.
2. For every $x_0 \in X$, the sequence defined by

$$\begin{aligned} x_1 &= T(x_0) \\ x_2 &= T(x_1) = T(T(x_0)) = T^2(x_0) \\ &\vdots \\ x_{n+1} &= T(x_n) = T^{n+1}(x_0) \end{aligned}$$

   converges to $x^*$.

Note that the theorem asserts both the **existence** and **uniqueness** of the fixed point, as well as giving an **algorithm** to find the fixed point of a contraction.

**Proof 3.3**

Define the sequence $\{x_n\}$ as above. Then,

$$\begin{aligned} d(x_{n+1}, x_n) &= d(T(x_n), T(x_{n-1})) \\ &\leq \beta d(x_n, x_{n-1}) \\ &\leq \beta^n d(x_1, x_0) \end{aligned}$$

Then for any $n > m$,

$$\begin{aligned} d(x_n, x_m) &\leq d(x_1, x_0) \sum_{i=m}^{n-1} \beta^i \\ &< d(x_1, x_0) \sum_{i=m}^{\infty} \beta^i \\ &= \frac{\beta^m}{1-\beta} d(x_1, x_0) \to 0 \text{ as } m \to \infty \end{aligned}$$

Fixed $\varepsilon > 0$, we can choose $N(\varepsilon)$ such that $\forall n, m > N(\varepsilon)$,

$$d(x_n, x_m) < \frac{\beta^m}{1-\beta} d(x_1, x_0) < \varepsilon$$

Therefore, $\{x_n\}$ is Cauchy. Since $(X, d)$ is complete, $x_n \to x^*$ for some $x^* \in X$.

Next we show that $x^*$ is a fixed point of $T$.

$$\begin{aligned} T\left(x^*\right) &= T\left(\lim_{n\to\infty} x_n\right) \\ &= \lim_{n\to\infty} T\left(x_n\right) \text{ since } T \text{ is continuous} \\ &= \lim_{n\to\infty} x_{n+1} \\ &= x^* \end{aligned}$$

so $x^*$ is a fixed point of $T$.

Finally, we show that there is at most one fixed point. Suppose $x^*$ and $y^*$ are both fixed points of $T$, so $T\left(x^*\right) = x^*$ and $T\left(y^*\right) = y^*$. Then

$$\begin{aligned} d\left(x^*, y^*\right) &= d\left(T\left(x^*\right), T\left(y^*\right)\right) \\ &\leq \beta d\left(x^*, y^*\right) \\ \Rightarrow (1-\beta) d\left(x^*, y^*\right) &\leq 0 \\ \Rightarrow d\left(x^*, y^*\right) &\leq 0 \end{aligned}$$

So $d\left(x^*, y^*\right) = 0$, which implies $x^* = y^*$.

### 3.2.3 Conditions for Fixed Point's Continuous Dependence on Parameters

**Theorem 3.4 (Continuous Dependence on Parameters)**

Let $(X, d)$ and $(\Omega, \rho)$ be two metric spaces and $T : X \times \Omega \to X$. For each parameter $\omega \in \Omega$ let $T_\omega : X \to X$ be defined by $T_\omega(x) = T(x, \omega)$.

Suppose (1). $(X, d)$ is complete, (2). $T$ is continuous in $\omega$ (that is $T(x, \cdot) : \Omega \to X$ is continuous for each $x \in X$), and (3). $\exists \beta < 1$ such that $T_\omega$ is a contraction of modulus $\beta$ $\forall \omega \in \Omega$.

Then the fixed point function (about parameter $\omega$) $x^* : \Omega \to X$ defined by $x^*(\omega) = T_\omega(x^*(\omega))$ is continuous.

## 3.3 Brouwer's Fixed Point Theorem (@ Lec 13 of ECON 204)

### 3.3.1 Simple One: One-dimension

**Theorem 3.5**

Let $X = [a, b]$ for $a, b \in \mathbb{R}$ with $a < b$ and let $f : X \to X$ be continuous. Then $f$ has a fixed point.

**Proof 3.4**

Easily proved by Intermediate Value Theorem.

### 3.3.2 ★ Brouwer's Fixed Point Theorem: continuous function has fixed point over compact, convex set

**Theorem 3.6 (Brouwer's Fixed Point Theorem)**

Let $X \subseteq \mathbb{R}^n$ be nonempty, **compact**, and **convex**, and let $f : X \to X$ be continuous. Then $f$ has a fixed point.

**Proof 3.5**

Consider the case when the set $X$ is the unit ball in $\mathbb{R}^n$.

Using a fact that "Let $B$ be the unit ball in $\mathbb{R}^n$. Then there is no continuous function $h : B \to \partial B$ such that $h(x_0) = x_0$ for every $x_0 \in \partial B$", which is intuitive but hard to prove. (See *J. Franklin, Methods of Mathematical Economics*, for an elementary (but long) proof.)

Then prove by contradiction: suppose $f$ has no fixed points in $B$. That is, $\forall x \in B$, $x \neq f(x)$. Since x and its image $f(x)$ are distinct points in $B$ for every $x$, we can carry out the following construction. For each $x \in B$, construct the line segment originating at $f(x)$ and going through $x$. Let $g(x)$ denote the intersection of this line segment with $\partial B$. This construction gives a continuous function $g : B \to \partial B$. Furthermore, notice that if $x_0 \in \partial B$, then $x_0 = g(x_0)$. Then, $g$ gives $g(x) = x, \forall x \in \partial B$. Since there are no such functions by the fact above, we have a contradiction.

# Chapter 4 Correspondence: $\Psi : X \to 2^Y$ (@ Lec 07 of ECON 204)

**Definition 4.1 (Correspondence)**

A **correspondence** $\Psi : X \to 2^Y$ from $X$ to $Y$ is a function from $X$ to $2^Y$, that is, $\Psi(x) \subseteq Y$ for every $x \in X$. ($2^Y$ is the set of all subsets of $Y$)

**Example 4.1**

Let $u : \mathbb{R}^n_+ \to \mathbb{R}$ be a continuous utility function, $y > 0$ and $p \in \mathbb{R}^n_{++}$, that is, $p_i > 0$ for each $i$. Define $\Psi : \mathbb{R}^n_{++} \times \mathbb{R}_{++} \to 2^{\mathbb{R}^n_+}$ by

$$\begin{aligned} \Psi(p, y) = &\operatorname{argmax} u(x) \\ &\text{s.t. } x \geq 0 \\ &\quad p \cdot x \leq y \end{aligned}$$

$\Psi$ is the demand correspondence associated with the utility function $u$; typically $\Psi(p, y)$ is multi-valued.

## 4.1 Continuity of Correspondences

### 4.1.1 Upper/Lower Hemicontinuous

Let $X \subseteq \mathbb{E}^n$, $Y \subseteq \mathbb{E}^m$, and $\Psi : X \to 2^Y$.

**Definition 4.2 (Upper Hemicontinuous)**

$\Psi$ is **upper hemicontinuous** (uhc) at $x_0 \in X$ if, for every open set $V$ with $\Psi(x_0) \subseteq V$, there is an open set $U$ with $x_0 \in U$ s.t.

$$x \in U \Rightarrow \Psi(x) \subseteq V$$

Upper hemicontinuity reflects the requirement that $\Psi$ doesn't "jump down/implode in the limit" at $x_0$. *(A set to "jump down" at the limit $x_0$: It should mean the set suddenly gets smaller – it "implodes in the limit" – that is, there is a sequence $x_n \to x_0$ and points $y_n \in \Psi(x_n)$ that are far from every point of $\Psi(x_0)$ as $n \to \infty$.)*

**Definition 4.3 (Lower Hemicontinuous)**

$\Psi$ is **lower hemicontinuous** (lhc) at $x_0 \in X$ if, for every open set $V$ with $\Psi(x_0) \cap V \neq \emptyset$, there is an open set $U$ with $x_0 \in U$ s.t.

$$x \in U \Rightarrow \Psi(x) \cap V \neq \emptyset$$

Lower hemicontinuity reflects the requirement that $\Psi$ doesn't "jump up/explode in the limit" at $x_0$. *(A set to* *"jump up" at the limit $x_0$: It should mean that the set suddenly gets bigger – it "explodes in the limit" – that is, there is a sequence $x_n \to x_0$ and a point $y_0 \in \Psi(x_0)$ that is far from every point of $\Psi(x_n)$ as $n \to \infty$.)*

**Definition 4.4 (Continuous Correspondence)**

$\Psi$ is **continuous** at $x_0 \in X$ if it is both **uhc** and **lhc** at $x_0$.

**Proposition 4.1**

$\Psi$ is upper hemicontinuous (respectively lower hemicontinuous, continuous) if it is uhc (respectively lhc, continuous) at every $x \in X$.

