Models in Political Economy

Juan Dodyk

Contents

1 Analysis

1.1 Supermodularity and complementarity

I follow Topkis (1998). Let X, T be partially ordered sets, $S \subset X \times T$, $S_t = \{x \mid (x,t) \in S\}$ for each $t \in T$, and $f : S \to \mathbb{R}$. We say that f has *increasing differences* if $f(\cdot, t') - f(\cdot, t)$ is increasing on $S_t \cap S_{t'}$ for all $t, t' \in T$ such that $t' \succ t$. Note that if we flip the role of X and T, the definition is equivalent. Let X_i be a partially ordered set for each $i \in I$, $X \subset \prod_{i \in I} X_i$, and $f : X \to \mathbb{R}$. We induce a partial order on X: for $x, x' \in X$, $x \succeq x'$ iff, for all $i \in I$, $x_i \succeq_i x'_i$. We say that f has *increasing differences* if, for each $i, j \in I$ such that $i \neq j$, and $\tilde{x} \in X$, $f|_{S_{\tilde{x}}}$, as a function of its projection on $X_i \times X_j$, has increasing differences, where $S_{\tilde{x}} = \{x \in X \mid \forall k \in I \setminus \{i,j\}.\, x_k = \tilde{x}_k\}$. If $X_i = \mathbb{R}$, X is open and convex, and f is twice differentiable, then f has increasing differences iff $\frac{\partial^2 f}{\partial x_i \partial x_j} \geqq 0$ for all $i \neq j$. We can similarly talk about *decreasing* differences, and *strict* increasing (decreasing) differences.

Let X be a set and $\preceq$ a partial order. The *supremum* (*infimum*) of $X' \subset X$ is its least upper bound (greatest lower bound). We say that X is a *lattice* if for each $x, x' \in X$ then $\{x, x'\}$ has a supremum $x \vee x'$ and an infimum $x \wedge x'$. We say that $X' \subset X$ is a *sublattice* if it is a lattice. We say that X is a *chain* if the order is complete. The lattice order on X induces a partial order on its sublattices: if S, S' are sublattices of X, we say that $S \sqsubseteq S'$ iff for all $x \in S, x' \in S'$ we have $x \wedge x' \in S$ and $x \vee x' \in S'$. (It is not entirely obvious that $\sqsubseteq$ is transitive and antisymmetric; the trick is to note that $x = x \vee (x \wedge x') = x \wedge (x \vee x')$.) If X is a lattice and $f : X \to \mathbb{R}$, we say that f is *supermodular* if $f(x) + f(x') \leqq f(x \vee x') + f(x \wedge x')$ for all $x, x' \in X$.

Proposition 1.1.1. Let $X_1, \ldots, X_n$ be lattices, X a sublattice of $\prod_{i=1}^n X_i$, and $f : X \to \mathbb{R}$. If f is supermodular then it has increasing differences. If, for each $i = 1, \ldots, n$ and $\tilde{x} \in X$, $f|_{S_{\tilde{x}}}$, as a function of its projection on X_i, is supermodular, where $S_{\tilde{x}} = \{x \in X \mid \forall j \in I \setminus \{i\}.\, x_j = \tilde{x}_j\}$, and f has increasing differences, then f is supermodular.

Proof. The first part is trivial. For the second, let $x, x' \in X$. We have

$$\begin{aligned}
&f(x \vee x') - f(x) = \\
&\quad = \sum_{i=1}^n \left(f(x_1 \vee x'_1, \ldots, x_i \vee x'_i, x_{i+1}, \ldots, x_n) - f(x_1 \vee x'_1, \ldots, x_{i-1} \vee x'_{i-1}, x_i, \ldots, x_n) \right) \geqq \\
&\quad \geqq \sum_{i=1}^n \big(f(x_1 \vee x'_1, \ldots, x_{i-1} \vee x'_{i-1}, x'_i, x_{i+1}, \ldots, x_n) - \\
&\qquad\quad - f(x_1 \vee x'_1, \ldots, x_{i-1} \vee x'_{i-1}, x_i \wedge x'_i, x_{i+1}, \ldots, x_n) \big) \geqq \\
&\quad \geqq \sum_{i=1}^n \left(f(x'_1, \ldots, x'_i, x_{i+1} \wedge x'_{i+1}, \ldots, x_n \wedge x'_n) - f(x'_1, \ldots, x'_{i-1}, x_i \wedge x'_i, \ldots, x_n \wedge x'_n) \right) = \\
&\quad = f(x') - f(x \wedge x'),
\end{aligned}$$

where the first inequality is by "point-wise" supermodularity and the second one by increasing differences. ∎

Proposition 1.1.2. If $f : X \to \mathbb{R}$ is supermodular on a lattice X, then $\arg\max f$ is a sublattice. If f is strictly supermodular, then $\arg\max f$ is a chain. If X is a compact sublattice of $\mathbb{R}^n$ and f is continuous, $\arg\max f$ has a least and a greatest element.

Proof. The first two claims are obvious. For the last, note that $A = \arg\max f$ is compact, non-empty, and a sublattice of $\mathbb{R}^n$. Let $\pi_i : \mathbb{R}^n \to \mathbb{R}$ be the projection to the i-th coordinate,

$x_i \in X$ such that $\pi_i(x_i)$ is minimum, and $x = \bigwedge_{i=1}^n x_i$; clearly $x = \min A \in X$. Similarly $\max A \in X$. ■

Theorem 1.1.3. If X is a lattice, T is a partially ordered set, for each $t \in T$, $S_t \subset X$ is a sublattice, $S_t \subset S_{t'}$ and $S_t \sqsubseteq S_{t'}$ for $t \preceq t'$, $S = \bigcup_{t\in T} S_t \times \{t\}$, $f : S \to \mathbb{R}$, $f(\cdot, t)$ is supermodular for each $t \in T$, and f has increasing differences on (x, t), then $A_t = \arg\max_{x\in S_t} f(x, t)$ is increasing in t on $\{t \in T : A_t \neq \varnothing\}$.

Proof. Let $t \preceq t'$, $A_t, A_{t'} \neq \varnothing$, $x \in A_t$, $x' \in A_{t'}$. Since $S_t \sqsubseteq S_{t'}$, $x \wedge x' \in S_t$ and $x \vee x' \in S_{t'}$; since $S_t \subset S_{t'}$, $x, x \wedge x' \in S_{t'}$. By the fact that $f(\cdot, t') - f(\cdot, t)$ is increasing on $S_t \cap S_{t'} = S_t$ and $f(\cdot, t')$ is supermodular, we have $0 \leqq f(x, t) - f(x \wedge x', t) \leqq f(x, t') - f(x \wedge x', t') \leqq f(x \vee x', t') - f(x', t') \leqq 0$, so $x \wedge x' \in A_t$ and $x \vee x' \in A_{t'}$, as desired. ■

If $S \subset \mathbb{R}^n$ is convex and f is twice differentiable, the sufficient condition is that $\frac{\partial^2 f}{\partial x_i \partial x_j} \geqq 0$ for all $i \neq j$ and $\frac{\partial^2 f}{\partial x_i \partial t_k} \geqq 0$ for all i, k.

1.2 Concave functions

Definition 1.2.1. Una función $f : A \to \mathbb{R}$, con $A \subset \mathbb{R}^n$ convexo, se dice *cóncava* si para todos $x, y \in A$ y $\lambda \in [0, 1]$ se cumple $f(\lambda x + (1 - \lambda)y) \geqq \lambda f(x) + (1 - \lambda)f(y)$.

Theorem 1.2.2. Sea $f : A \subset \mathbb{R}^n \to \mathbb{R}$ una función cóncava. Si $x_0 \in A^\circ$ entonces f es Lipschitz en un entorno de x_0, i.e., existe un abierto $\mathcal{U} \subset A$ con $x_0 \in \mathcal{U}$ y una constante $M > 0$ tales que, para todos $x, y \in \mathcal{U}$, $|f(x) - f(y)| \leqq M\|x - y\|$.

Proof. Primero veamos que es localmente acotada. Sea $\epsilon > 0$ tal que $B_{2\epsilon}(x_0) \subset A$, $S = \{x_0 \pm \epsilon e_i \mid i = 1, \ldots, n\}$, donde e_i es la fila i de la matriz I_n, y $m = \min_{x\in S}\{f(x)\}$. Tenemos que $f(x) \geqq m$ si $x \in \mathrm{co}(S)$ y $B_{\epsilon'}(x_0) \subset \mathrm{co}(S)$, con $\epsilon' = \frac{\epsilon}{\sqrt{n}}$. Si $x \in B_{\epsilon'}(x_0)$, $f(x_0) \geqq \frac{f(x)+f(2x_0-x)}{2} \geqq \frac{f(x)+m}{2}$, por lo que $f(x) \leqq 2f(x_0) - m$. Obtenemos que $|f(x)| \leqq M$ para todo $x \in B_{\epsilon'}(x_0)$, con $M = \max\{|m|, |2f(x_0) - m|\}$. Ahora si $x, y \in B_{\epsilon'/4}(x_0)$ con $x \neq y$, $z = y + \frac{\epsilon'}{2\|x-y\|}(x - y)$ está en $B_{\epsilon'}(x_0)$, y $f(x) = f(\frac{2\|x-y\|}{\epsilon'} z + (1 - \frac{2\|x-y\|}{\epsilon'})y) \geqq \frac{2\|x-y\|}{\epsilon'} f(z) + (1 - \frac{2\|x-y\|}{\epsilon'})f(y) = f(y) + \frac{2\|x-y\|}{\epsilon'}(f(z) - f(y)) \geqq f(y) - \frac{4M}{\epsilon'}\|x - y\|$. Obtenemos $|f(x) - f(y)| \leqq \frac{4M}{\epsilon'}\|x - y\|$ para todos $x, y \in B_{\epsilon'/4}(x_0)$, como queríamos. ■

Theorem 1.2.3. Sea $A \subset \mathbb{R}^n$ convexo y funciones $f_n, f : A \to \mathbb{R}$ para $n \in \mathbb{N}$, con f_n cóncavas. Si $f_n \to f$ puntualmente, entonces f es cóncava y, si $K \subset A^\circ$ es compacto, $f_n \to f$ uniformemente en K.

Proof. Sea $\epsilon > 0$. Sea $x_0 \in K$. Por la demostración del teorema anterior, tenemos que, para todo $x \in B_{\delta/4}(x_0)$, $|f_n(x) - f_n(x_0)| \leqq \frac{4M_n}{\delta}\|x - x_0\|$, donde $M_n = 2|f_n(x_0)| + \max_{x\in S}\{|f_n(x)|\}$ y $\delta > 0$ cumple que $B_{2\sqrt{n}\delta}(x_0) \subset A$. Si $n \geqq n_0$, tenemos $M_n \leqq M = 1 + 2|f(x_0)| + \max_{x\in S}\{|f(x)|\}$. Luego $|f_n(x) - f_n(x_0)| \leqq \frac{4M}{\delta}\|x - x_0\|$. En particular, $|f_n(x) - f(x)| \leqq |f_n(x) - f_n(x_0)| + |f_n(x_0) - f(x_0)| + |f(x_0) - f(x)| \leqq \frac{8M}{\delta}\|x - x_0\| + \frac{\epsilon}{2} < \epsilon$ si $n \geqq n_1(x_0)$ y $x \in B_{\delta'}(x_0)$, con $\delta' = \min\{\frac{\sqrt{n}\delta}{2}, \frac{\epsilon\delta}{16M}\}$; llamemos $\mathcal{U}(x_0)$ a $B_{\delta'}(x_0)$. Entonces cubrimos K con abiertos $\mathcal{U}(x_0)$, y obtenemos un subcubrimiento finito $\mathcal{U}(x_1), \ldots, \mathcal{U}(x_k)$. Tenemos, pues, que $|f_n(x) - f(x)| < \epsilon$ para todo $x \in K$ si $n \geqq n_2 = \max\{n_1(x_1), \ldots, n_1(x_k)\}$, como queríamos. ■

1.3 Función inversa y función implícita

Theorem 1.3.1 (Función inversa)**.** Sea $f : A \subset \mathbb{R}^n \to \mathbb{R}^n$ una función diferenciable, con A un entorno de a. Supongamos que $|Df(a)| \neq 0$ y que f es continuamente diferenciable en a. Entonces hay abiertos $V, W \subset \mathbb{R}^n$ tales que $a \in V$, $f(a) \in W$, $f : V \to W$ es biyectiva y $f^{-1} : W \to V$ es diferenciable. Además, si $f \in C^k(A)$, $f^{-1} \in C^k(W)$.