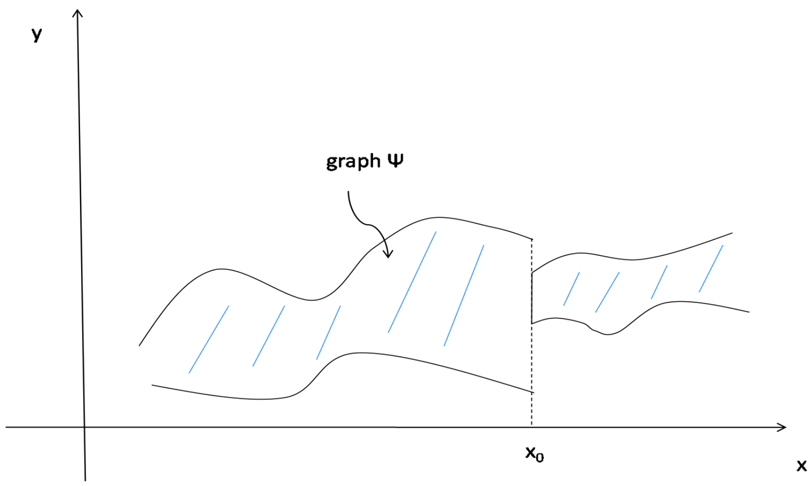


**Figure 4.1:** The correspondence $\Psi$ "implodes in the limit" at $x_0$. $\Psi$ is not upper hemicontinuous at $x_0$.

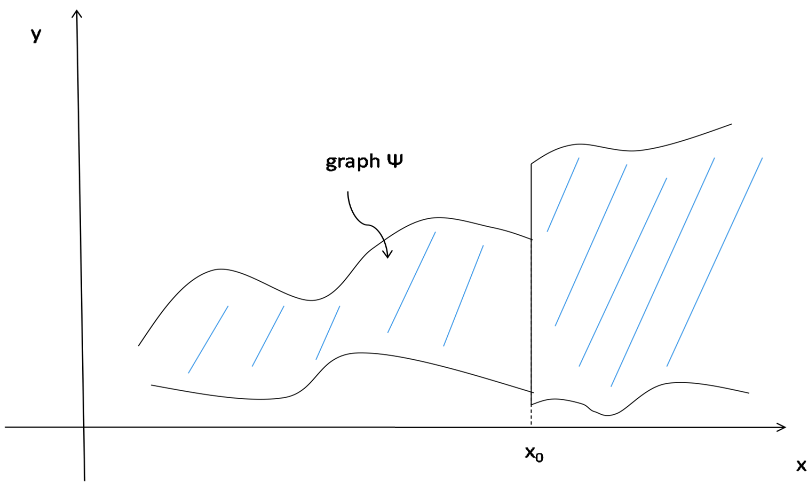


**Figure 4.2:** The correspondence $\Psi$ "explodes in the limit" at $x_0$. $\Psi$ is not lower hemicontinuous at $x_0$.

### 4.1.2 Theorem: $\Psi(x) = \{f(x)\}$ is uhc $\Leftrightarrow$ $f$ is continuous

**Theorem 4.1 ($\Psi(x) = \{f(x)\}$ is uhc $\Leftrightarrow$ $f$ is continuous)**

Let $X \subseteq \mathbb{E}^n$, $Y \subseteq \mathbb{E}^m$ and $f : X \to Y$. Let $\Psi : X \to 2^Y$ be defined by $\Psi(x) = \{f(x)\}$ for all $x \in X$. Then $\Psi$ is **uhc** if and only if $f$ is **continuous**.

### 4.1.3 Berge's Maximum Theorem: the set of maximizers is uhc with non-empty compact values

**Theorem 4.2 (Berge's Maximum Theorem)**

Let $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$. Consider the function $f : X \times Y \to \mathbb{R}$ and the correspondence $\Gamma : Y \to 2^X$. Define $v(y) = \max_{x\in\Gamma(y)} f(x, y)$ and the set of maximizers

$$\Omega(y) = \underset{x\in\Gamma(y)}{\operatorname{argmax}} f(x, y) = \{x : f(x, y) = v(y)\}$$

Suppose $f$ and $\Gamma$ are continuous, and that $\Gamma$ has non-empty compact values. Then, $v$ is continuous and $\Omega$ is uhc with non-empty compact values.

## 4.2 Graph of Correspondence

An alternative notion of continuity looks instead at properties of the graph of the correspondence.

**Definition 4.5 (Graph of Correspondence)**

The **graph** of a correspondence $\Psi : X \to 2^Y$ is the set

$$\text{graph}\Psi = \{(x, y) \in X \times Y : y \in \Psi(x)\}$$

### 4.2.1 Closed Graph

By the definition of continuous function $f : \mathbb{R}^n \to \mathbb{R}$, each convergent sequence $\{(x_n, y_n)\}$ in graph $f$ converges to a point $(x, y)$ in graph $f$, that is, graph $f$ is closed.

**Definition 4.6 (Closed Graph)**

Let $X \subseteq \mathbb{E}^n$, $Y \subseteq \mathbb{E}^m$. A correspondence $\Psi : X \to 2^Y$ has closed graph if its graph is a closed subset of $X \times Y$, that is, if for any sequences $\{x_n\} \subseteq X$ and $\{y_n\} \subseteq Y$ such that $x_n \to x \in X$, $y_n \to y \in Y$ and $y_n \in \Psi(x_n)$ for each $n$, then $y \in \Psi(x)$.

**Example 4.2**

Consider the correspondence $\Psi(x) = \begin{cases} \{\frac{1}{x}\}, & \text{if } x \in (0,1] \\ \{0\}, & \text{if } x = 0 \end{cases}$ ("implode in the limit")

Let $V = (-0.1, 0.1)$. Then $\Psi(0) = \{0\} \subseteq V$, but no matter how close $x$ is to $0$, $\Psi(x) = \{\frac{1}{x}\} \nsubseteq V$, so $\Psi$ is not uhc at $0$. However, note that $\Psi$ has closed graph.

## 4.3 Closed-valued, Compact-valued, and Convex-valued Correspondences

**Definition 4.7 (Closed-valued, Compact-valued, and Convex-valued Correspondences)**

Given a correspondence $\Psi : X \to 2^Y$,

1. $\Psi$ is **closed-valued** if $\Psi(x)$ is a closed subset of $Y$ for all $x$;
2. $\Psi$ is **compact-valued** if $\Psi(x)$ is compact for all $x$.
3. $\Psi$ is **convex-valued** if $\Psi(x)$ is convex for all $x$.

### 4.3.1 Closed-valued, uhc and Closed Graph

For closed-valued correspondences these concepts can be more tightly connected. A closed-valued and upper hemicontinuous correspondence must have closed graph. For a closed-valued correspondence with a compact range, upper hemicontinuity is equivalent to closed graph.

**Theorem 4.3**

Let $X \subseteq \mathbb{E}^n$, $Y \subseteq \mathbb{E}^m$, and $\Psi : X \to 2^Y$.

1. $\Psi$ is **closed-valued** and **uhc** $\Rightarrow$ $\Psi$ has **closed graph**.
2. $\Psi$ is **closed-valued** and **uhc** $\Leftarrow$ $\Psi$ has **closed graph**. (If $Y$ is **compact**)

**Theorem 4.4**

Let $X \subseteq \mathbb{E}^n$, $Y \subseteq \mathbb{E}^m$, and $\Psi : X \to 2^Y$. If $\Psi$ has **closed graph** and there is an **open set** $W$ with $x_0 \in W$ and a **compact set** $Z$ such that $x \in W \cap X \Rightarrow \Psi(x) \subseteq Z$, then $\Psi$ is **uhc** at $x_0$.

### 4.3.2 Theorem: compact-valued, uhs correspondence of compact set is compact

**Theorem 4.5**

Let $X$ be a compact set and $\Psi : X \to 2^X$ be a non-empty, compact-valued upper-hemicontinuous correspondence. If $C \subseteq X$ is compact, then $\Psi(C)$ is compact.

**Proof 4.1**

Given the compact-valued $\Psi$, we can have an open cover of $\Psi(C)$, $\{U_\lambda : \lambda \in \Lambda\}$. So $\forall x \in C$, there exists $U_{l(x)}, l(x) \in \Lambda$ such that $U_{l(x)}$ is an open cover of $\Psi(x)$.

Consider a $c \in C$. Since $\Psi$ is uhs and $\Psi(c) \subseteq U_{l(c)}$, there exists open set $V_c$ s.t. $c \in V_c$ and $\Psi(x) \subseteq U_{l(c)}, \forall x \in V_c \cap C$.

$\{V_c : c \in C\}$ is an open cover of $C$. Because $C$ is compact, there is a finite subcover $\{V_{c_i} : i = 1, ..., m\}, m \in \mathbb{N}$, where $\{c_i : i = 1, ..., m\} \subseteq C$.

Because $\Psi(x) \subseteq U_{l(c_i)}, \forall x \in V_{c_i} \cap C$ and $\{V_{c_i} : i = 1, ..., m\}, m \in \mathbb{N}$ is a open cover for $C$, we can infer $\{U_{l(c_i)} : i = 1, ..., m\}$ is a finite subcover of $\{U_{l(c)} : c \in C\}$ for $\Psi(C)$. Hence, $\Psi(C)$ is compact.

## 4.4 Fixed Points for Correspondences (@ Lec 13 of ECON 204)

### 4.4.1 Definition

**Definition 4.8 (Fixed Points for Correspondences)**

Let $X$ be nonempty and $\psi : X \to 2^X$ be a correspondence. A point $x^* \in X$ is a fixed point of $\psi$ if $x^* \in \psi(x^*)$.

**Note** *We only need $x^*$ to be in $\psi(x^*)$, not $\{x^*\} = \psi(x^*)$. That is, $\psi$ need not be single-valued at $x^*$. So $x^*$ can be a fixed point of $\psi$ but there may be other elements of $\psi(x^*)$ different from $x^*$.*

### 4.4.2 Kakutani's Fixed Point Theorem: uhs, compact, convex values correspondence has a fixed point over compact convex set

**Theorem 4.6 (Kakutani's Fixed Point Theorem)**

Let $X \subseteq \mathbb{R}^n$ be a non-empty, **compact**, **convex** set and $\psi : X \to 2^X$ be an **upper hemi-continuous** correspondence with non-empty, **compact**, **convex** values. Then $\psi$ has a fixed point in $X$.

### 4.4.3 Theorem: $\exists$ compact set $C = \cap_{i=0}^{\infty} \Psi^i(X)$ s.t. $\Psi(C) = C$

**Theorem 4.7**

Let $(X, d)$ be a compact metric space and let $\Psi(x) : X \to 2^X$ be a upper-hemicontinuous, compact-valued correspondence, such that $\Psi(x)$ is non-empty for every $x \in X$. There exists a compact non-empty subset $C \subseteq X$, such that $\Psi(C) \equiv \cup_{x \in C} \Psi(x) = C$.

**Proof 4.2**

Let's construct a sequence $\{C_n\}$ such that $C_0 = X$, $C_1 = \Psi(C_0)$, ..., $C_n = \Psi(C_{n-1})$, ... We claim that $C = \cap_{i=0}^{\infty} C_i$ is a non-empty compact set and satisfies $\Psi(C) = C$.