Proof. Pedimos una bola $V = B_\delta(a)$ tal que $|Df(x)| \neq 0$ y $\|f'(x) - f'(a)\| \leqq \frac{1}{2\|f'(a)^{-1}\|}$ para todo $x \in V$. Tenemos $\|x-y\| - \|f'(a)^{-1}(f(x)-f(y))\| \leqq \|(x - f'(a)^{-1}f(x)) - (y - f'(a)^{-1}f(y))\| = \|(1 - f'(a)^{-1}f'(z))(x-y)\| = \|f'(a)^{-1}(f'(a) - f'(z))(x-y)\| \leqq \|f(a)^{-1}\|\|f'(a) - f'(z)\|\|x-y\| \leqq \frac{1}{2}\|x-y\|$, donde $z \in [x, y]$ (valor medio en $g(x) = x - f'(a)^{-1}f(x)$); entonces $\|f(a)^{-1}\|\|f(x) - f(y)\| \geqq \|f'(a)^{-1}(f(x)-f(y))\| \geqq \|x-y\| - \frac{1}{2}\|x-y\| = \frac{1}{2}\|x-y\|$ y $\|f(x) - f(y)\| \geqq M\|x-y\|$, donde $M = \frac{1}{2\|f'(a)^{-1}\|}$. En particular, $f|_V$ es inyectiva.

Veamos que $f|_V$ manda abiertos a abiertos. Sea $U \subset V$ un abierto, y $x_0 \in U$; hay que ver que $f(U)$ es un entorno de $f(x_0)$. Hay $d > 0$ tal que $\overline{B_d(x_0)} \subset U$. La función $\|f(\cdot) - f(x_0)\|$ tiene un mínimo en $\partial B_d(x_0)$ que es $m > 0$. Veamos que $B_{m/2}(f(x_0)) \subset f(B_d(a))$. Si $y \in B_{m/2}(f(x_0))$, la función $g(x) = \|f(x) - y\|^2$ alcanza un mínimo en $x_* \in \overline{B_d(x_0)}$, y $x_* \notin \partial B_d(x_0)$ ya que en ese caso $\|f(x_*) - y\| \geqq \frac{m}{2}$, pero $\|f(x_0) - y\| < \frac{m}{2}$; luego $x \in B_d(x_0)$. Entonces $g'(x_*) = 0$, $f'(x_*)(f(x_*) - y) = 0$ y $f(x_*) = y$, como queríamos, ya que $f'(x_*)$ se suponía inversible. En particular obtenemos que $W = f(V)$ es abierto y $f : V \to W$ es biyectiva y con inversa continua.

Tenemos que $f(x) - f(y) - f'(y)(x-y) = E(x,y)$ con $\lim\limits_{x \to y} \frac{E(x,y)}{\|x-y\|} = 0$. Ahora $f^{-1}(x) - f^{-1}(y) - f'(f^{-1}(y))^{-1}(x-y) = -f'(f^{-1}(y))^{-1}(E(f^{-1}(x), f^{-1}(y)))$. Para ver que f^{-1} es diferenciable, hay que ver que $\lim\limits_{x \to y} \frac{E(f^{-1}(x), f^{-1}(y))}{\|x-y\|} = 0$. En efecto,

$$\lim_{x \to y} \frac{E(f^{-1}(x), f^{-1}(y))}{\|x-y\|} = \lim_{x \to y} \frac{E(f^{-1}(x), f^{-1}(y))}{\|f^{-1}(x) - f^{-1}(y)\|} \frac{\|f^{-1}(x) - f^{-1}(y)\|}{\|x-y\|} = 0,$$

porque el primer factor tiende a cero ya que f^{-1} es continua y el segundo está acotado por $\frac{1}{M}$. El tema de C^k viene de que $Df^{-1} = (Df)^{-1}$ por regla de la cadena y esto preserva continuidad de derivadas ya que es hacer un cociente de polinomios donde el denominador no se anula. ■

Theorem 1.3.2 (Función implícita)**.** Sean $A \subset \mathbb{R}^n$ y $B \subset \mathbb{R}^m$ abiertos, $f : A \times B \to \mathbb{R}^m$ diferenciable, y $(a, b) \in A \times B$ tal que f es continuamente diferenciable en (a, b) y $f(a, b) = 0$. Sea M la matriz $(D_{n+i}f_j(a,b))_{1 \leqq i,j \leqq m}$. Si $|M| \neq 0$, hay $g : A' \to B'$ donde $A' \subset A$ y $B \subset B'$ son abiertos, $(a, b) \in A' \times B'$ y $f(x, g(x)) = 0$ si $x \in A$. Además, g es diferenciable.

Proof. Sea $F : A \times B \to \mathbb{R}^n \times \mathbb{R}^m$ con $F(x, y) = (x, f(x, y))$. Entonces $|DF(a, b)| = |M| \neq 0$. Por el teorema de la función inversa, hay $V = A' \times B' \subset A \times B$ y $W \subset \mathbb{R}^n \times \mathbb{R}^m$ abiertos tales que $F : V \to W$ tiene una inversa diferenciable $h : W \to V$. Claramente $h(x, y) = (x, k(x, y))$ con k diferenciable. Ponemos $g(x) = k(x, 0)$ y $f(x, g(x)) = f(x, k(x, 0)) = f(h(x, 0)) = 0$, como queríamos. ■

1.4 Correspondencias y el teorema de Berge

Una *correspondencia* $F : X \rightrightarrows Y$ es un conjunto $F \subset X \times Y$ tal que para todo $x \in X$ existe $y \in Y$ (no necesariamente único) con $(x, y) \in F$. Se puede pensar como una función $F : X \to \mathcal{P}(Y) \smallsetminus \{\varnothing\}$.

Definition 1.4.1. Una correspondencia $F : X \rightrightarrows Y$ entre espacios topológicos X, Y se dice *upper hemicontinuous* en $x \in X$ si para todo abierto $V \subset Y$ con $F(x) \subset V$ existe un abierto

$U \subset X$ con $x \in U$ y $F(U) \subset V$; se dice *lower hemicontinuous* en $x \in X$ si para todo abierto $V \subset Y$ tal que $F(x) \cap V \neq \varnothing$ existe un abierto $U \subset X$ con $x \in U$ tal que para todo $x' \in U$, $F(x') \cap V \neq \varnothing$; se dice *continua* en $x \in X$ si es upper- y lower-hemicontinuous en x.

Proposition 1.4.2. Sean X, Y espacios métricos, $K \subset X$ compacto, $F : X \rightrightarrows Y$ compact-valued (es decir, $F(x)$ es compacto para todo $x \in X$) y upper-hemicontinuous en K, entonces $F(K)$ es compacto.

Proof. Sea $F(K) \subset \bigcup_{i \in I} V_i$ un cubrimiento abierto. Sea $x \in K$; como $F(x)$ es compacto, hay $i_1, \ldots, i_{n_x} \in I$ con $F(x) \subset \bigcup_{j=1}^{n_x} V_{i_j} = V_x$; como F es upper-hemicontinuous en x, hay $U_x \ni x$ abierto con $F(U_x) \subset V_x$. Como K es compacto, hay $x_1, \ldots, x_n \in K$ con $K \subset \bigcup_{i=1}^{n} U_{x_i}$; luego $F(K) \subset \bigcup_{i=1}^{n} F(U_{x_i}) \subset \bigcup_{i=1}^{n} \bigcup_{j=1}^{n_{x_i}} V_{i_j}$ y encontramos un subcubrimiento finito. ∎

Proposition 1.4.3. Sean X, Y espacios métricos, $F : X \rightrightarrows Y$ y $x \in X$.

(a) Supongamos que para toda secuencia $x_n \to x$ en X y toda secuencia $y_n \in F(x_n)$ existe una subsecuencia convergente $y_{n_k} \to y$ con $y \in F(x)$. Entonces F es upper-hemicontinuous en x. Vale el recíproco si F es compact-valued y upper-hemicontinuous en un entorno de x. También vale el recíproco si $F(x)$ es compacto y Y es localmente compacto.

(b) F es lower-hemicontinuous en x si y sólo si para todo $y \in F(x)$ y toda secuencia $x_n \to x$ existe una secuencia $y_n \to y$ con $y_n \in F(x_n)$.

Proof. (a) Primera implicación. Supongamos que F no es upper-hemicontinuous en x. Entonces existe un abierto $V \supset F(x)$ tal que para todo abierto $U \ni x$ vale $F(U) \setminus V \neq \varnothing$; tomamos $U_n = B_{\frac{1}{n}}(x)$ y obtenemos $x_n \in U_n$, $x_n \to x$, $y_n \in F(x_n)$ con $y_n \notin V$; tomamos una subsecuencia convergente $y_{n_k} \to y$ con $y \in F(x) \subset V$ por hipótesis y tenemos $y_n \in V$ para cierto n, absurdo.

Segunda implicación. Sea $x_n \to x$ y $y_n \in F(x_n)$. El conjunto $S = \{x_n \mid n \in \mathbb{N}\} \cup \{x\}$ es compacto, luego por la hipótesis y la proposición anterior $K = F(S)$ es compacto. Como $y_n \in K$, hay una subsecuencia convergente $y_{n_k} \to y \in K$. Si $y \notin F(x)$, hay abiertos disjuntos $V \supset F(x), W \ni y$ que los separan. Ahora sea $U \ni x$ abierto tal que $F(U) \subset V$, por hipótesis, y n_0 tal que $x_n \in U$ si $n \geqq n_0$; tenemos $y_{n_k} \in V$ si $n_k \geqq n_0$, pero dado que $y \in W$, tenemos que $y_{n_k} \in W$ si k es grande, lo cual es absurdo.

Si $F(x)$ es compacto y Y es localmente compacto, para cada $y \in F(x)$ existe $V_y \ni y$ abierto y $K_y \supset V_y$ compacto; luego hay finitos, $y_1, \ldots, y_n$ con $F(x) \subset V = \bigcup_{i=1}^{n} V_{y_i} \subset \bigcup_{i=1}^{n} K_{y_i} = K$, con K compacto. Luego por upper-hemicontinuity en x, hay $U \ni x$ abierto con $F(U) \subset V$. Si $x_n \to x$ y $y_n \in F(x_n)$, para n grande vale $x_n \in U$, $y_n \in F(x_n) \subset F(U) \subset V \subset K$, luego hay una subsecuencia convergente $y_{n_k} \to y$. Si $y \notin F(x)$, hay abiertos que los separan, y por upper-hemicontinuity obtenemos un absurdo.

(b) Primera implicación. Sea $V \subset Y$ abierto con $F(x) \cap V \neq \varnothing$ de manera que para todo $U \ni x$ abierto hay $x' \in U$ con $F(x') \cap V = \varnothing$. Sea $y \in F(x) \cap V$, $U_n = B_{\frac{1}{n}}(x)$ y $x_n \in U_n$ con $F(x_n) \cap V = \varnothing$. Tenemos $x_n \to x$; luego, por hipótesis, hay $y_n \to y$ con $y_n \in F(x_n)$; como $y \in V$, para $n \geqq n_0$ tenemos $y_n \in V$, y $y_n \in F(x_n) \cap V$, absurdo.

Segunda implicación. Sea $y \in F(x)$ y $x_n \to x$. Dado k sea $V_k = B_{\frac{1}{k}}(y)$; como $y \in F(x) \cap V_k$, hay $U_k \ni x$ abierto tal que si $x' \in U_k$, hay $y' \in F(x') \cap V_k$. Esto implica que podemos construir una secuencia y_n inductivamente de manera que $x_n \in U_k$ y $y_n \in F(x_n) \cap V_k$ si $n \geqq n_k$. Entonces $y_n \to y$ y $y_n \in F(x_n)$. ∎

Definition 1.4.4. Dados X, Y métricos, decimos que $F : X \rightrightarrows Y$ tiene *gráfico cerrado* en $x \in X$ si para toda sucesión convergente $(x_n, y_n) \to (x, y)$ con $y_n \in F(x_n)$ vale $y \in F(x)$.