1. Because we can infer $\Psi(X_1) \subseteq \Psi(X_2)$ if $X_1 \subseteq X_2$, $X = C_0 \supseteq C_1 \Rightarrow C_1 = \Psi(C_0) \supseteq C_2 = \Psi(C_1)$,...., so $C_0 \supseteq C_1 \supseteq \cdots C_n \supseteq \cdots$. Hence, $C$ is not empty.
2. Because $X$ is compact, by the theorem 4.5, we can infer $C_n$ is compact for all $n$. Then, $C_n$ is closed for all $n$, so $C$ is closed. Because $C$ is a closed set of compact set $X$, $C$ is compact.
3. $C \subseteq C_n, \forall n \Rightarrow \Psi(C) \subseteq \Psi(C_n), \forall n \Rightarrow \Psi(C) \subseteq C$
4. Assume $C \subseteq \Psi(C)$ doesn't hold, that is $\exists y \in C$ s.t. $y \notin \Psi(C)$. Because $y \in C$ and $C_0 \supseteq C_1 \supseteq \cdots C_n \supseteq \cdots$, there exists $k \in C_n$ for all $n$ s.t. $y \in \Psi(k)$. $k \in \cap_{i=1}^{\infty} C_i = C$, so $\Psi(k) \subseteq \Psi(C)$, which contradicts to $y \notin \Psi(C)$. Hence, $C \subseteq \Psi(C)$.

All in all the claim "$C = \cap_{i=0}^{\infty} C_i$ is a non-empty compact set and satisfies $\Psi(C) = C$" is proved.

# Chapter 5 Bayesian Persuasion: Extreme Points and Majorization

Based on

- Kleiner, A., Moldovanu, B., & Strack, P. (2021). Extreme points and majorization: Economic applications. *Econometrica*, 89(4), 1557-1593.
- 

## 5.1 Extreme Points

### 5.1.1 Extreme Points of Convex Set

**Definition 5.1 (Extreme Points)**

An **extreme point** of a convex set $A$ is a point $x \in A$ that cannot be represented as a convex combination of points in $A$.

### 5.1.2 Krein-Milman Theorem: Existence of Extreme Points

**Theorem 5.1 (Krein-Milman Theorem)**

Every non-empty **compact convex** subset of a Hausdorff locally convex topological vector space (for example, a normed space) is the closed, convex hull of its extreme points.

In particular, this set has extreme points.

### 5.1.3 Bauer's Maximum Principle: Usefulness of Extreme Points for Optimization

**Theorem 5.2 (Bauer's Maximum Principle)**

Any function that is **convex and continuous**, and defined on a set that is **convex and compact**, attains its maximum at some extreme point of that set.

## 5.2 Majorization

### 5.2.1 Majorization and Weak Majorization

**Definition 5.2 (Majorization of Non-decreasing Functions)**

Consider right-continuous functions that map the unit interval $[0,1]$ into the real numbers. For two non-decreasing functions $f, g \in L^1$, we say that $f$ **majorizes** $g$, denoted by $g \prec f$, if the following two conditions hold:

$$\int_x^1 g(s)ds \leq \int_x^1 f(s)ds, \forall x \in [0,1] \quad \text{(Condition 1)}$$

$$\int_0^1 g(s)ds = \int_0^1 f(s)ds \quad \text{(Condition 2)}$$

**Definition 5.3 (Weak Majorization)**

$f$ **weakly majorizes** $g$, denoted by $g \prec_w f$, if Condition 1 holds (not necessarily Condition 2).

### 5.2.2 How to work for non-monotonic functions? – Non-Decreasing Rearrangement

**Note How this work with non-monotonic functions?**

Suppose $f, g$ are non-monotonic, we compare their non-decreasing rearrangements $f^*$, $g^*$.

**Definition 5.4 (Rearrangement)**

Given a function $f$, let $m(x)$ denote the Lebesgue measure of the set $\{s \in [0,1] : f(s) \leq x\}$, that is $m(x) = \int_{s\in\{s\in[0,1]:f(s)\leq x\}} 1ds$ (the "length" of the set). The non-decreasing rearrangement of $f$, $f^*$, is defined by

$$f^*(t) = \inf\{x \in \mathbb{R} : m(x) \geq t\},\ t \in [0,1]$$

### 5.2.3 Theorem: $F$ majorizes $G \Leftrightarrow G$ is a mean-preserving spread of $F$

Based on

- Shaked, M., & Shanthikumar, J. G. (2007). *Stochastic orders*. New York, NY: Springer New York.

**Definition 5.5 (Generalized Inverse)**

Suppose $G$ is defined on the interval $[0,1]$, we can define the **generalized inverse**

$$G^{-1}(x) = \sup\{s : G(s) \leq x\}, x \in [0,1]$$

Let $X_F$ and $X_G$ be now random variables with distributions $F$ and $G$, defined on the interval $[0,1]$.

**Theorem 5.3 (Shaked & Shanthikumar (2007), Section 3.A)**

$$G \prec F \Leftrightarrow F^{-1} \prec G^{-1} \Leftrightarrow X_G \leq_{ssd} X_F \text{ and } \mathbb{E}[X_G] = \mathbb{E}[X_F]$$

where $\leq_{ssd}$ denotes the standard second-order stochastic dominance.

Based on Theorem 1.1 and the Condition 2 of Majorization, we can conclude

**Corollary 5.1 (Majorization $\Leftrightarrow$ Mean-preserving Contraction)**

$F$ majorizes $G \Leftrightarrow F$ is a mean-preserving contraction of $G$ ($G$ is a mean-preserving spread of $F$)

That is, we can construct random variables $X_F$, $X_G$, jointly distributed on some probability space, such that $X_F \sim F$, $X_G \sim G$ and such that $X_F = \mathbb{E}[X_G \mid X_F]$.

## 5.3 Capture Extreme Points in Economic Applications

Let $L^1$ denote the real-valued and integrable functions defined on $[0, 1]$.

In this section, we focus on **non-decreasing (weakly increasing) functions**, for example, a cumulative distribution function in Bayesian persuasion, or an incentive-compatible allocation in mechanism design.

### 5.3.1 Definitions of $\mathcal{MPS}(f)$, $\mathcal{MPS}_w(f)$, $\mathcal{MPC}(f)$

Based on Corollary 5.1, we can define following sets

**Definition 5.6**

1. The set of non-decreasing functions that are majorized by $f$ is denoted by
$$\begin{aligned}\mathcal{MPS}(f) &= \text{MPS}(f) \cap \{g \in L^1 \mid g \text{ is non-decreasing}\} \\ &= \{g \in L^1 \mid g \text{ is non-decreasing and } g \prec f\}\end{aligned}$$
2. The set of non-negative, non-decreasing functions that are weakly majorized by $f$ is denoted by
$$\mathcal{MPS}_w(f) = \{g \in L^1 \mid g \text{ is non-negative, non-decreasing and } g \preceq f\}$$
3. The set of non-decreasing functions that majorize $f$ and satisfy $f(0) \leq g \leq f(1)$ is denoted by
$$\begin{aligned}\mathcal{MPC}(f) &= \text{MPC}(f) \cap \{g \in L^1 \mid g \text{ is non-decreasing and } f(0) \leq g \leq f(1)\} \\ &= \{g \in L^1 \mid g \text{ is non-decreasing and } g \succ f \text{ and } f(0) \leq g \leq f(1)\}\end{aligned}$$
where $f(0) \leq g \leq f(1)$ is used to ensure compactness.

### 5.3.2 Proposition: $\mathcal{MPS}(f)$, $\mathcal{MPS}_w(f)$, $\mathcal{MPC}(f)$ have extreme points and any element is a combination of extreme points

Following two propositions are the Proposition 1 of the Kleiner et al. (2021).

**Proposition 5.1 (Non-decreasing $f \Rightarrow \mathcal{MPS}(f)$, $\mathcal{MPS}_w(f)$, and $\mathcal{MPC}(f)$ have extreme points)**

Suppose $f \in L^1$ is non-decreasing. Then $\mathcal{MPS}(f)$, $\mathcal{MPS}_w(f)$, and $\mathcal{MPC}(f)$ are convex and compact in the norm topology $\Rightarrow$ (by Krein-Milman Theorem 5.1) they all have non-empty set of extreme points.

**Note** *We use* ext$A$ *to denote the set of extreme points of set* $A$*.*

**Proposition 5.2 (Non-decreasing $f \Rightarrow$ any distribution is a combination of extreme points)**

Suppose $f \in L^1$ is non-decreasing. For any $g \in \mathcal{MPS}(f)$, $\exists$ a probability measure $\lambda_g$ over ext$\mathcal{MPS}(f)$ such that

$$g = \int_{\text{ext}\mathcal{MPS}(f)} h \, d\lambda_g(h)$$

(also hold for any $g \in \mathcal{MPS}_w(f)$ and $g \in \mathcal{MPC}(f)$).

### 5.3.3 Extreme Points in $\mathcal{MPS}(f)$

**Theorem 5.4 (Form of Extreme Points in MPS$(f)$: Kleiner et al. (2021), Theorem 1)**

Let $f$ be non-decreasing. Then $g$ is an **extreme point** in $\mathcal{MPS}(f)$ <u>if and only if</u> there exists a collection of disjoint intervals $\{[\underline{x}_i, \overline{x}_i)\}_{i \in I}$ such that

$$g(x) = \begin{cases} f(x), & \text{if } x \notin \cup_{i \in I}[\underline{x}_i, \overline{x}_i) \\ \frac{\int_{\underline{x}_i}^{\overline{x}_i} f(s)ds}{\overline{x}_i - \underline{x}_i}, & \text{if } x \in [\underline{x}_i, \overline{x}_i) \end{cases}$$

$g$ is an extreme point of $\mathcal{MPS}(f)$ implies either that $g(x) = f(x)$ or that $g$ is constant at $x$.

**Definition 5.7 (Exposed Element)**

An element $x$ of a convex set $A$ is **exposed** if there exists a continuous linear functional that attains its maximum on $A$ uniquely at $x$.

**Note** *Every exposed point is extreme, but the converse is not true in general.*

**Corollary 5.2 (Kleiner et al. (2021), Corollary 1)**

Every extreme point of $\mathcal{MPS}(f)$ is exposed.

### 5.3.4 Extreme Points in $\mathcal{MPS}_w(f)$

For a set $A \subseteq [0,1]$, we use $\mathbf{1}_A(x)$ denote the indicator function of set $A$: it equals to 1 if $x \in A$ and 0 otherwise.