Proposition 1.4.5. Si X, Y métricos, $x \in X$, $F : X \rightrightarrows Y$ es upper-hemicontinuous en x y $F(x)$ es cerrado, entonces F tiene gráfico cerrado en x.

Proof. Sean $x_n \to x$, $y_n \to y$ con $y_n \in F(x_n)$. Si $y \notin F(x)$ hay abiertos disjuntos $W \ni y$, $V \supset F(x)$. Por upper-hemicontinuity, hay $U \ni x$ abierto con $F(U) \subset V$. Si n es grande, $x_n \in U$ y $y_n \in F(x_n) \subset F(U) \subset V$, pero además $y_n \in W$, luego $y_n \in V \cap W$, absurdo. ■

Proposition 1.4.6. Si X, Y métricos, $F : X \rightrightarrows Y$ tiene gráfico cerrado en $x \in X$ y existe un entorno U de x tal que $F(U)$ está contenido en un compacto entonces F es upper-hemicontinuous en x.

Proof. Si no, existen $V \supset F(x)$ abierto, $x_n \to x$, $y_n \in F(x_n)$ con $y_n \notin V$. Ahora si n es grande, $x_n \in U$ y $y_n \in F(U)$, luego hay $y_{n_k} \to y$. Ahora por gráfico cerrado en x, como $(x_{n_k}, y_{n_k}) \to (x, y)$, resulta $y \in F(x)$ y $y \in V$, por lo que para k grande $y_{n_k} \in V$, absurdo. ■

Theorem 1.4.7 (Berge)**.** Sean X, Y espacios métricos, $f : X \times Y \to \mathbb{R}$ continua y $G : X \rightrightarrows Y$ compact-valued y continua en un entorno de $x_0 \in X$ (o continua en x_0 y con Y localmente compacto). Entonces $M(x) = \max_{y \in G(x)} f(x, y)$ es continua en x_0, y $\Pi(x) = \arg\max_{y \in G(x)} f(x, y)$ es una correspondencia compact-valued, upper-hemicontinuous y tiene gráfico cerrado en x_0.

Proof. Primero, veamos que M es continua en x_0. Hay que probar que para todo $\epsilon > 0$ existe $U \ni x_0$ abierto tal que $M(x_0) - \epsilon < M(x) < M(x_0) + \epsilon$ para todo $x \in U$. Sea $y_0 \in G(x_0)$ con $f(x_0, y_0) = M(x_0)$. Busco en primer lugar U que cumpla la primera desigualdad. Por continuidad de f, el conjunto $W = \{f > M(x_0) - \epsilon\}$ es abierto y $(x_0, y_0) \in W$, por lo que hay abiertos $U_1 \ni x_0$, $V \ni y_0$ con $U_1 \times V \subset W$. Por hipótesis (G es lower-hemicontinuous en x_0), y dado que $G(x_0) \cap V \neq \varnothing$, existe $U_1 \ni x_0$ abierto tal que para todo $x \in U_1$, $G(x) \cap V \neq \varnothing$. Entonces si $x \in U = U_1 \cap U_2$, existe $y \in G(x) \cap V$, luego $(x, y) \in W$, $f(x, y) > M(x_0) - \epsilon$, y dado que $M(x) \geqq f(x, y)$ resulta $M(x) > M(x_0) - \epsilon$, como queríamos.

Busco ahora $U \ni x_0$ abierto tal que $M(x) < M(x_0) + \epsilon$ para todo $x \in U$. Si no existe, hay $x_n \to x$ con $M(x_n) \geqq M(x_0) + \epsilon$, es decir, hay $y_n \in G(x_n)$ con $f(x_n, y_n) \geqq f(x_0, y_0) + \epsilon$. Por la proposición anterior, hay $y_{n_k} \to y \in G(x_0)$. Por continuidad de f, $f(x_{n_k}, y_{n_k}) \to f(x_0, y)$; luego $f(x_0, y) \geqq f(x_0, y_0) + \epsilon$, absurdo.

Segundo, veamos que Π es upper-hemicontinuous en x_0 (que es compact-valued es obvio). Sean $x_n \to x_0$ y $y_n \in \Pi(x_n)$. Como $\Pi(x_n) \subset G(x_n)$ y G es compact-valued y upper-hemicontinuous en un entorno de x_0, hay $y_{n_k} \to y \in G(x_0)$. Ahora por continuidad de f tenemos $f(x_{n_k}, y_{n_k}) \to f(x_0, y)$, es decir, $M(x_{n_k}) \to f(x_0, y)$, pero por continuidad de M tenemos $M(x_{n_k}) \to M(x_0)$, luego $f(x_0, y) = M(x_0)$ y $y \in \Pi(x_0)$. Esto prueba que Π es upper-hemicontinuous en x_0 por una proposición anterior. Por otra, resulta que Π tiene gráfico cerrado en x_0. ■

1.5 Fixed points

Let $\Delta^n = \{\lambda \in [0,1]^{n+1} : \sum_{i=1}^{n+1} \lambda_i = 1\}$ be the n-simplex. A simplicial subdivision of Δ^n is a finite partition into small simplices of dimension n, called cells, such that any two cells share a full face of a certain dimension.

Lemma 1.5.1 (Sperner)**.** Let V be the set of vertices of a simplicial subdivision of Δ^n and $c : V \to \{1, \ldots, n+1\}$ a coloring such that if $v \in V$, $v_i = 0$ then $c(v) \neq i$. Then there is a cell such that each vertex has a different color.

Proof. We prove that the number of such cells is odd. For $n = 0$ there is nothing to prove. For $n \geqq 1$ we proceed by induction. Let R denote the number of cells with $n+1$ colors, and Q the number of cells with colors $\{1, \ldots, n\}$. We call n-faces the $n-1$-dimensional faces of cells with colors $\{1, \ldots, n\}$. Let X be the number of n-faces faces on the boundary of Δ^n and Y the number of those faces on the inside. Each cell of type R contributes one n-face and each cell of type Q contributes two. Inside n-faces appear in two cells, and boundary n-faces appear in only one. Hence $R + 2Q = X + 2Y$. Now, n-faces on the boundary are in Δ^{n-1}, which is properly colored. Hence by induction X is odd, and R is odd as well. ∎

Theorem 1.5.2 (Brouwer's Fixed Point)**.** Let $X \subset \mathbb{R}^d$ be convex, compact and nonempty, and $f : X \to X$ continuous. Then f has a fixed point.

Proof. There is a homeomorphism with Δ^n, so assume $X = \Delta^n$. Suppose that there is no fixed point. Take simplicial subdivisions with diameters tending to zero. Define the coloring $c : X \to \{1, \ldots, n+1\}$ by $c(x) = \min\{i : f(x)_i < x_i\}$. We can apply Sperner, and we get a cell S_k with vertices of all colors. Take a convergent subsequence of the centroids $x_k \to x^*$. The vertices converge to x^* as well, and taking limits we get $f(x^*) \leqq x^*$, so $f(x^*) = x^*$, absurd. ∎

Theorem 1.5.3 (Kakutani)**.** Let $X \subset \mathbb{R}^d$ be convex, compact and nonempty, and $f : X \rightrightarrows X$ convex-valued with closed graph. Then f has a fixed, point, i.e., there is $x \in X$ with $x \in f(x)$.

Proof. There is a homeomorphism with Δ^k, so assume $X = \Delta^k$. Take simplicial subdivisions S_n with diameters tending to zero. Define $f_n : X \to X$ by taking $f_n(v) \in f(v)$ arbitrary for any vertex v and extrapolating linearly in every cell: $f_n(\sum_{i=1}^{k+1} \lambda_i v_i) = \sum_{i=1}^{k+1} \lambda_i f_n(v_i)$. The maps f_n are continuous and have fixed points x_n by Brouwer. Pass to a convergence subsequence $x_n \to x^*$. Let $v_{n1}, \ldots, v_{n,k+1}$ be the vertices of the cell of x_n, and pass to a subsequence so that $y_{ni} = f_n(v_{ni})$ converges to y_i^*. We have $(v_{ni}, y_{ni}) \to (x^*, y_i^*)$, so $y_i^* \in f(x^*)$ since f has closed graph. We have $x_n = f_n(x_n) = \sum_{i=1}^{k+1} \lambda_{ni} y_{ni}$, and passing to a subsequence we have $\lambda_n \to \lambda^*$, hence $x^* = \sum_{i=1}^{k+1} \lambda_i^* y_i^* \in f(x^*)$, since it is assumed convex. ∎

We define the *convex hull* of a set $S \subset \mathbb{R}^d$, denoted $\mathrm{co}(S)$, as the set of convex combinations of points of S, i.e., sums $\sum_{i=1}^n \lambda_i x_i$ with $n \in \mathbb{N}$, $\lambda \in [0,1]^N$, $\sum_{i=1}^n \lambda_i = 1$, $x_i \in S$.

Theorem 1.5.4 (Carathéodory)**.** Let $S \subset \mathbb{R}^n$ and $x \in \mathrm{co}(S)$; then x is a convex combination of $n+1$ points from S.

Proof. Let $T = \{x_1, \ldots, x_m\} \subset S$ be a set s.t. $x \in \mathrm{co}(T)$ with m minimal, and let $\lambda_i \in (0,1)$ with sum 1 s.t. $x = \sum_{i=1}^m \lambda_i x_i$. Suppose that $m > n+1$. Then there is $a \in \mathbb{R}^m$, $a \neq 0$, s.t. $\sum_{i=1}^m a_i x_i = 0$, $\sum_{i=1}^m a_i = 0$. Let $t = \min\{-\frac{\lambda_i}{a_i} \mid i = 1, \ldots, m \text{ and } a_i < 0\}$. We have $t > 0$, $\mu_i = \lambda_i + t a_i \geqq 0$ for all i and there is j s.t. $\mu_j = \lambda_j + t a_j = 0$. Now $\mu_i \in [0,1]$, $\sum_{i=1}^m \mu_i = 1$ and $x = \sum_{i=1}^m \mu_i x_i$, but $\mu_j = 0$, hence $x \in \mathrm{co}(T \setminus \{x_j\})$, absurd. ∎

Theorem 1.5.5 (Helly)**.** Let $X_1, \ldots, X_n \subset \mathbb{R}^d$ convex, $n \geqq d+2$, such that $\bigcap_{i \in L} X_i \neq \varnothing$ for all $L \subset N \equiv \{1, \ldots, n\}$ with $|L| = d+1$. Then $\bigcap_{i \in N} X_i \neq \varnothing$.

Proof. First, assume $n = d+2$. Let $x_i \in \bigcap_{j \neq i} X_j$. There is $a \in \mathbb{R}^n$ s.t. $a \neq 0$, $\sum_{i=1}^n a_i x_i = 0$, $\sum_{i=1}^n a_i = 0$. This implies $\mathrm{co}(\{x_i : a_i > 0\}) \cap \mathrm{co}(\{x_i : a_i \leqq 0\}) \neq \varnothing$, and clearly a point in that intersection is in $\bigcap_{i \in N} X_i \neq \varnothing$. For larger n we proceed by induction, using the result on $X_1' = X_1, \ldots, X_{n-1}' = X_{n-1} \cap X_n$. ∎

Theorem 1.5.6 (Knaster-Kuratowski-Mazurkiewicz Lemma)**.** Let $a_1, \ldots, a_{n+1} \in \mathbb{R}^d$, $S_1, \ldots, S_{n+1} \subset \mathbb{R}^d$ closed such that for every $L \subset N \equiv \{1, \ldots, n+1\}$ we have $\mathrm{co}(\{a_i : i \in L\}) \subset \bigcup_{i \in L} S_i$. Then $\bigcap_{i \in N} S_i \neq \varnothing$.