**Corollary 5.3 (Kleiner et al. (2021), Corollary 2)**

Suppose that $f$ is non-decreasing and non-negative. A function $g$ is an extreme point of $\mathcal{MPS}_w(f)$ <u>if and only if</u> there is $\theta \in [0,1]$ such that $g$ is an extreme point of $\mathcal{MPS}(f)$ and $g(x) = 0, \forall x \in [0,\theta)$.

### 5.3.5 Extreme Points in $\mathcal{MPC}(f)$

**Theorem 5.5 (Kleiner et al. (2021), Theorem 2)**

Let $f$ be non-decreasing and continuous. Then $g \in \mathcal{MPC}(f)$ is an extreme point of $\mathcal{MPC}(f)$ <u>if and only if</u> there exists a collection of intervals $[\underline{x}_i, \bar{x}_i)$, (potentially empty) sub-intervals $\left[\underline{y}_i, \bar{y}_i\right) \subseteq [\underline{x}_i, \bar{x}_i)$, and numbers $v_i$ indexed by $i \in I$ such that for all $x \in [0,1]$,

$$g(x) = \begin{cases} f(x) & \text{if } x \notin \bigcup_{i\in I} [\underline{x}_i, \bar{x}_i) \\ f(\underline{x}_i) & \text{if } x \in \left[\underline{x}_i, \underline{y}_i\right) \\ v_i & \text{if } x \in \left[\underline{y}_i, \bar{y}_i\right) \\ f(\bar{x}_i) & \text{if } x \in [\bar{y}_i, \bar{x}_i) \end{cases} \tag{5.1}$$

Moreover, a function $g$ as defined in (5.1) is in $\mathcal{MPC}(f)$ if the following three conditions are satisfied:

$$\left(\bar{y}_i - \underline{y}_i\right) v_i = \int_{\underline{x}_i}^{\bar{x}_i} f(s)\mathrm{d}s - f(\underline{x}_i)\left(\underline{y}_i - \underline{x}_i\right) - f(\bar{x}_i)(\bar{x}_i - \bar{y}_i) \tag{5.2}$$

$$f(\underline{x}_i)(\bar{y}_i - \underline{x}_i) + f(\bar{x}_i)(\bar{x}_i - \bar{y}_i) \leq \int_{\underline{x}_i}^{\bar{x}_i} f(s)\mathrm{d}s \leq f(\underline{x}_i)\left(\underline{y}_i - \underline{x}_i\right) + f(\bar{x}_i)\left(\bar{x}_i - \underline{y}_i\right) \tag{5.3}$$

If $v_i \in \left(f\left(\underline{y}_i\right), f(\bar{y}_i)\right)$, then for an arbitrary point $m_i$ satisfying $f(m_i) = v_i$ it must hold that

$$\int_{m_i}^{\bar{x}_i} f(s)\mathrm{d}s \leq v_i(\bar{y}_i - m_i) + f(\bar{x}_i)(\bar{x}_i - \bar{y}_i) \tag{5.4}$$

Condition (5.2) in the theorem ensures that $g$ and $f$ have the same integrals for each sub-interval $[\underline{x}_i, \bar{x}_i)$, analogously to the condition imposed in Theorem 5.3.3. Condition (5.3) ensures that $v_i \in (f(\underline{x}_i), f(\bar{x}_i))$, ensuring that $g$ is non-decreasing. If $f$ crosses $g$ in the interval $\left[\underline{y}_i, \bar{y}_i\right]$, then there is $m_i \in \left[\underline{y}_i, \bar{y}_i\right]$ such that $f(m_i) = v_i$. In this case, Condition (5.4) ensures that $\int_s^{\bar{x}_i} f(t)\mathrm{d}t \leq \int_s^{\bar{x}_i} g(t)\mathrm{d}t$ for all $s \in [\underline{x}_i, \bar{x}_i)$ and thus that $f \prec g$. If $v_i \notin \left(f\left(\underline{y}_i\right), f(\bar{y}_i)\right)$, Condition (5.3) is enough to ensure that $f \prec g$ and thus Condition (5.4) is not necessary.

# Chapter 6 Bayesian Persuasion: Bi-Pooling

Based on

- ★ Arieli, I., Babichenko, Y., Smorodinsky, R., & Yamashita, T. (2023). Optimal persuasion via bi-pooling. *Theoretical Economics*, 18(1), 15-36.
- Gentzkow, Matthew and Emir Kamenica (2016), "A Rothschild-Stiglitz approach to Bayesian persuasion." *American Economic Review*, 106, 597-601.
- Kolotilin, Anton (2018), "Optimal information disclosure: A linear programming ap-proach." *Theoretical Economics*, 13, 607-635.

## 6.1 Persuation Model

Consider a persuation model where the state space is the interval $[0, 1]$ with a common prior $F \in \Delta([0, 1])$ that has full support (i.e., $[0, 1]$ is the smallest closed set that has probability one). The sender knows the realized state and the receiver is uninformed.

1. Singaling: Prior to the realization of the state, the sender commits to a **signaling policy**

$$\pi : [0, 1] \to \Delta(S)$$

   where $S$ is an arbitrary measurable space. Once the state $\omega \in [0, 1]$ is realized, the sender sends a signal $s \in S$ to the receiver based on the committed signaling policy, i.e., $s \sim \pi(\omega)$. Without loss of generality, we may assume that $S = [0, 1]$, and that the posterior mean of the state, given signal $s$, is $s$ itself.
   Hence, the distribution of the posterior mean $s$ given the signal policy $\pi$, denoted by $F_\pi \in \Delta([0, 1])$ is a *mean-preserving contraction* of $F$.
   It is also easy to note that for any $G \in \text{MPC}(F)$, there exists a signaling policy $\pi$ (may not be unique) that makes $F_\pi = G$ (e.g., Gentzkow and Kamenica(2016), Kolotilin (2018)).
2. Persuasion problem: The sender's indirect utility is denoted by $u : [0, 1] \to \mathbb{R}$, where $u(x)$ is the sender's expected utility in case the receiver's posterior mean is $x$. $u$ is assume to be upper semicontinuous. $(F, u)$ is referred as a **persuation problem**. The sender's problem takes the form:

$$\max_{G \in \text{MPC}(F)} \mathbb{E}_{x \sim G}[u(x)]$$

## 6.2 Bi-Pooling

### 6.2.1 Bi-pooling Distribution

**Note** *For a distribution $H \in \Delta([0,1])$ and a measurable set $C \subseteq [0,1]$ we denote by $H|_C$ the distribution of $h \sim H$ conditional on the event that $h \in C$.*

**Definition 6.1 (Bi-pooling Distribution (Arieli et al. (2023), Definition 1))**

A distribution $G \in \text{MPC}(F)$ is called a **bi-pooling distribution** (with respect to $F$) if there exists a collection of pairwise disjoint open intervals $\{(\underline{y}_i, \overline{y}_i)\}_{i\in A}$ such that

- For every $i \in A$,

$$G((\underline{y}_i, \overline{y}_i)) = F((\underline{y}_i, \overline{y}_i))$$

  where $G((\underline{y}_i, \overline{y}_i)) = G(\overline{y}_i) - G(\underline{y}_i) = \int_{\underline{y}_i}^{\overline{y}_i} g(x)dx$, $F((\underline{y}_i, \overline{y}_i)) = F(\overline{y}_i) - F(\underline{y}_i) = \int_{\underline{y}_i}^{\overline{y}_i} f(x)dx$.
- The remaining intervals are the same:

$$G|_{[0,1]\setminus \cup_{i\in A}(\underline{y}_i, \overline{y}_i)} = F|_{[0,1]\setminus \cup_{i\in A}(\underline{y}_i, \overline{y}_i)}$$

- For every $i \in A$,

$$|\text{supp}(G|_{(\underline{y}_i, \overline{y}_i)})| \leq 2$$

  which means there are at most two different values of $G$ over $(\underline{y}_i, \overline{y}_i)$. If $|\text{supp}(G|_{(\underline{y}_i, \overline{y}_i)})| = 2$, we call $(\underline{y}_i, \overline{y}_i)$ a **bi-pooling interval**; If $|\text{supp}(G|_{(\underline{y}_i, \overline{y}_i)})| = 1$, we call $(\underline{y}_i, \overline{y}_i)$ a **pooling interval**. In the case where all intervals are pooling intervals, we say that $G$ is a **pooling distribution** (with respect to $F$).

**Example 6.1**

Consider the persuasion problem $(F, u)$, where $F = U[0,1]$ is the uniform distribution over $[0,1]$ and $u : [0,1] \to \mathbb{R}$ is an arbitrary function satisfying $u(\frac{1}{3}) = u(\frac{2}{3}) = 0$ and $u(x) < 0, \forall x \notin \{\frac{1}{3}, \frac{2}{3}\}$. Consider using a binary signal space $S = \{s_1, s_2\}$, where $s_1$ is sent with probability 1 over the interval $(\frac{1}{12}, \frac{7}{12})$ and $s_2$ is sent with probability 1 over the interval $[0, \frac{1}{12}] \cup [\frac{7}{12}, 1]$. This policy is a bi-pooling policy for the singleton collection $\{[0,1]\}$.

## 6.3 Applying Bi-pooling Distributions to Persuasion Problems

### 6.3.1 It works for all

**Theorem 6.1 (Arieli et al. (2023), Theorem 1)**

Every persuasion problem $(F, u)$ admits an optimal bi-pooling distribution.

**Proposition 6.1 (Arieli et al. (2023), Proposition 1)**

The set of extreme points of MPC$(F)$ is precisely the set of bi-pooling distributions.

**Theorem 6.2 (Arieli et al. (2023), Theorem 2)**

For every bi-pooling distribution $G \in$ MPC$(F)$ there exists a continuous utility function $u$ for which $G$ is the unique optimal solution of $\max_{G \in \text{MPC}(F)} \mathbb{E}_{x \sim G}[u(x)]$. That is, every extreme point of MPC$(F)$ is exposed.

### 6.3.2 How it works

**Definition 6.2 (Bi-pooling Policy (Arieli et al. (2023), Definition 3))**

A signaling policy $\pi$ is called a **bi-pooling policy** if there exists a collection of pairwise disjoint intervals $\{(\underline{y}_i, \overline{y}_i)\}_{i \in A}$ such that

- for every state $\omega \in (\underline{y}_i, \overline{y}_i)$ we have $\text{supp}(\pi(\omega)) \subseteq \{\underline{z}_i, \overline{z}_i\}$ (either $\pi(\omega) = \overline{z}_i$ or $\pi(\omega) = \underline{z}_i$) for some $\underline{z}_i \leq \overline{z}_i$ and $\underline{z}_i, \overline{z}_i \in [\underline{y}_i, \overline{y}_i]$;
- for every $\omega \notin \cup_{i \in A}(\underline{y}_i, \overline{y}_i)$, the policy sends the signal $\pi(\omega) = \omega$ (i.e., it reveals the state).

In the case where $\underline{z}_i = \overline{z}_i$ for all $i \in A$, we refer to $\pi$ as a **pooling policy**.

**Definition 6.3 (Monotonic Signaling Policy (Arieli et al. (2023), Definition 4))**

A (possibly mixed) signaling policy, $\pi : [0, 1] \to \Delta([0, 1])$, is **monotonic** if

$$\pi(x) \text{ first-order stochastically dominates } \pi(y) \text{ for every } x \geq y.$$

**Proposition 6.2 (Arieli et al. (2023), Proposition 2)**

Every persuasion problem admits an optimal (mixed) monotonic signaling policy.

**Lemma 6.1 (Arieli et al. (2023), Lemma 3)**

A persuasion problem $(F, u)$ admits an optimal pure monotonic signaling policy <u>if and only if</u> it admits an optimal pooling policy.

**Definition 6.4 (Double-Interval Nested Structure)**

A pure signaling policy: for each bi-pooling interval $(\underline{y}_i, \overline{y}_i)$, we can find a sub-interval $(\underline{w}_i, \overline{w}_i) \subseteq (\underline{y}_i, \overline{y}_i)$ such that $\pi$ is constant over the interval $(\underline{w}_i, \overline{w}_i)$ as well as over its complement $(\underline{y}_i, \overline{y}_i) \backslash (\underline{w}_i, \overline{w}_i)$.

**Corollary 6.1 (Arieli et al. (2023), Corollary 2)**

Every persuasion problem $(F, u)$ has an optimal bi-pooling policy that has a double-interval nested structure.

# Chapter 7 Optimization Methods

## 7.1 Generalized Neyman-Pearson Lemma

Based on

- Chernoff, H., & Scheffe, H. (1952). A generalization of the Neyman-Pearson fundamental lemma. *The Annals of Mathematical Statistics*, 213-225.
- Dantzig, G. B., & Wald, A. (1951). On the fundamental lemma of Neyman and Pearson. *The Annals of Mathematical Statistics*, 22(1), 87-93.

Given

- $n+m$ real integrable functions $g_1, ..., g_n, f_1, ..., f_m$ of a point $x$ in a Euclidean space $X$;
- a real function $\phi : \mathbb{R}^n \to \mathbb{R}$ of $n$;
- and $m$ constants $c_1, ..., c_m$.

The problem considered is about the existence, necessary conditions, and sufficient conditions of

$$\begin{aligned} S_0 = \arg\max_{S \subset X} \phi \left( \int_S g_1 dx, ..., \int_S g_n dx \right) \\ \text{s.t. } \int_S f_i dx = c_i, i = 1, ..., m \end{aligned}$$

Notations: $y_j(S) \triangleq \int_S f_j dx, j = 1, ..., m$ and $z_i(S) \triangleq \int_S g_i dx, i = 1, ..., n$.

### 7.1.1 The Neyman-Pearson Lemma

The Neyman-Pearson lemma refers to the case $n = 1, \phi(z_1) = z_1$, $X$ is 1-dimensional Euclidean space.

$$\begin{aligned} &\max_{S \subset X} \int_S g(x) dx \\ \text{s.t. } &\int_S f_i(x) dx = c_i, i = 1, ..., m \end{aligned} \tag{S1}$$

# Chapter 8 Calculus of Variations

Based on:

- Advanced Mathematical Economics Paulo B. Brito PhD in Economics Lecture 4 3.11.2021
- Minimization and Constraints of Partial Differential Equations Cathal Ormond.
- 

Calculus of variations primarily focuses on finding functions that make certain integral expressions reach their maximum or minimum values.

## 8.1 Generalized Calculus

### 8.1.1 Functional

**Definition 8.1 (functional)**

A **functional** is a mapping between a normed vector space (e.g. a space of functions) and the space of real numbers.

**Example 8.1**

Specifically, the input of a functional can be a function, for example,

$$F(y) := \int_a^b y(x)dx$$

where $y \in X \subseteq \mathbb{R} \to \mathbb{R}$ is a function in the space of functions $\mathcal{Y}$ which map $X \subseteq \mathbb{R} \to \mathbb{R}$. The $F$ is the functional between $\mathcal{Y}$ and $\mathbb{R}$.

### 8.1.2 Gâteux Derivative

We consider a functional $F$ over the space of functions $\mathcal{Y}$ in the following.

The variation of the functional is denoted as

$$\Delta F(y) = F(y + dy) - F(y)$$

In particular, the variation of the functional in the direction $h(x) \in \mathcal{Y}$ is

$$DF(y) = F(y + \epsilon h) - F(y)$$

**Definition 8.2 (Gâteaux Derivative / First Variation)**

The ***Gâteaux derivative*** *(or the* ***first variation****) of the functional* $F$ *at* $y$ *in the direction* $h(x) \in \mathcal{Y}$ is defined as the variation of the functional in the direction $h(x) \in \mathcal{Y}$ when the constant $\epsilon$ is infinitesimal

$$\begin{aligned}\delta F(y,h) :&= \lim_{\epsilon \to 0} \frac{F(y+\epsilon h) - F(y)}{\epsilon} \\ &= \frac{d}{d\epsilon} F(y+\epsilon h)\bigg|_{\epsilon=0}\end{aligned}$$

**Corollary 8.1 (Corollary to Riesz-Frechet theorem (Riesz and Sz.-Nagy, 1955, p. 61))**

If we consider $\mathcal{Y}$ as a *space of distributions*, we can represent the Gâteaux derivative as a linear functional as regards any perturbation $h(\cdot)$ by

$$\delta F(y,h) = \int_X \frac{dF(y)}{dy(x)} h(x)dx$$

**Definition 8.3 (Second-order Gâteaux derivative)**

The second-order Gâteaux derivative associated to perturbations $h_1(x)$ and $h_2(x)$ is given by

$$\begin{aligned}\delta F(y,h_1,h_2) :&= \lim_{\epsilon_2 \to 0} \lim_{\epsilon_1 \to 0} \frac{F(y+\epsilon_1 h_1 + \epsilon_2 h_2) - F(y+\epsilon_1 h_1) - F(y+\epsilon_2 h_2) - F(y)}{\epsilon_1 \epsilon_2} \\ &= \int_X \int_X \frac{d^2 F(y)}{dy(x_1)dy(x_2)} h_1(x_1) h_2(x_2) dx_1 dx_2\end{aligned}$$

Specifically, we write $\delta^2 F(y,h) := \delta F(y,h,h)$.

From now on we will consider the following types of functionals which are common in economics.

### 8.1.3 Linear Functionals

Consider the linear functional

$$F(y) = \int_X f(y(x))dx$$

where $f(\cdot)$ is assumed to be smooth and the integral exists, and

$$G(y) = g(F(y)) = g\left(\int_X f(y(x))dx\right)$$

Then, the Gâteaux derivatives of these two functionals are

$$\begin{aligned}\delta F(y,h) &= \int_X \frac{dF(y)}{dy(x)} h(x)dx = \int_X \frac{df(y(x))}{dy} h(x)dx \\ \delta G(y,h) &= \int_X \frac{dG(y)}{dy(x)} h(x)dx = \int_X g'(F(y)) \frac{df(y(x))}{dy} h(x)dx\end{aligned}$$

and the second-order Gâteux derivatives are

$$\delta^2 F(y,h) = \int_X \frac{d^2F(y)}{dy(x)^2} h^2(x)dx = \int_X \frac{d^2 f(y(x))}{dy^2} h(x)^2 dx$$

$$\delta^2 G(y,h) = \int_X \frac{d^2G(y)}{dy(x)^2} h^2(x)dx = \int_X \left[ g''(F(y)) \left( \frac{df(y(x))}{dy} \right)^2 + g'(F(y)) \frac{d^2 f(y(x))}{dy^2} \right] h(x)^2 dx$$

### 8.1.4 Functionals involving first-order derivatives

Consider the functional that involves first-order derivatives

$$F(y) = \int_{\underline{x}}^{\bar{x}} f(x, y(x), y'(x))dx$$

where $f(\cdot)$ is continuous and continuous differentiable in $(y, y')$. The Gâteux derivative is given by

$$\begin{aligned}
\delta F(y,h) &= \frac{d}{d\epsilon} F(y + \epsilon h) \bigg|_{\epsilon=0} \\
&= \frac{d}{d\epsilon} \int_X f(x, y(x) + \epsilon h(x), y'(x) + \epsilon h'(x))dx \bigg|_{\epsilon=0} \\
&= \int_{\underline{x}}^{\bar{x}} \left[ \frac{\partial f(x, y(x), y'(x))}{\partial y} h(x) + \frac{\partial f(x, y(x), y'(x))}{\partial y'} h'(x) \right] dx \\
&= \int_{\underline{x}}^{\bar{x}} \frac{\partial f(x, y(x), y'(x))}{\partial y} h(x)dx + \int_{\underline{x}}^{\bar{x}} \frac{\partial f(x, y(x), y'(x))}{\partial y'} dh(x)
\end{aligned}$$

By integration by parts, the second integral can be written as

$$\int_{\underline{x}}^{\bar{x}} \frac{\partial f(x, y(x), y'(x))}{\partial y'} dh(x) = \frac{\partial f(x, y(x), y'(x))}{\partial y'} h(x) \bigg|_{\underline{x}}^{\bar{x}} - \int_{\underline{x}}^{\bar{x}} h(x) d\frac{\partial f(x, y(x), y'(x))}{\partial y'}$$

Therefore,

$$\delta F(y,h) = \int_{\underline{x}}^{\bar{x}} \left[ \frac{\partial f(x, y(x), y'(x))}{\partial y} - \frac{d}{dx} \left( \frac{\partial f(x, y(x), y'(x))}{\partial y'} \right) \right] h(x)dx + \frac{\partial f(x, y(x), y'(x))}{\partial y'} h(x) \bigg|_{\underline{x}}^{\bar{x}}$$

**Note** *By choosing the $h(\cdot)$ that is a differentiable function vanishing on its boundary, i.e., $h(\underline{x}) = h(\bar{x}) = 0$, we have*

$$\delta F(y,h) = \int_{\underline{x}}^{\bar{x}} \left[ \frac{\partial f(x, y(x), y'(x))}{\partial y} - \frac{d}{dx} \left( \frac{\partial f(x, y(x), y'(x))}{\partial y'} \right) \right] h(x)dx \tag{GD}$$

*That is,* $\frac{dF(y^*)}{dy(x)} = \frac{\partial f(x,y(x),y'(x))}{\partial y} - \frac{d}{dx}\left(\frac{\partial f(x,y(x),y'(x))}{\partial y'}\right)$.