Proof. Define the coloring $c : \Delta^n \to N$ by $c(\lambda) = \min\{i : \lambda_i \neq 0, \sum_{j=1}^{n+1} \lambda_j a_j \in S_i\}$. Take simplicial subdivisions with diameters tending to zero, apply Sperner, obtain a sequence of cells with vertices of all colors, and take a convergent subsequence of the centroids $x_k \to x^*$. The vertices form sequences $x_k^i \to x^*$ with $x_k^i \in S_i$, so $x^* \in \bigcap_{i \in N} S_i$, as desired. ∎

Application (existence of Nash equilibria) There are n players. Each player i choosea a strategy from the set $X_i \neq \varnothing$. She obtains a payoff $u_i(x)$, where $x \in X = \prod_{i=1}^n X_i$ is the profile of strategies chosen by all the players, and $u_i : X \to \mathbb{R}$ is a utility function. We say that $x^* \in X$ is a *Nash equilibrium* of the game if, for all i, $x_i^* \in \arg\max_{x_i \in X_i} u_i(x_i, x_{-i}^*)$, where x_{-i}^* is x^* omitting x_i^*.

Theorem 1.5.7. If, for every i, X_i is convex and compact, u_i is continuous and $u_i(\cdot, x_{-i})$ is quasi-concave [1] for every $x_{-i} \in X_{-i}$, then a Nash equilibrium exists.

Proof. Given $x \in X$ let $F(x) = \prod_{i=1}^n BR_i(x_{-i})$, where $BR_i(x_{-i}) = \arg\max_{x \in X_i} u_i(x, x_{-i})$. Note that x is a Nash equilibrium if and only if $x \in F(x)$. Now $F : X \rightrightarrows X$; since X_i is convex and $u_i(\cdot, x_{-i})$ is quasi-concave, F is convex-valued. By Berge's theorem, F has closed graph. The result then follows from Kakutani. ∎

Suppose that the sets X_i are finite. A *mixed strategy* for player i is a probability distribution over X_i, i.e., a function $p_i \in \Delta X_i = \{p : X_i \to [0,1] \mid \sum_{x_i \in X_i} p(x_i) = 1\}$. Given a profile of mixed strategies $p \in \prod_{i=1}^n \mathcal{P}(X_i)$, there are payoffs $u_i(p) = \mathbb{E}(u_i(x)) = \sum_{x \in X} (\prod_{j=1}^n p_j(x_j)) u_i(x)$. Note that we can apply the previous theorem, and we obtain that every finite game has a Nash equilibrium in mixed strategies.

1.6 Teorema de Hahn-Banach

Theorem 1.6.1 (Hahn-Banach)**.** Sea X un $\mathbb{R}$-espacio vectorial y $m : X \to \mathbb{R}$ tal que $m(ax) = am(x)$ si $a \in \mathbb{R}_{\geqq 0}$ y $m(x+y) \leqq m(x) + m(y)$. Si $Y \subset X$ es un subespacio, $\varphi : Y \to \mathbb{R}$ es lineal y $\varphi(x) \leqq m(x)$, hay una extensión $\tilde{\varphi} : X \to \mathbb{R}$ que también cumple $\tilde{\varphi}(x) \leqq m(x)$.

Proof. Por Zorn tomemos φ maximal, y $u \in X \setminus Y$. Buscamos $k \in \mathbb{R}$ de manera que la extensión $\tilde{\varphi}$ con $\tilde{\varphi}(u) = k$ cumpla $\tilde{\varphi} \leqq m$. Queremos $\tilde{\varphi}(x + au) \leqq m(x + au)$ para todo $x \in Y$. Se ve que k debe cumplir $-m(x-u) + \varphi(x) \leqq k \leqq m(y+u) - \varphi(y)$ para todos $x, y \in Y$. Ahora $-m(x-u) + \varphi(x) \leqq m(y+u) - \varphi(y)$, ya que es $\varphi(x+y) \leqq m(x-u) + m(y+u)$, que vale, luego $k = \sup_{x \in Y} \{-m(x-u) + \varphi(x)\}$ funciona. ∎

Recordemos que si X es normado X^*, el *espacio dual*, es el conjunto de las funciones lineales $f : X \to \mathbb{R}$ continuas, con la norma $\|f\| = \sup_{x \in X \setminus 0} \frac{|f(x)|}{\|x\|}$.

Theorem 1.6.2 (Hahn-Banach geométrico 1)**.** Sea X normado y $U, V \subset X$ convexos no vacíos disjuntos, con U abierto. Entonces hay $\varphi \in X^*$ y $t \in \mathbb{R}$ con $\varphi(x) < t \leqq \varphi(y)$ para todos $x \in U$, $y \in V$.

Proof. Sean $x_0 \in U$, $y_0 \in V$, $z_0 = y_0 - x_0$ y $C = U - V + z_0$; C es abierto convexo y $0 \in C$. Sea $m(x) = \inf\{s > 0 \mid x \in sC\}$. Como $z_0 \notin C$, $m(x_0) \geqq 1$. Definimos φ_0 en $\langle z_0 \rangle$ por $\varphi_0(z_0) = 1$ y tenemos $\varphi_0 \leqq m$. Por Hahn-Banach extendemos φ_0 a $\varphi : X \to \mathbb{R}$ lineal con $\varphi \leqq m$; se ve que $\varphi \in X^*$. Si $x \in U$, $y \in V$, $x - y + z_0 \in C$ y $\phi(x - y + z_0) < 1$, por lo que $\phi(x) < \phi(y)$. Sea $t = \sup_{x \in U} \phi(x)$; vale $\phi(x) \leqq t \leqq \phi(y)$ si $x \in U, y \in V$. Ahora $(1 + \frac{1}{n})x \in U$ si n es grande, luego $\phi(x) < t$. ∎

Theorem 1.6.3 (Hahn-Banach geométrico 2)**.** Sea X normado y $U, V \subset X$ convexos con $U^\circ \neq \varnothing$ y $U^\circ \cap V = \varnothing$. Entonces hay $\varphi \in X^*$ y $t \in \mathbb{R}$ con $\varphi(x) \leqq t \leqq \varphi(y)$ para todos $x \in U$, $y \in V$.

[1] If X is convex, $f : X \to \mathbb{R}$ is called *quasi-concave* if for all $x, y \in X$, $\lambda \in [0,1]$, we have $f(\lambda x + (1-\lambda) y) \geqq \min\{f(x), f(y)\}$.

Proof. Por HBG1 tenemos $\varphi \in X^*$ y $t \in \mathbb{R}$ con $\varphi(x) < t \leqq \varphi(y)$ para todos $x \in U^\circ$, $y \in V$. Sea $x \in U$. Como $U^\circ \neq \varnothing$, hay $B_\epsilon(x_0) \subset U$. Ahora $B_{\lambda\epsilon}(\lambda x_0 + (1-\lambda)x) = \lambda B_\epsilon(x_0) + (1-\lambda)x \subset U$, luego $\lambda x_0 + (1-\lambda)x \in U^\circ$ si $\lambda > 0$, así que poniendo $\lambda = \frac{1}{n}$ tenemos $x_n \to x$ con $x_n \in U^\circ$; tenemos $\varphi(x_n) < t$, luego $\varphi(x) \leqq t$, como queríamos. ∎

Theorem 1.6.4 (Hahn-Banach geométrico 3)**.** Sea X normado, K convexo compacto y V convexo cerrado, con $K \cap V = \varnothing$. Entonces hay $\varphi \in X^*$ y $t \in \mathbb{R}$ con $\varphi(x) < t < \varphi(y)$ para todos $x \in K$, $y \in V$.

Proof. $S = K - V$ es cerrado y $0 \notin S$, por lo que hay $\epsilon > 0$ con $B_\epsilon(0) \cap S = \varnothing$. Ahora $\tilde{K} = K + \frac{1}{2}U$ y $\tilde{V} = V + \frac{1}{2}U$ son abiertos convexos con $\tilde{K} \cap \tilde{V} = \varnothing$. Hay $\phi \in X^*$ que separa $\tilde{K}$ y $\tilde{V}$, luego a K y V. ∎

1.7 Optimization

Let $X \subset \mathbb{R}^k$ be convex, $f : X \to \mathbb{R}$, $g : X \to \mathbb{R}^n$. We want to

$$\begin{aligned} \text{maximize} \quad & f(x) \\ \text{subject to} \quad & g(x) \leqq 0, \quad x \in X. \end{aligned}$$

Given $\lambda \in \mathbb{R}^n$, let $\mathcal{L}(x, \lambda) = f(x) - \lambda^\top g(x)$. We say that (x, λ) such that $x \in X \cap \{g \leqq 0\}$, $\lambda \in \mathbb{R}^n_{\geqq 0}$ is a *saddle point* if x maximizes $\mathcal{L}(\cdot, \lambda)$ over X and λ minimizes $\mathcal{L}(x, \cdot)$ over $\mathbb{R}^n_{\geqq 0}$.

Theorem 1.7.1. Let f be concave and g convex. Then x is a solution if and only if there exists $\lambda \in \mathbb{R}^n_{\geqq 0}$ such that (x, λ) is a saddle point of $\mathcal{L}$. In that case, $\lambda_i g_i(x) = 0$ for all $i = 1, \dots, n$.

Proof. Let x^* be a solution, $U = \{(a, b) \mid a > f(x^*), b < 0\}$ and $V = \bigcup_{x \in X} \{(a, b) \mid a \leqq f(x), b \geqq g(x)\}$. By convexity of X, concavity of f and convexity of g, V is convex. Since x^* is a solution, $U \cap V = \varnothing$. By the geometric Hahn-Banach theorem there is $\varphi \in \mathbb{R}^{n+1}$ and $t \in \mathbb{R}$ s.t. $\varphi^\top x < t \leqq \varphi^\top y$ for all $x \in U$, $y \in V$. Let $\rho \in \mathbb{R}$, $\lambda \in \mathbb{R}^n$ s.t. $(\rho, \lambda) = \varphi$. If $y \in \mathbb{R}^n_{\leqq 0}$, $(f(x^*), y) \in \overline{U}$ and $(f(x^*), 0) \in V$, then $\lambda^\top y \leqq 0$ and $\lambda \geqq 0$; since $(f(x^*) + 1, 0) \in \overline{U}$, $\rho \leqq 0$. If $\rho < 0$ we are done, since, letting $\lambda^* = -\frac{1}{\rho}\lambda$, it is clear that (x^*, λ^*) is a saddle point of $\mathcal{L}$ and $\lambda^* \cdot g(x^*) = 0$. If $\rho = 0$ we have $\lambda^\top g(x) \geqq 0$ for all $x \in X$. Since $\varphi \neq 0$, there is $\lambda_i > 0$, so for all $x \in X \cap \{g \leqq 0\}$ we have $g_i(x) = 0$. Remove iteratively those i. At the end we have $\tilde{g}$ s.t. $\tilde{g}(x) \leqq 0$ iff $g(x) \leqq 0$, but we must have $\rho < 0$ for $\tilde{g}$. It is easy to see that we can extend the $\tilde{\lambda}^*$ we obtain by adding zeroes, and the resulting (x^*, λ^*) is a saddle point of $\mathcal{L}$, as desired. The converse is obvious. ∎

Let $c \in \mathbb{R}^k$, $A \in \mathbb{R}^{n \times k}$. A *linear programming* problem is to

$$\begin{aligned} \text{maximize} \quad & c^\top x \\ \text{subject to} \quad & Ax \leqq b, \quad x \geqq 0. \end{aligned}$$

The *dual problem* is to minimize $b^\top y$ subject to $A^\top y \geqq c$, $y \geqq 0$.

Theorem 1.7.2 (Duality)**.** A linear programming problem has a solution x^* iff its dual has a solution y^*, and the objectives are equal: $c^\top x^* = b^\top y^*$.

Proof. We apply the previous Theorem and get y^* such that x^* maximizes $c^\top x - (Ax - b)^\top y^*$ s.t. $x \geqq 0$ and y^* minimizes $c^\top x^* - (Ax^* - b)^\top y$ s.t. $y \geqq 0$. This implies $A^\top y^* \geqq c$ and $c^\top x^* = b^\top y^*$. If y satisfies $A^\top y \geqq c$ and $y \geqq 0$, we have $c^\top x^* \leqq b^\top y$. Since y^* satisfies this with equality, it solves the dual problem, as desired. ∎

Sea $X \subset \mathbb{R}^k$ convexo, $f : X \to \mathbb{R}$, $g : X \to \mathbb{R}^n$, $h : X \to \mathbb{R}^m$ funciones C^1 en X°. Buscamos soluciones al siguiente problema:

$$\begin{aligned} \text{maximizar} \quad & f(x) \\ \text{sujeto a} \quad & g(x) \leqq 0, \quad h(x) = 0, \quad x \in X. \end{aligned}$$

Decimos que la condición $g_i(x) \leqq 0$ está *activa* si $g_i(x) = 0$.