## 8.2 Optimization of Functionals

### 8.2.1 Extremes of Functionals

Given an extreme $y^* \in \mathcal{Y}$ of functional $F(y)$, locally we have

$$\delta F(y^*, h) = 0, \forall h \in \mathcal{Y}$$

According to $\delta F(y,h) = \int_X \frac{dF(y)}{dy(x)} h(x)dx$, we have that

$$\frac{dF(y^*)}{dy(x)} = 0, \forall x \tag{N1}$$

is a necessary condition for a maximum.

Since the maximum requires $F[y^*] \geq F[y]$ for all $y \in \mathcal{Y}$, by the generalized Taylor expansion that $F(y+\epsilon h) = F(y) + \delta F(y,h)\epsilon + \frac{1}{2}\delta^2 F(y,h)\epsilon^2 + o(\epsilon^2)$,

$$\delta^2 F(y^*,h) \leq 0, \forall h \in \mathcal{Y} \tag{N2}$$

is a necessary condition for a maximum.

### 8.2.2 Euler-Lagrange Equation

**Proposition 8.1 (Euler-Lagrange Equation)**

Let $y^* : \mathbb{R} \to \mathbb{R}$ be an extremum of the functional that involves first-order derivatives

$$F(y) = \int_{\underline{x}}^{\bar{x}} f(x, y(x), y'(x))dx$$

where $f(\cdot)$ is continuous and continuous differentiable in $(y, y')$. Then, $y^*$ must satisfy the **Euler-Lagrange Equation** for $f$, i.e.,

$$\frac{\partial f(x, y^*(x), y^{*\prime}(x))}{\partial y} = \frac{d}{dx}\left(\frac{\partial f(x, y^*(x), y^{*\prime}(x))}{\partial y'}\right) \text{ for each } x \in [\underline{x}, \bar{x}]$$

**Proof 8.1**

By choosing the $h(\cdot)$ that is a differentiable function vanishing on its boundary, i.e., $h(\underline{x}) = h(\bar{x}) = 0$, we have the equation (GD), where $\frac{dF(y^*)}{dy(x)} = \frac{\partial f(x,y(x),y'(x))}{\partial y} - \frac{d}{dx}\left(\frac{\partial f(x,y(x),y'(x))}{\partial y'}\right)$. Then, the Euler-Lagrange Equation is directly given by (N1).

### 8.2.3 Constrained Maximum of Functionals

**Problems with functional constraints** Consider the two functionals over function $y : X \to \mathbb{R}$, $F(y) = \int_X f(y(x))dx$ and $G(y) = \int_X g(y(x))dx$.

The optimization problem is

$$\begin{aligned} \max_{y(\cdot)} F(y) &:= \int_X f(y(x))dx \\ \text{s.t. } G(y) &:= \int_X g(y(x))dx = 0 \end{aligned} \tag{P1}$$

We can define a generalized Lagrangian functional

$$\begin{aligned} \mathcal{L}(y;\lambda) :&= F(y) + \lambda G(y) \\ &= \int_X L(y(x), \lambda)dx \end{aligned}$$

where $\lambda \in \mathbb{R}$ is the Lagrangian multiplier and $L(y(x), \lambda) := f(y(x)) + \lambda g(y(x))$ is a Lagrangian (function).

The necessary conditions for an optimum $y^*(x)$ are

$$\left.\frac{\partial \mathcal{L}(y; \lambda)}{\partial y(x)}\right|_{y^*(x)} = \frac{\partial f(y^*(x))}{\partial y} + \lambda \frac{\partial g(y^*(x))}{\partial y} = 0, \text{ for each } x \in X \tag{P1-N1}$$

and

$$\left.\frac{\partial \mathcal{L}(y; \lambda)}{\partial \lambda}\right|_{y^*(x)} = \int_X g(y^*(x))dx = 0 \tag{P1-N2}$$

**Problems with local constraints** Now consider the problem that has infinity of constraints, i.e., for each $x \in X$.

$$\begin{aligned} &\max_{y(\cdot), z(\cdot)} \int_X f(y(x), z(x))dx \\ &\text{s.t. } g(y(x), z(x)) = 0 \text{ for each } x \in X \end{aligned} \tag{P2}$$

Therefore, we introduce a Lagrangian function $\lambda : X \to \mathbb{R}$ instead of a Lagrange multiplier. The Lagrangian functional is

$$\mathcal{L}(y, z; \lambda) := \int_X \left[ f(y(x), z(x)) + \lambda(x) g(y(x), z(x)) \right] dx$$

The necessary conditions are

$$\left.\frac{\partial \mathcal{L}(y, z; \lambda)}{\partial y(x)}\right|_{y^*(x), z^*(x)} = \frac{\partial f}{\partial y}(y^*(x), z^*(x)) + \lambda(x) \frac{\partial g}{\partial y}(y^*(x), z^*(x)) = 0, \text{ for each } x \in X$$

$$\left.\frac{\partial \mathcal{L}(y, z; \lambda)}{\partial z(x)}\right|_{y^*(x), z^*(x)} = \frac{\partial f}{\partial z}(y^*(x), z^*(x)) + \lambda(x) \frac{\partial g}{\partial z}(y^*(x), z^*(x)) = 0, \text{ for each } x \in X$$

$$\left.\frac{\partial \mathcal{L}(y, z; \lambda)}{\partial \lambda(x)}\right|_{y^*(x), z^*(x)} = g(y^*(x), z^*(x)) = 0, \text{ for each } x \in X$$

# Chapter 9 Politics Models

## 9.1 Voting Model: Implicit Function Theroem

Consider an incumbent $I$ and a citizen/voter $v$.

- $I$ picks $x_1 \in \mathbb{R}$;
- $v$ observes $u_1 = -x_1^2 + \epsilon$, where $\epsilon \sim f$ and $f$ is uninormal of 0, symmetric, continuous, and differentiable. $f'(z)$ is positive for $z < 0$, negative for $z > 0$, and zero for $z = 0$.
- $v$ re-elects or not
- (new) $I$ chooses $x_2$
- ...

### 9.1.1 Case 1

Incumbents have $\alpha \in (0, 1)$ probability to be "good" type who picks $x_1 = x_2 = 0$ and $1 - \alpha$ probability to be "bad" type who picks $\hat{x} = x_1 = x_2 > 0$.

Bayesian posterior beliefs are

$$\Pr(\text{good} \mid u_1) = \frac{\alpha f(u_1)}{\alpha f(u_1) + (1 - \alpha) f(u_1 + \hat{x}^2)}$$

where $\Pr(\text{good} \mid u_1) \geq \alpha \Leftrightarrow f(u_1) \geq f(u_1 + \hat{x}^2)$.

By our assumption about $f$, $f(u_1) \geq f(u_1 + \hat{x}^2)$ means $u_1$ is closer to zero than $u_1 + \hat{x}^2 \Rightarrow u_1^2 \leq (u_1 + \hat{x}^2)^2 = u_1^2 + 2u_1\hat{x}^2 + \hat{x}^4$, that is, $u_1 > -\frac{\hat{x}^2}{2}$.

### 9.1.2 Case 2: Moral Hazard Version

All incumbents are "bad": ideal policy is 1. Assume voters re-elect if and only if $u_1 \geq k$, where $k$ is endogenous. Based on this rule, the probability of an incumbent being re-elected is

$$\Pr(\text{re-elect}|x_1) = \Pr(-x_1^2 + \epsilon \geq k) = 1 - F(k + x_1^2)$$

Suppose the utility of the incumbent is

$$U_I(x_1, x_2) = w - (1 - x_1)^2 + \delta(w - (1 - x_2)^2)\mathbf{1}_{\text{re-elect}}$$

Specifically, the expected utility with $u_2 = 1$ is

$$U_I(x_1, x_2 = 1) = w - (1 - x_1)^2 + \delta w \left[1 - F(k + x_1^2)\right]$$

Then, $x_1^*$ should solve

$$\frac{\partial U_I}{\partial x_1} = 2(1-x_1) - 2\delta w x_1 f(k+x_1^2) = 0$$
$$\Rightarrow f(k+x_1^2) = -\frac{1}{\delta w} + \frac{1}{x_1}\frac{1}{\delta w}$$

**Apply Implicit Function Theorem**

Let $g(k,x) = f(k+x_1^2) + \frac{1}{\delta w} - \frac{1}{x_1}\frac{1}{\delta w}$.