Theorem 1.7.3 (Kuhn-Tucker)**.** Si $x \in X^\circ$ es una solución, $J = \{i \in \{1, \dots, n\} \mid g_i(x) = 0\}$, $n' = |J|$, $\bar{h} : X \to \mathbb{R}^{n'+m}$ está dado por $\bar{h}(x) = ((g_i(x))_{i\in J}, h(x))$, y $D\bar{h}(x)$ tiene rango $n'+m < k$, entonces existen $\lambda \in \mathbb{R}^n_{\geqq 0}$, $\mu \in \mathbb{R}^m$ tales que

$$Df(x) - \lambda^t Dg(x) - \mu^t Dh(x) = 0$$

y $\lambda_i g_i(x) = 0$ para todo $i = 1, \dots, n$.

Proof. Sea x^* la solución. Como $D\bar{h}(x)$ tiene rango $n' + m$, por función implícita hay $U \subset \mathbb{R}^{k-n'-m}$ y $V \subset \mathbb{R}^{n'+m}$ abiertos, $U \times V \subset X$, $x^* \in U \times V$, $\varphi : U \to V$ diferenciable con $\bar{h}(x, \varphi(x)) = 0$ para todo $x \in U$, y $x^* = (x_1^*, x_2^*)$, $x_1^* \in U$, $x_2^* \in V$, $x_2^* = \varphi(x_1^*)$. Sea $\bar{f} : U \to \mathbb{R}$ dada por $\bar{f}(x) = f(x, \varphi(x))$. En un entorno de x_1^* vale que $(x, \varphi(x))$ cumple las restricciones, luego $D\bar{f}(x_1^*) = 0$, y $D_1 f(x^*) + D_2 f(x^*) D\varphi(x_1^*) = 0$. Como $\bar{h}(x, \varphi(x)) = 0$ para todo $x \in U$, $D_1\bar{h}(x^*) + D_2\bar{h}(x^*) D\varphi(x_1^*) = 0$. Ahora por hipótesis $D_2\bar{h}(x^*)$ es inversible, luego

$$\begin{aligned} D_1 f(x^*) &= - D_2 f(x^*) D\varphi(x_1^*) = \\ &= - D_2 f(x^*) D_2\bar{h}(x^*)^{-1} D_2\bar{h}(x^*) D\varphi(x_1^*) = D_2 f(x^*) D_2\bar{h}(x^*)^{-1} D_1\bar{h}(x^*). \end{aligned}$$

Si $\tilde{\lambda} = D_2 f(x^*) D_2\bar{h}(x^*)^{-1}$, tenemos $Df(x^*) = \tilde{\lambda} D\bar{h}(x^*)$. Con esto, y rellenando λ con ceros, tenemos λ y μ como buscamos.

Sólo falta ver que $\lambda_i \geqq 0$ sobre las condiciones activas. Esto es, que $\tilde{\lambda} e_i \geqq 0$ para $i = 1, \dots, n'$. Ahora sea $\tilde{h} : V \to \mathbb{R}^{n'+m}$ dada por $\tilde{h}(x_2) = \bar{h}(x_1^*, x_2)$; vale $\tilde{h}(x_2^*) = 0$. Tenemos que $D\tilde{h}(x_2^*) = D_2\bar{h}(x^*)$ es inversible. Entonces localmente tiene una inversa $\tilde{h}^{-1}$ y tenemos $D\tilde{h}^{-1}(0) = D\tilde{h}(x_2^*)^{-1} = D_2\bar{h}(x^*)$. Sea $\tilde{f} : V \to \mathbb{R}$ dada por $\tilde{f}(x_2) = f(x_1^*, x_2)$. Tenemos $D\tilde{f}(x_2^*) = D_2 f(x^*)$. Ahora $\tilde{\lambda} = D_2 f(x^*) D_2\bar{h}(x^*)^{-1} = D\tilde{f}(x_2^*) D\tilde{h}^{-1}(0) = D(\tilde{f} \circ \tilde{h}^{-1})(0)$. Entonces queremos ver que para todo $i = 1, \dots, n'$ vale $\partial_i(\tilde{f} \circ \tilde{h}^{-1})(0) \geqq 0$. Ahora si $\partial_i(\tilde{f} \circ \tilde{h}^{-1})(0) < 0$, hay $\epsilon > 0$ con $\tilde{f}(\tilde{h}^{-1}(-\epsilon e_i)) > \tilde{f}(\tilde{h}^{-1}(0)) = f(x^*)$. Ahora $\tilde{f}(\tilde{h}^{-1}(-\epsilon e_i)) = f(x_1^*, x_2)$ tal que (x_1^*, x_2) cumple todas las restricciones: $g_i(x_1^*, x_2) = -\epsilon < 0$. Entonces x^* no es solución, absurdo. ∎

Ahora sean $X \subset \mathbb{R}^k$ e $I \subset \mathbb{R}$ convexos, $f : X \times I \to \mathbb{R}$, $g : X \times I \to \mathbb{R}^n$, $h : X \times I \to \mathbb{R}^m$ funciones C^1 en $X^\circ \times I$. Dado $p \in I$ buscamos soluciones al siguiente problema:

$$\begin{aligned} \text{maximizar} \quad & f(x, p) \\ \text{sujeto a} \quad & g(x, p) \leqq 0, \quad h(x, p) = 0, \quad x \in X. \end{aligned}$$

Se comprueba fácilmente lo siguiente.

Theorem 1.7.4 (Envelope theorem)**.** Supongamos que existe una función diferenciable $x^* : I \to X$ tal que $x^*(p)$ es solución del problema para todo $p \in I$ y cumple las condiciones del teorema anterior. Sea $v(p) = f(x^*(p), p)$. Tenemos

$$v'(p) = \partial_2 f(x^*(p), p) - \lambda^t D_2 g(x^*(p), p) - \mu^t D_2 h(x^*(p), p).$$

¿Bajo qué condiciones existe la función $x^*(\cdot)$ y es diferenciable? Supongamos que en $x^*(p)$ se cumplen las condiciones de Kuhn-Tucker y sea s el vector de condiciones activas y condiciones h. Tenemos que $Df - \lambda Ds = 0$ y $s = 0$ son $k + n' + m$ ecuaciones sobre $k + n' + m$ incógnitas. Sea $F(x, \lambda) = (Df - \lambda Ds, s)$; si $|DF| \neq 0$, tenemos $x^*(\cdot)$ diferenciable por función implícita, como queremos. Ahora $|DF| = |H(f - \lambda s)||-DsH(f - \lambda s)^{-1}Ds^t| \neq 0$ en x^* si asumimos $H(f-\lambda s)(x^*)$ definida negativa y que $Ds(x^*)$ tiene rango $n'+m$. Notar que $D(f-\lambda s)(x) = 0$, $s(x) = 0$ y $H(f-\lambda s)(x)$ definida positiva implican que x es un máximo local de f sobre $\{s = 0\}$, así que esto funciona.

1.8 Dynamic programming

Given a metric space X, $G : X \rightrightarrows X$, $f : G \to \overline{\mathbb{R}}$, $\beta \in (0, 1)$ y $x_0 \in X$, a *dynamic programming* problem is

$$\begin{aligned} \text{maximize} \quad & \sum_{t=0}^{\infty} \beta^t f(x_t, x_{t+1}) \\ \text{subject to} \quad & x_{t+1} \in G(x_t) \text{ for all } t \in \mathbb{N}_0. \end{aligned}$$

Define $\Omega(x_0) = \{x \in X^{\mathbb{N}} \mid (\forall t \in \mathbb{N}_0)\, x_{t+1} \in G(x_t)\}$ and $F(x) = \sum_{t=0}^{\infty} \beta^t f(x_t, x_{t+1})$ for $x \in \Omega(x_0)$. We assume that for all $x \in \Omega(x_0)$, $F(x) \in \overline{\mathbb{R}}$ holds. Let $V : X \to \overline{\mathbb{R}}$ be given by $V(x_0) = \sup_{x \in \Omega(x_0)} F(x)$.

Assumption 1.8.1. (A1) f is continuous and bounded. (A2) G is compact-valued and continuous.

Proposition 1.8.1 (Bellman)**.** Let $x \in \Omega(x_0)$. If x maximizes F then $V(x_t) = f(x_t, x_{t+1}) + \beta V(x_{t+1})$ for all $t \in \mathbb{N}_0$. Under assumption (A1) the reciprocal also holds.

Proof. The first assertion is obvious, since if $V(x_t) \neq f(x_t, x_{t+1}) + \beta V(x_{t+1})$ for the first t, we have $V(x_t) = \sum_{s \geqq t} \beta^{s-t} f(x_s, x_{s+1})$, so $V(x_{t+1}) > \sum_{s \geqq t+1} \beta^{s-t} f(x_s, x_{s+1})$, there is x' with $F(x') > F(x_{t+1}, \ldots)$ and $x'_0 = x_{t+1}$, we can modify x to be x' from x_{t+1} on and we obtain a higher value for F, absurd.

For the second assertion, note that $V(x_0) = f(x_0, x_1) + \beta V(x_1) = \cdots = \sum_{t=0}^{T-1} \beta^t f(x_t, x_{t+1}) + \beta^T V(x_T)$. Now by (A1) we have that F and therefore V is bounded, so if $T \to +\infty$ we obtain $V(x_0) = \sum_{t=0}^{\infty} \beta^t f(x_t, x_{t+1})$, and x maximizes F. ∎

Given $x \in X$ let $P(x) = \arg\max_{y \in G(x)} \{f(x, y) + \beta V(y)\}$. The last proposition tells us that, under (A1), $x \in \Omega(x_0)$ iff $x_{t+1} \in P(x_t)$ for all $t \in \mathbb{N}_0$.

Proposition 1.8.2 (Optimality principle)**.** If $W : X \to \mathbb{R}$ is bounded and, for all $x \in X$, it satisfies $W(x) = \max_{y \in G(x)} \{f(x, y) + \beta W(y)\}$ then $W = V$ and a solution to the problem exists.

Proof. Let $x_0 \in X$. Given $x \in \Omega(x_0)$ we have $W(x_0) \geqq f(x_0, x_1) + W(x_1) \geqq \cdots \geqq \sum_{t=0}^{T-1} \beta^t f(x_t, x_{t+1}) + \beta^T W(x_T)$. Since W is bounded, taking $T \to \infty$ we get $W(x_0) \geqq F(x)$. Since this holds for all $x \in \Omega(x_0)$, we get $W(x_0) \geqq V(x_0)$. On the other hand, by assumption there exists $x_1 \in G(x_0)$ with $W(x_0) = f(x_0, x_1) + \beta W(x_1)$, and in general $x \in \Omega(x_0)$ with $W(x_0) = \sum_{t=0}^{T-1} \beta^t f(x_t, x_{t+1}) + \beta^T W(x_T)$; taking $T \to \infty$ we get $W(x_0) = F(x) \leqq V(x_0)$. Combining the two inequalities we obtain $W(x_0) = V(x_0)$. Since this holds for all $x_0 \in X$, we obtain $W = V$. ∎

To prove existence of a solution we will use Banach's fixed point theorem, so let's recall it.

Definition 1.8.3. Let X be a metric space; we call $f : X \to X$ *contractive* if there exists $k \in (0,1)$ such that $d(f(x), f(y)) \leqq kd(x,y)$ for all $x, y \in X$. (This implies continuity trivially.)

Theorem 1.8.4 (Banach)**.** Let X be a complete metric space and $f : X \to X$ contractive; then it has an unique fixed point.

Proof. Let $x_1 \in X$ and $x_{n+1} = f(x_n)$ for $n \in \mathbb{N}$. We have $d(x_{n+1}, x_{n+2}) = d(f(x_n), f(x_{n+1})) \leqq kd(x_n, x_{n+1})$ y $d(x_n, x_{n+1}) \leqq k^{n-1}d(x_1, x_2)$; if $n \leqq m$, $d(x_n, x_m) \leqq d(x_n, x_{n+1}) + \cdots + d(x_{m-1}, x_m) \leqq (k^{n-1} + \cdots + k^{m-2})d(x_1, x_2) \leqq \frac{k^{n-1}}{1-k}d(x_1, x_2)$, then x_n is Cauchy and has a limit $x_n \to x$. We have $f(x_n) \to f(x)$, so $x = f(x)$, and f has a fixed point. Uniqueness is obvious. ■

Note that if f^n is contractive for some $n \in \mathbb{N}$, f has a fixed point: if $f^n(x) = x$, $f^n(f(x)) = f(f^n(x)) = f(x)$, $f(x)$ is also a fixed point of f^n and $f(x) = x$ follows by uniqueness.

Proposition 1.8.5. Let $T \neq \varnothing$ be a set and X a non-empty subset of the set of bounded functions $T \to \mathbb{R}$, closed by the sum of positive constants. Assume that $F : X \to X$ satisfies $F(f) \leqq F(g)$ if $f \leqq g$, and there exists $\delta \in (0,1)$ such that $F(f+\alpha) \leqq F(f)+\delta\alpha$ for all $f \in X$, $\alpha > 0$. Then F is a contraction.

Proof. Let $f, g \in X$. We have $-\|f-g\|_\infty \leqq f - g \leqq \|f-g\|_\infty$. Then $f \leqq g + \|f-g\|_\infty$, $F(f) \leqq F(g + \|f-g\|_\infty) \leqq F(g) + \delta\|f-g\|_\infty$. Similarly $F(g) \leqq F(f) + \delta\|f-g\|_\infty$. Then $\|F(f) - F(g)\|_\infty \leqq \delta\|f-g\|_\infty$, as desired. ■

Proposition 1.8.6 (Existence)**.** Under assumptions (A1) y (A2) there exists a solution to the problem. Moreover, P is compact-valued and upper-hemicontinuous.

Proof. Given a continuous and bounded $W : X \to \mathbb{R}$, let $\varphi(W) : X \to \mathbb{R}$ be given by $\varphi(W)(x) = \max_{y \in G(x)} \{f(x,y) + \beta W(y)\}$. By Berge's theorem and the assumptions, $\varphi(W)$ is continuous and bounded. Then $\varphi : CB(X) \to CB(X)$; by Proposition 1.8.5 we have that φ is contractive; then by Banach's fixed point theorem there exists W that satisfies the optimality principle, and we obtain the existence of solutions. The rest follows from Berge's theorem. ■

Assumption 1.8.2. (A3) X is a TVS, G is convex (as a subset of $X \times X$), f is concave in G and $f(\lambda(x,y) + (1-\lambda)(x',y')) > \lambda f(x,y) + (1-\lambda)f(x',y')$ if $\lambda \in (0,1)$ and $(x,y), (x',y') \in G$ with $x \neq x'$.

Proposition 1.8.7 (Uniqueness)**.** Under assumptions (A1)-(A3), there exists an unique solution to the problem. Moreover, P is a function and V is strictly concave.

Proof. We have to prove that, for all $x \in X$, $P(x)$ has cardinality one. Since $f(x,\cdot)$ is concave, it suffices to prove that V is strictly concave. It's easily verified that φ preserves concavity. Let $W \in CB(X)$ be constant; we have that $\varphi^n(W) \to V$ by Banach's fixed point theorem; since the pointwise limit of concave functions is concave, it results that V is concave. It remains to be proven that if $x \neq x'$ and $\lambda \in (0,1)$, $V(\lambda x + (1-\lambda)x') > \lambda V(x) + (1-\lambda)V(x')$. Now using $V = \varphi(V)$ and the last condition of (A3) we obtain the result. ■

Assumption 1.8.3. (A4) X is ordered, $f(\cdot, y)$ is increasing for all $y \in X$, and $G(x) \subset G(x')$ if $x \leqq x'$.

Proposition 1.8.8 (Monotonicity)**.** If assumptions (A1), (A2) and (A4) hold, then V in increasing.

Proof. Let $x < x'$. We have that, for a certain $y \in G(x)$, $V(x) = f(x,y) + \beta V(y) < f(x',y) + \beta V(y) \leqq \max_{y \in G(x')} \{f(x',y) + \beta V(y)\} = V(x')$. ■

Proposition 1.8.9 (Differentiability)**.** If assumptions (A1)-(A3) hold, $X \subset \mathbb{R}^k$, and $f(\cdot, y)$ is C^1 in X° for each $y \in X$ then V is differentiable in X° and $DV(x) = D_1 f(x, P(x))$ for all $x \in X^\circ$. Moreover, under (A4) and $X \subset \mathbb{R}^k_{\geqq 0}$, $x \in \Omega(x_0)$ solves the problem if it satisfies

$$D_2 f(x_t, x_{t+1}) + \beta D_1 f(x_{t+1}, x_{t+2}) = 0 \tag{1.8.1}$$

and the *transversality condition*

$$\lim_{t\to\infty} \beta^t D_1 f(x_t, x_{t+1})^\top x_t = 0. \tag{1.8.2}$$

Proof. Let $x_0 \in X^\circ$ and $x_1 = P(x_0)$. By (A2) G is continuous, and therefore if h is small, $x_1 \in G(x_0 + he_i)$. We have $V(x_0 + he_i) \geqq f(x_0 + he_i, x_1) + \beta V(x_1)$, so $V(x_0 + he_i) - V(x_0) \geqq f(x_0 + he_i, x_1) - f(x_0, x_1) = \partial_i f(x_0 + \xi e_i, x_1)h$. By Proposition 1.8.7, V is concave, thus

$$\partial_i f(x_0 + \xi e_i, x_1) \leqq \frac{V(x_0 + he_i) - V(x_0)}{h} \leqq \frac{V(x_0 - he_i) - V(x_0)}{-h} \leqq \partial_i f(x_0 + \xi' e_i, x_1)$$

for $h > 0$. With $h \to 0$ we get $\partial_i V(x_0) = \partial_i f(x_0, x_1)$, and the first result follows.

If $x \in \Omega(x_0)$ satisfies (1.8.1) and (1.8.2), and $x' \in \Omega(x_0)$,

$$\begin{aligned}
F(x) - F(x') = \sum_{t=0}^{\infty} \beta^t (f(x_t, x_{t+1}) - f(x'_t, x'_{t+1})) \geqq \\
&\geqq \liminf_{T\to\infty} \sum_{t=0}^{T} \beta^t (D_1 f(x_t, x_{t+1})^\top (x_t - x'_t) + D_2 f(x_t, x_{t+1})^\top (x_{t+1} - x'_{t+1})) = \\
&= \liminf_{T\to\infty} \Bigg[\sum_{t=0}^{T} (D_2 f(x_t, x_{t+1}) + \beta D_1 f(x_{t+1}, x_{t+2}))^\top (x_{t+1} - x'_{t+1}) - \\
&\qquad - \beta^{T+1} D_1 f(x_{T+1}, x_{T+2})^\top (x_{T+1} - x'_{T+1}) \Bigg] = \\
&= \liminf_{T\to\infty} \beta^{T+1} D_1 f(x_{T+1}, x_{T+2})^\top x'_{T+1} \geqq 0,
\end{aligned}$$

and x solves the problem. ■

Application (optimal growth) We have an economy with initial capital stock $x_0 > 0$. There is a production function $f : \mathbb{R}_{\geqq 0} \to \mathbb{R}_{\geqq 0}$ such that, given a capital stock x, generates $y = f(x)$ in a unit of time. A fraction $s_t \in [0, 1]$ of the production is invested (or saved) and gives a new capital stock $x_{t+1} = x_t + s_t y_t$. The rest, $c_t = (1 - s_t) y_t$, is consumed. From consumption we get a utility $\beta^t u(c_t)$ at present value, where $u : \mathbb{R}_{\geqq 0} \to \mathbb{R}$ and $\beta \in (0, 1)$ is the discount rate. The planner's problem is to maximize $\sum_{t=0}^{\infty} \beta^t u(c_t)$ given x_0. It's easier to state in terms of x: we want to

$$\begin{aligned}
\text{maximize} \quad & \sum_{t=0}^{\infty} \beta^t u(f(x_t) - x_{t+1} + x_t) \\
\text{subject to} \quad & 0 \leqq x_{t+1} \leqq x_t + f(x_t).
\end{aligned}$$

We assume (1) $u \in C^2(\mathbb{R}_{\geqq 0})$ is bounded with $u'_+(0) = +\infty$, $u' > 0$ y $u'' < 0$; (2) $f \in C^2(\mathbb{R}_{\geqq 0})$ is bounded, $f(0) = 0$, $f'_+(0) = +\infty$, $f' > 0$ y $f'' < 0$. We are under the conditions of Proposition 1.8.7, and therefore there is a unique solution to the problem. We have

$$\frac{u'(c_{t-1})}{u'(c_t)} = \beta(1 + f'(x_t)).$$

1.9 Calculus of variations

Theorem 1.9.1 (Euler-Lagrange)**.** Sea $\Omega \subset \mathbb{R}^2$ convexo, $\varphi : [0,T] \times \Omega \to \mathbb{R}$ continua y C^2 en el interior y $x_0 \in \mathbb{R}$. Sea $\mathcal{F} = \{f \in C^1[0,T] : f(0) = x_0, (f, f') \in \Omega\}$. Entonces si $f \in \mathcal{F}$ maximiza $f \in \mathcal{F} \mapsto \int_0^T \varphi(t, f, f')\, dt$, y $(f, f') \in \Omega^\circ$, vale

$$\partial_2\varphi(t, f, f') = \partial_t\partial_3\varphi(t, f, f') \tag{1.9.1}$$

para todo $t \in [0,T]$ y $\partial_3\varphi(T, f(T), f'(T)) = 0$.

Si además $\varphi(t, \cdot, \cdot)$ es cóncava para todo $t \in [0,T]$ y $f \in \mathcal{F}$ con $(f, f') \in \Omega^\circ$ cumple la ecuación (1.9.1), f maximiza $f \in \mathcal{F} \mapsto \int_0^T \varphi(t, f, f')\, dt$. Si $\varphi(t, \cdot, \cdot)$ es estrictamente cóncava para todo $t \in [0,T]$ y existe un máximo, es único.

Proof. Sea $g \in C^1(0,T)$ con soporte compacto. Veamos que existe $\epsilon > 0$ tal que $f + hg \in \mathcal{F}$ para todo $h \in (-\epsilon, \epsilon)$. Si no, para todo $n \in \mathbb{N}$ existe $h_n \in (-\frac{1}{n}, \frac{1}{n})$ con $f + h_n g \notin \mathcal{F}$, y por lo tanto existe $t_n \in (0,T)$ con $(f(t_n) + h_n g(t_n), f'(t_n) + h_n g'(t_n)) \notin \Omega$. Ahora, pasando a una subsecuencia, tenemos $t_n \to t \in [0,T]$. Entonces $(f(t_n) + h_n g(t_n), f'(t_n) + h_n g'(t_n)) \to (f(t), f'(t)) \in \Omega^\circ$, absurdo.

Entonces la función $F(h) = \int_0^T \varphi(t, f + hg, f' + hg')\, dt$ está definida para $h \in (-\epsilon, \epsilon)$ y tiene un máximo en $h = 0$. Es derivable, por lo que $F'(0) = 0$, es decir, $\int_0^T (\partial_2\varphi(t, f, f')g + \partial_3\varphi(t, f, f')g')\, dt = 0$. Ahora $\int_0^T \partial_3\varphi(t, f, f')g'\, dt = -\int_0^T \partial_t\partial_3\varphi(t, f, f')g\, dt$ por partes, así que tenemos

$$\int_0^T (\partial_2\varphi(t, f, f') - \partial_t\partial_3\varphi(t, f, f'))g\, dt = 0.$$

Esto vale para toda $g \in C^1(0,T)$ con soporte compacto, luego $\partial_2\varphi(t, f, f') - \partial_t\partial_3\varphi(t, f, f') = 0$ para todo $t \in (0,T)$, como queríamos. Si admitimos $g(T) \neq 0$, la misma cuenta da $\partial_3\varphi(T, f(T), f'(T)) = 0$. ■

Theorem 1.9.2. Sea $\Omega \subset \mathbb{R}^2$ convexo, $\varphi : \mathbb{R}_{\geq 0} \times \Omega \to \mathbb{R}$ continua y C^2 en el interior y $x_0 \in \mathbb{R}$. Sea $\mathcal{F} = \{f \in C^1(\mathbb{R}_{\geqq 0}) : f(0) = x_0, (f, f') \in \Omega\}$. Entonces si $f \in \mathcal{F}$ maximiza $f \in \mathcal{F} \mapsto \int_0^\infty \varphi(t, f, f')\, dt$, y $(f, f') \in \Omega^\circ$, vale

$$\partial_2\varphi(t, f, f') = \partial_t\partial_3\varphi(t, f, f')$$

para todo $t \in \mathbb{R}_{\geqq 0}$. Si además $\varphi(t, \cdot, \cdot)$ es cóncava para todo $t \in \mathbb{R}_{\geqq 0}$, vale el recíproco.