The goal of the voter is to find the $k$ that minimizes $x_1^*$. By the implicit function theorem

$$\frac{\partial x_1^*}{\partial k} = -\frac{\frac{\partial g}{\partial k}\big|_{x_1^*}}{\frac{\partial g}{\partial x}\big|_{x_1^*}}$$

As $\frac{\partial g}{\partial k} = f'(k+x_1^2)$ and $\frac{\partial g}{\partial x} = 2x_1 f'(k+x_1^2) + \frac{1}{x_1^2}\frac{1}{\delta w}$, we can conclude the optimal $k$ satisfies $k = -x_1^{*2}$. Then, $f(0) = -\frac{1}{\delta w} + \frac{1}{x_1^*}\frac{1}{\delta w} \Rightarrow$

$$x_1^* = \frac{1}{1+\delta w f(0)},\; k^* = -\left(\frac{1}{1+\delta w f(0)}\right)^2$$

### 9.1.3 Case 3

Suppose the incumbent has probability $\alpha$ being "good" with $y_I = 0$ and probability $1-\alpha$ being "bad" with $y_I = 1$. He chooses $x_2 = y_I$ at stage 2.

Given the strategy $x_g$ and $x_b$ Bayesian posterior beliefs are

$$\Pr(\text{good} \mid u_1) = \frac{\alpha f(x_g^2+u_1)}{\alpha f(x_g^2+u_1) + (1-\alpha) f(x_b^2+u_1)}$$

Hence, $\Pr(\text{good} \mid u_1) \geq \alpha$ if and only if $f(x_g^2+u_1) \geq f(x_b^2+u_1)$.

The voter's strategy is also represented by "re-elect" iff $u_1 \geq k$. At the critical point $u_1 = k$,

$$f(x_g^2+k) = f(x_b^2+k) \Rightarrow k = -\frac{x_g^2+x_b^2}{2}$$

Suppose the expected utility (constructed based on avoiding deviations from the incumbent's true type) of the incumbent is

$$\mathbb{E}U_I(x_1, x_2 = y_I) = w - (x_1 - y_I)^2 + \delta w\left(1 - F(k+x_1^2)\right)$$

Obviously, $x_1^* = 0$ for good incumbent. (i.e., $x_g = 0$). Then, $k = -\frac{x_b^2}{2}$, and

$$\mathbb{E}U_b(x_1) = w - (x_1-1)^2 + \delta w\left(1 - F(k+x_1^2)\right)$$

which has derivative

$$-2(x_1-1) - 2\delta w x_1 f(k+x_1^2)$$

So, the optimal $x_1^*$ of "bad" type should satisfy

$$f(k+x_1^2)+\frac{1}{\delta w}-\frac{1}{\delta w x_1}=0$$

Consider the $x_1=\sqrt{-2k}$ (by what we induced, $k=-\frac{x_b^2}{2}$), the optimal $k$ should be solved by

$$\begin{aligned}H(k)&=f(-k)+\frac{1}{\delta w}-\frac{1}{\delta w\sqrt{-2k}}\\&=f(k)+\frac{1}{\delta w}-\frac{1}{\delta w\sqrt{-2k}}=0\end{aligned}$$

By our assumption about $f$, $f(k)=f(-k)$.

Also, by the implicit function theorem, we can analyze how the $w$ affects $k$

$$\frac{\partial k}{\partial w}=-\frac{\frac{\partial H}{\partial w}\big|_{k^*}}{\frac{\partial H}{\partial k}\big|_{k^*}}$$

## 9.2 Two Period Accountability Model: Normal-Normal Learning

### 9.2.1 Normal-Normal Learning

Suppose $\theta$ has a prior $N(\mu_\theta,\sigma_\theta^2)$. We observe $s=\theta+\varepsilon$, where $\varepsilon\sim N(0,\sigma_\varepsilon^2)$.

**Proposition 9.1 (Normal-Normal Learning)**

The posterior beliefs about $\theta$ given $s$ is also normal with mean $\mu_1=\lambda\mu_\theta+(1-\lambda)s$ and variance $\lambda\sigma_\theta^2$,

$$\theta\mid s\sim N(\lambda\mu_\theta+(1-\lambda)s,\lambda\sigma_\theta^2)$$

where $\lambda=\frac{\sigma_\theta^{-2}}{\sigma_\theta^{-2}+\sigma_\varepsilon^{-2}}=\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_\theta^2}$ is the precision weight.

### 9.2.2 Two Period Accountability Model

1. Nature chooses $\theta\in\mathbb{R}$, which follows distribution $N(\mu_\theta,\sigma_\theta^2)$.
2. Incumbent takes the first action $a_1\geq 0$.
3. All observe $y_1=\theta+a_1+\epsilon_1$, where $\epsilon_1\sim N(0,\sigma_\epsilon^2)$.
4. Citizens choose $s_1\in\mathbb{R}$.
5. Incumbent takes the second action $a_2\geq 0$.
6. Citizens observe $a_1$ and $y_2=\theta+a_2+\epsilon_2$.
7. Citizens choose $s_2\in\mathbb{R}$.

The utility of the incumbent is

$$U_I=s_1-ka_1^2+s_2-ka_2^2,k>0$$

and the utility of the citizens is

$$U_C = y_1 - (s_1 - \theta)^2 + y_2 - (s_2 - \theta)^2$$

1. **Period 1 Belief:** Given $y_1$ and the conjecture $\tilde{a}_1$, we have

$$y_1 - \tilde{a}_1 = \theta + \epsilon_1$$

Based on the normal-Normal learning (9.1), the posterior belief about $\theta$ is

$$N(\underbrace{\lambda_1\mu_\theta + (1-\lambda_1)(y_1 - \tilde{a}_1)}_{\bar{\mu}_\theta}, \underbrace{\lambda_1\sigma_\theta^2}_{\bar{\sigma}_\theta^2})$$

where $\lambda_1 = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_\theta^2}$

2. **Period 2 Belief:** Given $y_2$, $a_1$ (substitute $\tilde{a}_1$ in $\bar{\mu}_\theta$ and $\bar{\sigma}_\theta^2$), and the conjecture $\tilde{a}_2$, we have

$$y_2 - \tilde{a}_2 = \theta + \epsilon_2$$

Based on the normal-Normal learning (9.1), the posterior belief about $\theta$ is

$$N(\underbrace{\lambda_2\bar{\mu}_\theta + (1-\lambda_2)(y_2 - \tilde{a}_2)}_{\bar{\bar{\mu}}_\theta}, \underbrace{\lambda_2\bar{\sigma}_\theta^2}_{\bar{\bar{\sigma}}_\theta^2})$$

where $\lambda_1 = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\bar{\sigma}_\theta^2}$

The optimal $s_2^* = \bar{\bar{\mu}}_\theta$. Then,

$$\begin{aligned}
U_{I,2} &= s_2^* - ka_2^2 \\
&= \lambda_2\bar{\mu}_\theta + (1-\lambda_2)(y_2 - \tilde{a}_2) - ka_2^2 \\
&= \lambda_2\bar{\mu}_\theta + (1-\lambda_2)(\theta + a_2 + \epsilon_2 - \tilde{a}_2) - ka_2^2 \\
\frac{\partial U_{I,2}}{\partial a_2} &= 1 - \lambda_2 - 2ka_2 \\
a_2^* &= \frac{1-\lambda_2}{2k}
\end{aligned}$$

Similarly,

$$a_1^* = \frac{1-\lambda_2}{2k} > \frac{1-\lambda_2}{2k}$$

## 9.3 Motivated Beliefs

1. The objective probability distribution is $\Pi = (\pi_1, \pi_2, ..., \pi_n)$;
2. The motivated belief $\Pi' = (\pi_1', \pi_2', ..., \pi_n')$ maximizes

$$f(\Pi') = \underbrace{-\alpha D_{KL}\left(\Pi' \| \Pi\right)}_{\text{accuracy}} + \underbrace{v(\Pi')}_{\text{directional}}$$

$$\text{s.t. } g(\Pi') = 1 - \sum_{i=1}^{n} \pi_i' = 0$$

where $D_{KL}(\Pi' \| \Pi) \triangleq \sum_{i=1}^n \pi_i' \log\left(\frac{\pi_i'}{\pi_i}\right)$ is the KL-divergence.

The Lagrangian is

$$\begin{aligned} L(\Pi') &= f(\Pi') - \lambda g(\Pi') \\ &= -\alpha D_{KL}(\Pi' \| \Pi) + v(\Pi') - \lambda(1 - \sum_{i=1}^n \pi_i') \\ \frac{\partial L(\Pi')}{\partial \pi_i'} &= -\alpha\left(1 + \log\left(\frac{\pi_i'}{\pi_i}\right)\right) + \frac{\partial v(\Pi')}{\partial \pi_i'} + \lambda = 0 \end{aligned}$$

Let $v(\Pi') = \sum_{i=1}^n v_i \pi_i'$, then we have

$$\pi_i' = e^{\frac{\lambda}{\alpha}-1} e^{\frac{v_i}{\alpha}} \pi_i$$

By the constraint $1 - \sum_{i=1}^n \pi_i' = 0$, $e^{\frac{\lambda}{\alpha}-1} = \frac{1}{\sum_{j=1}^n e^{\frac{v_j}{\alpha}} \pi_j}$. Then,

$$\pi_i' = \frac{e^{\frac{v_i}{\alpha}} \pi_i}{\sum_{i=j}^n e^{\frac{v_j}{\alpha}} \pi_j}$$

### 9.3.1 Normal Distribution

Suppose there is a $\theta \sim N(\mu, \sigma^2)$, the real density is

$$f(\theta) \propto e^{-\frac{1}{2}\left(\frac{\theta-\mu}{\sigma}\right)^2}$$

The motivated density is

$$\begin{aligned} \tilde{f}(\theta) &= \underset{f'(\theta)}{\operatorname{argmax}} -D_{KL}(f' \| f) + \int_\theta v(\theta) f'(\theta) d\theta \\ \Rightarrow \tilde{f}(\theta) &= \frac{f(\theta) e^{v(\theta)}}{\int_{\theta'} f(\theta') e^{v(\theta')} d\theta'} \propto f(\theta) e^{v(\theta)} \end{aligned}$$

where $\int_{\theta'} f(\theta') e^{v(\theta')} d\theta'$ is assumed to be finite.