Proof. Si f_* maximiza, veamos que, para todo $T > 0$, $f_*|_{[0,T]}$ maximiza $f \in \mathcal{F}_T \mapsto \int_0^T \varphi(t, f, f')\, dt$, con $\mathcal{F}_T = \{f \in C^1[0,T] : f(0) = x_0, f(T) = f_*(T), (f, f') \in \Omega\}$. En efecto, si no, hay $h \in \mathcal{F}_T$ con $\int_0^T \varphi(t, h, h')\, dt > \int_0^T \varphi(t, f_*, f_*')\, dt$, y podemos formar $g \in C(\mathbb{R}_{\geqq 0}) \cap C^1([0,T) \cup (T, +\infty))$ dada por $g = h$ en $[0,T]$ y $g = f_*$ en $[T, +\infty)$; hay $g^* \in C^1(\mathbb{R}_{\geqq 0})$ con $\int_0^\infty \varphi(t, g^*, g^{*\prime})\, dt$ arbitrariamente cerca de $\int_0^\infty \varphi(t, g, g')\, dt$, y por lo tanto con $\int_0^\infty \varphi(t, g^*, g^{*\prime})\, dt > \int_0^\infty \varphi(t, f_*, f_*')\, dt$, absurdo. ■

Aplicación (crecimiento óptimo) Tenemos una economía con stock de capital inicial $x_0 > 0$. Existe una función de producción $f : \mathbb{R}_{\geqq 0} \to \mathbb{R}_{\geqq 0}$ que, dado un stock de capital x, genera una producción $f(x)$ por unidad de tiempo. La variación del capital en una unidad de tiempo, x', surge de la producción, que se divide en inversión (o ahorro) s y consumo c, de manera que $x' = s$, $s + c = f(x)$, $s \geqq 0$, $c \geqq 0$. Del consumo se obtiene una utilidad a tiempo presente $e^{-\rho t}u(c)$, donde $u : \mathbb{R}_{\geqq 0} \to \mathbb{R}$ es la función de utilidad y $\rho > 0$ es la tasa de descuento del futuro.

El problema del "planeador" es maximizar $\int_0^\infty e^{-\rho t} u(c)\,dt$ bajo la condición $0 \leqq c$, $x' = f(x) - c$, $x(0) = x_0$. Es más fácil plantear el problema en términos de x: buscamos

$$\begin{aligned} \text{maximizar} \quad & \int_0^\infty e^{-\rho t} u(f(x) - x')\,dt \\ \text{sujeto a} \quad & x \geqq 0, \quad x' \leqq f(x), \quad x(0) = x_0. \end{aligned}$$

Asumimos (1) $u \in C^2(\mathbb{R}_{\geqq 0})$ es acotada con $u'_+(0) = +\infty$, $u' > 0$ y $u'' < 0$; (2) $f \in C^2(\mathbb{R}_{\geqq 0})$ es acotada, $f(0) = 0$, $f'_+(0) = +\infty$, $f' > 0$ y $f'' < 0$. Si $\varphi(t, x, y) = e^{-\rho t} u(f(x) - y)$ estamos en las condiciones del teorema anterior; de hecho $\varphi(t, \cdot, \cdot)$ es cóncava. Por lo tanto una solución interior satisface $u'(f(x) - x')(f'(x) - \rho) = -u''(f(x) - x')(f'(x)x' - x'')$; poniendo $c = f(x) - x'$, esto es

$$\frac{c'}{c} = -\frac{u'(c)}{u''(c)c}(f'(x) - \rho).$$

1.10 Optimal control

Given T, $f : [0, T) \times R^d \times \mathcal{U} \to \mathbb{R}^d$, $g : [0, T) \times \mathbb{R}^d \times \mathcal{U} \to \mathbb{R}$ and x_0, we want to find $u : [0, T) \to \mathcal{U} \subset \mathbb{R}^n$ to

$$\begin{aligned} \text{maximize} \quad & \int_0^T g(t, x(t), u(t))\,dt \\ \text{subject to} \quad & x(0) = x_0, \\ & \dot{x}(t) = f(t, x, u). \end{aligned}$$

We call x the *state variable*. Let

$$V(t, x_0) = \max \left\{ \int_t^T g(s, x, u)\,ds : u : [t, T) \to \mathcal{U}, x(t) = x_0, \dot{x} = f(t, x, u) \right\}.$$

If V exists, it satisfies the *optimality principle*: for any $h > 0$,

$$V(t, x_0) = \max \left\{ \int_t^{t+h} g(s, x, u)\,ds + V(t + h, x(t + h)) : x(t) = x_0, \dot{x} = f(t, x, u) \right\}$$

and $V(T, \cdot) = 0$. If it is C^1, then this clearly implies the *Hamilton-Jacobi-Bellman (HJB) equation*

$$-\partial_t V(t, x) = \max_{u \in \mathcal{U}} \left\{ g(t, x, u) + \partial_x V(t, x) \cdot f(t, x, u) \right\}.$$

Given V, we can define $\mu(t, x) \in \arg\max_{u \in \mathcal{U}} \{ g(t, x, u) + \partial_x V(t, x) \cdot f(t, x, u) \}$, x given by $x(0) = x_0$ and $\dot{x} = f(t, x, \mu(t, x))$, and $u(t) = \mu(t, x)$. Now, the HJB equation implies that for an arbitrary u' with state x', we have $g(t, x', u') + \frac{d}{dt} V(t, x'(t)) \leqq 0$, so $\int_0^T g(t, x', u')\,dt \leqq V(0, x_0)$. For u and x this holds with equality, hence u solves the problem.

For $p : [0, T) \to \mathbb{R}^d$, let $H(t, x, u, p) = g(t, x, u) + p(t) \cdot f(t, x, u)$, the *Hamiltonian*. Take the HJB equation, and take $u = \mu(t, x)$. By the envelope theorem, differentiating in x and assuming $V \in C^2$, it implies

$$\dot{p} = -\partial_x H(t, x^*, u^*, p),$$

where u^* is a solution, x^* is its state, and $p = \partial_x V(t, x^*)$ is its *co-state*. Re-stating the HJB equation,

$$u^*(t) \in \arg\max_{u \in \mathcal{U}} H(t, x, u, p).$$

The boundary condition $V(T, \cdot) = 0$ implies $p(T) = 0$.

Consider the same problem, but with the additional constraint $x(T) = x_1$. The optimality principle and the HJB hold, but $p(T)$ is no longer necessarily 0, so the $x(T) = x_1$ replaces the boundary condition $p(T) = 0$. Now consider the same problem, but with the following objective function: $\int_0^T g(t, x(t), u(t))\, dt + q(T, x(T))$, and $T = \min\{t \geqq 0 : l(t, x(t)) = 0\}$, with l defining an n-dimensional smooth manifold in $\mathbb{R}_+ \times \mathbb{R}^d$. Again, the optimality principle holds, we get the boundary condition $V(T, x) = q(T, x)$ if $l(T, x) = 0$, from which we can derive a boundary condition for $p(T)$.

We can consider the infinite-horizon optimal control problem:

$$\begin{aligned} \text{maximize} \quad & \int_0^\infty e^{-rt} g(x(t), u(t))\, dt \\ \text{subject to} \quad & x(0) = x_0, \\ & \dot{x}(t) = f(x, u). \end{aligned}$$

Now the present-value $e^{rt}V(t, x)$ is constant, and if we call it $V(x)$ we have the Bellman equation

$$rV(x) = \max_{u \in \mathcal{U}} \{g(x, u) + V'(x) \cdot f(x, u)\}.$$

The present-value Hamiltonian is $H(x, u, p) = g(x, u) + p \cdot f(x, u)$ where $p = V'(x)$, and we have the equation

$$\dot{p} = rp - \partial_x H(x^*, u^*, p).$$

The transversality condition $\lim_{t \to \infty} e^{-rt} p(t) = 0$ could hold, but that is not always the case.

2 Game Theory

2.1 Dynamic games with perfect information

Tenemos un conjunto de *jugadores* $\mathcal{N}$. Hay un conjunto de *estados* $K \subset \mathbb{R}^n$ con $n \in \mathbb{N}$. Para todo jugador $i \in \mathcal{N}$ y todo estado $k \in K$ hay un conjunto de *acciones* $A_{ik} \subset \mathbb{R}^{n_i}$ con $n_i \in \mathbb{N}$, que forman un conjunto de perfiles de acciones $A_k = \prod_{i \in \mathcal{N}} A_{ik}$. Sea $A_i = \bigcup_{k \in K} A_{ik}$ y $A = \prod_{i \in \mathcal{N}} A_i$. Dado el estado k_t y el perfil de acciones $a_t \in A_{k_t}$, el estado k_{t+1} sigue una distribución de probabilidad con densidad $q(\cdot \mid k_t, a_t)$. Dado un tiempo $t \in \mathbb{N}$ $(= \{0, 1, \ldots\})$, tenemos el conjunto de *historias* $H_t = \{(k_0, a_0, \ldots, k_t, a_t) \mid a_s \in A_{k_s}$ para $s = 0, \ldots, t$ y $q(k_{s+1}|k_s, a_s) > 0$ para $s = 0, \ldots, t-1\}$. Una *estrategia pura* para el jugador i en el tiempo t es una función $\sigma_{ti} : H_{t-1} \times K \to A_i$ tal que $\sigma_{ti}(h_{t-1}, k_t) \in A_{ik_t}$ para todos $h_{t-1} \in H_{t-1}$, $k_t \in K$. Una *estrategia mixta* es una función $\sigma_{ti} : (h_{t-1}, k_t) \in H_{t-1} \times K \mapsto p \in \Delta(A_{ik_t})$, donde $\Delta(A_{ik_t})$ es el conjunto de distribuciones de probabilidad sobre A_{ik_t}. Sea S_{ti} el conjunto de estrategias σ_{ti}. Nos quedamos con estrategias mixtas ya que las puras son un caso particular, pero conservamos la distinción. Una estrategia para el jugador i es un vector $\sigma_i \in S_i = \prod_{t \in \mathbb{N}} S_{ti}$. Un perfil de estrategias es un vector $\sigma \in S = \prod_{i \in \mathcal{N}} S_i$. Para cada jugador i tenemos una *función de utilidad* instantánea $u_i : K \times A \to \mathbb{R}$, que asumimos continua y acotada. En el tiempo t, dada la realización de una historia $h_{t-1} \in H_{t-1}$ y un estado $k_t \in K$, el jugador i va a buscar maximizar su utilidad descontada a partir del tiempo t, esto es,

$$U_{it}(\sigma | h_{t-1}, k_t) = \sum_{s=0}^{\infty} \beta^s \mathbb{E}[u_i(k_{t+s}, a_{t+s}) | h_{t-1}, k_t, \sigma],$$

donde $\beta \in (0, 1)$ es el *factor de descuento*, $a_{t'i} \sim \sigma_{t'i}(h_{t'-1}, k_{t'})$ y $h_{t'} = (k_0, a_0, \ldots, k_{t'}, a_{t'})$ para todo $t' \in \mathbb{N}_{\geqq t}$. Dado $i \in \mathcal{N}$, $t \in \mathbb{N}$, $h_{t-1} \in H_{t-1}$, $k_t \in K$ y $\sigma^* \in S$, definimos

$$BR(\sigma^*_{-i} | h_{t-1}, k_t) = \operatorname*{arg\,max}_{\sigma_i \in S_i} \left\{ U_i((\sigma_i, \sigma^*_{-i}) | h_{t-1}, k_t) \right\}.$$

Definition 2.1.1. Un *subgame perfect equilibrium* (SPE) es un perfil de estrategias σ^* tal que, para todos $i \in \mathcal{N}$, $t \in \mathbb{N}$, $h_{t-1} \in H_{t-1}$ y $k_t \in K$, se tiene que $\sigma_i^* \in BR(\sigma_{-i}^*|h_{t-1}, k_t)$. Un *Markov perfect equilibrium* (MPE) es un SPE σ^* tal que $\sigma_{ti}^*(\cdot, k_t)$ es constante para todo $k_t \in K$, es decir, la estrategia sólo depende del estado, no de la historia.