**Proof 9.1**

The optimization problem is

$$\begin{aligned} \max_{f'(\cdot)} & \int_{-\infty}^{\infty} \left[v(\theta) - \log\left(\frac{f'(\theta)}{f(\theta)}\right)\right] f'(\theta) d\theta \\ \text{s.t.} & \int_{-\infty}^{\infty} f'(\theta) d\theta = 1 \end{aligned}$$

The generalized Lagrangian functional can be defined as

$$\mathcal{L}(f'; \lambda) = \int_{-\infty}^{\infty} \left[v(\theta) - \log\left(\frac{f'(\theta)}{f(\theta)}\right)\right] f'(\theta) d\theta + \lambda\left(\int_{-\infty}^{\infty} f'(\theta) d\theta - 1\right)$$

The necessary conditions for a maximum $f^*$ is

$$\begin{aligned} \frac{\partial \mathcal{L}(f^*; \lambda)}{\partial f'(\theta)} &= -1 + v(\theta) - \log\left(\frac{f^*(\theta)}{f(\theta)}\right) + \lambda = 0, \text{ for each } \theta \in \mathbb{R} \\ \frac{\partial \mathcal{L}(f^*; \lambda)}{\partial \lambda} &= \int_{-\infty}^{\infty} f^*(\theta) d\theta - 1 = 0 \end{aligned}$$

Thus, we have

$$f^*(\theta) \propto f(\theta)e^{v(\theta)}$$

We take any quadratic $v(\theta) = v_0 + v_1\theta + v_2\theta^2$. Then,

$$\tilde{f}(\theta) \propto e^{-\frac{1}{2}\left(\frac{\theta-\mu}{\sigma}\right)^2+v_0+v_1\theta+v_2\theta^2} = ke^{-\frac{1}{2}\left(\frac{\theta-\mu_d}{\sigma_d}\right)^2}$$

where $\mu_d = \frac{v_1+\sigma^{-2}\mu}{\sigma^{-2}-2v_2}$, $\sigma_d = \left(\sigma^{-2} - 2v_2\right)^{-\frac{1}{2}}$, and $k$ is a constant that is not a function of $\theta$.

### 9.3.2 Accountability Model with Motivated Reasoning

There is an incumbent (with $\theta_I \sim N(\mu_I = 0, \sigma_\theta^2)$), finite set of voters and a (non-strategic) challenger (with $\theta_C$).

The incumbent takes action $e \geq 0$. Public signal is $s = \theta_I + e + \epsilon$, where $\epsilon \sim N(0, \sigma_\epsilon^2)$. Voters decide whether to retain the incumbent after observing $s$.

$$U_I(e, R) = R - c(e)$$

$$U_j(R) = s + a_j + R(\theta_I + a_j + v_I) + (1-R)(\theta_C + v_C)$$

$R = 1$ if the incumbent stays at $t = 2$, $R = 0$ otherwise.

$v_I, v_C$ are candidate-specific utility shocks common to all voters ($v_I - v_C$ has mean 0 and variance $\sigma_v^2$).

$a_j$ is the affinity of voter $j$ to the incumbent.

**Motivated Reasoning:** (a simpler version), the voter is maximizing $\log f_{\theta|s}(\tilde{\theta}_I|s) + \delta v(a_j, \tilde{\theta}_I)$.

Assumptions: weakly concave in $\tilde{\theta}_I$ and $\frac{\partial^2 v(a_j,\tilde{\theta}_I)}{\partial a_j \partial\tilde{\theta}_I} \geq 0$.

A more general version:

$$\tilde{f}(\theta) = \underset{f'(\theta)}{\text{argmax}} -D_{KL}\left(f' \| f\right) + \delta \int_\theta v(\theta)f'(\theta)d\theta$$

$$\Rightarrow \tilde{f}(\theta) = \frac{f(\theta)e^{\delta v(\theta)}}{\int_{\theta'} f(\theta')e^{\delta v(\theta')}d\theta'} \propto f(\theta)e^{\delta v(\theta)}$$

**Example 9.1**

Spatial Bias: $v(a_j, \theta_I) = -(a_j - \theta_I)^2$.

$\tilde{\mu}_I = \frac{1}{1+2\delta\sigma_\theta^2}\mu_I + \frac{2\delta\sigma_\theta^2}{1+2\delta\sigma_\theta^2}a_j$ and the variance is $\tilde{\sigma}_\theta^2 = \left(\sigma_\theta^2 + 2\delta\right)^{-1} < \sigma_\theta^2$.

Given the conjecture $\hat{e}$, the posterior belief of mean upon receiving $s$ is $\lambda\tilde{\mu}_I + (1-\lambda)(s - \hat{e})$.

A voter votes to re-elect if and only if: $\tilde{\mu}_I(s, a_j, \delta, \hat{e}) + a_j + v_I \geq \mu_C + v_C$

## 9.4 Stochastic Game

A “stochastic game” consist of:

1. A set of states $K$;
2. A set of players $N$;
3. Action for player $i$: $A_i(k)$ $(k \in K)$;
4. “Period Payoffs”: $u_i(A, k)$;
5. Law of motion: $\Pr(k_{t+1} \mid k_t, a_t)$, where $a_t$ is the action taken at $t$. (“Markov”)
6. Discount Rate $\delta$;
7. Utility for the entire game is given by
$$U_i = \sum_{t=1}^{\infty} \delta^{t-1} u_i(a_t, k_t)$$
8. History: $h_t \triangleq (a_1, k_1, ..., a_t, k_t)$ and the set of possible history is $H_t$.

**Definition 9.1 (Markovian Strategy)**

A **strategy** for $i$ is a mapping $\sigma_i : H_t \times K \to \Delta A_i(k)$ for all $t$.

A **Markovian strategy** is a mapping $\sigma_i : K \to \Delta A_i(k)$.

Game starting at $t$ is a subgame; A strategy profile $\sigma^*$ is a SPNE if all players play BR, starting at each $t$.

**Definition 9.2 (Markov Perfect Equilibrium)**

A $\sigma^*$ is a **Markov Perfect Equilibrium (MPE)** iff it is a SPNE satisfying Markovian.

### 9.4.1 Prison Dilemma as a stochastic game

Consider PD (Prison Dilemma) as a stochastic game,

1. $K = \{PD\}$;
2. $A_i(PD) = \{0, 1\}$, $u_i(a, PD) = 1 - a_{i,t} + 2a_{-i,t}$;
3. $\Pr(PD|PD, a_t) = 1$

Markov Strategy is defined by $\sigma_i = \Pr(a_{i,t} = 1) \in [0, 1]$.

In any MPE, the Markov Strategy at $t$ should maximize $U_i$ starting at $t'$,

$$\begin{aligned}\sigma_i &= \operatorname*{argmax}_{\sigma_i' \in [0,1]} \delta^{t'}(1 - \sigma_i' + 2\sigma_{-i}) + \sum_{t=t'+1}^{\infty} \delta^{t-1}(1 - \sigma_i + 2\sigma_{-i}) \\ &= 0\end{aligned}$$

This a SPNE and MPE.

### 9.4.2 Revised Prison Dilemma

1. $K = \{PD, WPD\}$;
2. $A_i(PD) = \{0, 1\}$, $u_i(a, PD) = 1 - a_{i,t} + 2a_{-i,t}$;
3. $A_i(WPD) = \{0, 1\}$, $u_i(a, WPD) = u_i(a, PD) - x$, where $x \in \mathbb{R}_+$;
4. $\Pr(k_{t+1} = WPD \mid k_t = WPD) = 1$;
5. $\Pr(k_{t+1} = WPD \mid k_t = PD, (1, 1)) = 0$;
6. $\Pr(k_{t+1} = WPD \mid k_t = PD, \{(0, 1), (1, 0), (0, 0)\}) = q, q \in [0, 1]$.

Markov Strategy is defined by $\sigma_i(k_t)$. Obviously, $\sigma_i^*(WPD) = 0$ in any MPE.

"Value function" $v(PD, \sigma)$ represents the net present value of starting a period in state $PD$ given $\sigma$. The most desirable situation that both players choose 1:

$$v(PD, \sigma^*) = 2 + \delta v(PD, \sigma^*) \Rightarrow v(PD, \sigma^*) = \frac{2}{1 - \delta}$$

Check one-period devotion from changing 1 to 0 at this stage:

$$\begin{aligned} v'(PD, \sigma^*) &= 3 + \delta \left[ qv(WPD, \sigma^*) + (1 - q)v(PD, \sigma^*) \right] \\ &= 3 + \delta \left[ q\frac{1 - x}{1 - \delta} + (1 - q)\frac{2}{1 - \delta} \right] \end{aligned}$$

This deviation is not profitable if

$$v'(PD, \sigma^*) \leq v(PD, \sigma^*)$$
$$\text{i.e. } q \geq \frac{1 - \delta}{(1 + x)\delta}$$

### 9.4.3 Dynamic Commitment Problem

1. $K = \{l, h, w_C, w_R\}$;
2. $N = \{C, R\}$;
3. In state $k \in \{l, h\}$:

   $R$ makes an offer $x_k \leq 1$;

   $C$ accepts ($R$ and $C$ get period payoffs $(1 - x_k, x_k)$ and $\Pr(k_{t+1} = h) = q$, $\Pr(k_{t+1} = l) = 1 - q$) or rejects ($R$ and $C$ get period payoffs $((1 - p_k)(1 - f), p_k(1 - f))$ and $\Pr(k_{t+1} = w_C) = p_k$, $\Pr(k_{t+1} = w_R) = 1 - p_k$), where $f \in (0, 1)$ and $0 < p_l < p_h < 1$.
4. If enter $w_C$ ($R$ and $C$ get period payoffs $(0, 1 - f)$); If enter $w_R$ ($R$ and $C$ get period payoffs $(1 - f, 0)$); Game over.

MPE

If offer accepted in $l$ and $h$, $x_k$:

$$v_C(l;p) = x_k + \delta\left(qv_C(h;p) + (1-q)v_C(l;p)\right)$$

$$= v_C(h;p) = x_k + \delta\left(qv_C(h;p) + (1-q)v_C(l;p)\right)$$

If offer rejected, $\frac{p_k(1-f)}{1-\delta}$.

In equilibrium $\frac{p_k(1-f)}{1-\delta} = v_C(l;p) = v_C(h;p)$, then we have

$$x_k = (p_k - \delta\bar{p})\frac{1-f}{1-\delta}, \text{ where } \bar{p} = qp_h + (1-q)p_l$$