Proposition 2.1.2. $\sigma^* \in S$ es un SPE sii para todos $i \in \mathcal{N}$, $t \in \mathbb{N}$, $h_{t-1} \in H_{t-1}$ y $k_t \in K$ vale que $U_{it}(\sigma^*|h_{t-1}, k_t)$ es el máximo de

$$\mathbb{E}[u_i(k_t, a_t)|h_{t-1}, k_t, \sigma_t] + \beta\mathbb{E}[U_{i,t+1}(\sigma^*|(h_{t-1}, k_t, a_t), k_{t+1})|h_{t-1}, k_t, \sigma_t]$$

sobre todos los perfiles de estrategias $\sigma_t = (\sigma_{ti}, \sigma_{t,-i}^*)$ con $\sigma_{ti} \in S_{ti}$.

Proof. Si σ^* es un SPE pero no sucede lo segundo, en (h_{t-1}, k_t) el jugador i tiene una estretegia mejor, absurdo. Si sucede lo segundo, sea σ' otro perfil de estrategias con $\sigma_{-i}^* = \sigma_{-i}'$. Veamos que $U_{it}(\sigma^*|h_{t-1}, k_t) \geqq U_{it}(\sigma'|h_{t-1}, k_t)$. Ahora, por la condición,

$$\begin{aligned}
U_{it}(\sigma^*|h_{t-1}, k_t) &\geqq \mathbb{E}[u_i(k_t, a_t)|h_{t-1}, k_t, \sigma_t'] + \beta\mathbb{E}[U_{i,t+1}(\sigma^*|(h_{t-1}, k_t, a_t), k_{t+1})|h_{t-1}, k_t, \sigma_t'] \geqq \\
&\geqq \mathbb{E}[u_i(k_t, a_t)|h_{t-1}, k_t, \sigma_t'] + \beta\mathbb{E}\Big[\mathbb{E}[u_i(k_{t+1}, a_{t+1})|h_t, k_{t+1}, \sigma_{t+1}'] + \\
&+ \beta\mathbb{E}[U_{i,t+2}(\sigma^*|h_{t+1}, k_{t+2})|h_t, k_{t+1}, \sigma_{t+1}'] \mid h_{t-1}, k_t, \sigma_t'\Big] = \\
&= \mathbb{E}[u_i(k_t, a_t)|h_{t-1}, k_t, \sigma_t'] + \beta\mathbb{E}[u_i(k_{t+1}, a_{t+1})|h_{t-1}, k_t, \sigma_t'] + \\
&+ \beta^2\mathbb{E}[U_{i,t+2}(\sigma^*|h_{t+1}, k_{t+2})|h_{t-1}, k_t, \sigma_t', \sigma_{t+1}'].
\end{aligned}$$

Siguiendo inductivamente obtenemos

$$U_{it}(\sigma^*|h_{t-1}, k_t) \geqq \sum_{s=0}^{T-1} \beta^s\mathbb{E}[u_i(k_{t+s}, a_{t+s})|h_{t-1}, k_t, \sigma'] + \beta^T\mathbb{E}[U_{i,t+T}(\sigma^*|h_{t+T-1}, k_{t+T})|h_{t-1}, k_t, \sigma'].$$

Con $T \to \infty$, usando que u_i es acotada, obtenemos $U_{it}(\sigma^*|h_{t-1}, k_t) \geqq U_{it}(\sigma'|h_{t-1}, k_t)$, como queríamos. ■

Proposition 2.1.3. Sea σ_{-i}^* de Markov, $h_{t-1} \in H_{t-1}$ y $k_t \in K$ tales que $BR(\sigma_{-i}^*|h_{t-1}, k_t) \neq \varnothing$. Entonces existe $\sigma_i^* \in BR(\sigma_{-i}^*|h_{t-1}, k_t)$ de Markov.

Proof. Es obvio en vista de la proposición anterior ya que si σ_{-i}^* es de Markov la historia no entra en el problema de minimización que cumple $\sigma_{t'i}^*$ para ningún $t' \geqq t$. ■

Theorem 2.1.4 (Existencia de MPE)**.** Si $\mathcal{N}$, K y A son finitos entonces existe un MPE. Si $\mathcal{N}$ y K son finitos, y para todos $i \in \mathcal{N}$ y $k \in K$ vale que A_{ik} es convexo, $u_i(k, \cdot)$ es continua y, para todo $a_{-i} \in A_{-ik}$, $u_i(k, \cdot, a_{-i})$ es cuasicóncava, entonces existe un MPE en estrategias puras.

Proof. Consideramos el juego en el que los jugadores son (i, k) con $i \in \mathcal{N}$, $k \in K$, y el conjunto de estrategias de (i, k) es $\Delta(A_{ik})$ con pagos $v_{(i,k)}(\sigma) = U_{i0}(\sigma|k)$. Hay un equilibrio de Nash σ. Construimos $\sigma^* \in S$ dado por $\sigma_{ti}^*(\cdot, k) = \sigma_{(i,k)}$. En base a los lemas anteriores se ve que es un MPE. Lo segundo sale igual. ■

Los juegos repetidos pueden verse como juegos dinámicos con un único estado. Un MPE es un equilibrio de Nash del juego estático. Con SPEs podemos obtener algo mejor.

Theorem 2.1.5 (Folk theorem)**.** Sea $\mathcal{N}$ finito, A_i compactos para cada $i \in \mathcal{N}$, $a_* \in A$ un equilibrio de Nash del juego estático y $a \in A$ tal que $u_i(a) > u_i(a_*)$ para todo $i \in \mathcal{N}$. Entonces existe $\beta_* \in [0, 1)$ tal que si $\beta \geqq \beta_*$ hay un SPE en el que en cada paso los jugadores eligen a.

Proof. Defino una estrategia σ en la que el jugador i en el tiempo t dada la historia h_{t-1} elige a_i si se eligió a en el paso y elige a_{*i} si no. Veamos que es un SPE usando la Proposición 2.1.2. En el primer caso el jugador opta entre recibir $\frac{a_i}{1-\beta}$ y $v_i + \frac{\beta}{1-\beta} a_{*i}$. Si $\beta \leqq \beta_i = \frac{\max u_i - a_i}{\max u_i - a_{*i}} < 1$, elige a_i. En el segundo caso, a_{*i} es su mejor opción. Entonces si $\beta \leqq \beta_* = \max_{i \in \mathcal{N}} \beta_i$, σ es un SPE, como queríamos. ■

Theorem 2.1.6 (Folk theorem)**.** Sea $\mathcal{N}$ finito y A_i compactos para cada $i \in \mathcal{N}$. Para cada jugador i sea $\underline{v}_i = \min_{a_{-i} \in A_{-i}} \max_{a_i \in A_i} \{u_i(a_i, a_{-i})\}$. Sean $a, a'_j \in A$ para cada $j \in \mathcal{N}$ tales que $u_i(a) > u_i(a'_j) > \underline{v}_i$ y $u_i(a'_j) > u_i(a'_i)$ si $i \neq j$ para todos $i, j \in \mathcal{N}$. Entonces existe $\beta_* \in [0, 1)$ tal que si $\beta \geqq \beta_*$ hay un SPE en el que en cada paso los jugadores eligen a.

Proof. Sea $m_i \in \arg\min_{a_{-i} \in A_{-i}} \max_{a_i \in A_i} u_i(a_i, a_{-i})$, $v_i = u_i(a)$, $v'_i = \min_{j \neq i} \{u_i(a'_j)\}$, $v''_i = u_i(a'_i)$, $\overline{v}_i = \max_a u_i(a)$. Tengo $v_i > v'_i > v''_i > \underline{v}_i$. Sea $T \in \mathbb{N}$ tal que $(T+1)v''_i > \overline{v}_i + T\underline{v}_i$ para todo i. La estrategia para i es como sigue. Si se hizo a en el pasado, elige a_i. Si en el último paso j se desvió, elige m_i^j T veces, y luego elige a'_{ji} por siempre, mientras los demás hagan lo mismo. Veamos que es un SPE.

Primero, si se hizo a en el pasado, tiene que darse $v_i > (1-\beta)\overline{v}_i + \beta(1-\beta^T)\underline{v}_i + \beta^{T+1}v''_i$; si $\beta = 1$ se cumple, luego para $\beta < 1$ suficientemente grande también. Segundo, si $j \neq i$ se desvió y resta hacer $T' \leqq T$ veces m_j, tiene que darse $(1-\beta^{T'})u_i(m_j) + \beta^{T'}v'_i > (1-\beta)\overline{v}_i + (1-\beta^T)\underline{v}_i + \beta^{T+1}v''_i$; si $\beta = 1$ se cumple, luego para $\beta < 1$ suficientemente grande también. Tercero, si i se desvió y resta hacer $T' \leqq T$ veces m_i, tiene que darse $(1-\beta^{T'})\underline{v}_i + \beta^{T'}v''_i > (1-\beta^{T+1})\underline{v}_i + \beta^{T+1}v''_i$, o sea $\sum_{t=0}^{T'-1} \beta^t \underline{v}_i + \sum_{t=T'}^{T} \beta^t v''_i > \sum_{t=1}^{T} \beta^t \underline{v}_i$, que se cumple para todo β. Cuarto, si hay que elegir a'_j, tiene que darse $v''_i > (1-\beta)\overline{v}_i + (1-\beta^T)\underline{v}_i + \beta^{T+1}v''_i$, es decir, $(1-\beta^{T+1})v''_i > (1-\beta)\overline{v}_i + (1-\beta^T)\underline{v}_i$, o sea $\sum_{t=0}^{T} \beta^t v''_i > \overline{v}_i + \sum_{t=1}^{T} \beta^t \underline{v}_i$; si $\beta = 1$ es $(T+1)v''_i > \overline{v}_i + T\underline{v}_i$, que se cumple; luego para $\beta < 1$ suficientemente grande también. ■

2.2 Games with imperfect information

There is a set of players $\mathcal{N}$. There is a set of *histories* H such that $\varnothing \in H$, if $(a_1, \ldots, a_t) \in H$ then $(a_1, \ldots, a_s) \in H$ for all $s \leqq t$, and if $h = (a_t)_{t \in \mathbb{N}}$ satisfies that $(a_t)_{t=1}^T \in H$ for all $T \in \mathbb{N}$ then $h \in H$. If $h \in H$ is a history, let $A(h) = \{a : (h, a) \in H\}$ be the set of actions available after h. If $A(h) = \varnothing$, which happens if h is infinite, then h is *terminal*. The set of terminal histories is denoted Z. There is a function $P : H \smallsetminus Z \to \mathcal{N}$ that assigns to each non-terminal history the next player. For each player i let $\mathcal{I}_i$ be a partition of the set $\{h \in H \smallsetminus Z : P(h) = i\}$ such that $A(h) = A(h')$ for each $h, h' \in I \in \mathcal{I}_i$; each $I \in \mathcal{I}_i$ is an information set; let's call $A(I) = A(h)$ for any $h \in I$. Each player has an utility function over terminal histories $u_i : Z \to \mathbb{R}$. A pure strategy for $i \in \mathcal{N}$ is a function $\sigma_i : I \in \mathcal{I}_i \mapsto \sigma_i(I) \in A(I)$; a strategy is a function $\sigma_i : I \in \mathcal{I}_i \mapsto \sigma_i(I) \in \Delta A(I)$; let σ be the profile of strategies $(\sigma_i)_{i \in \mathcal{N}}$. A *belief system* is a function μ that maps each information set $I \in \mathcal{I}_i$ to a probability distribution over I.

We assume that there is *perfect recall*, which is defined as follows. Let $h \in H$, $h = (a_1, \ldots, a_T)$. For $t = 1, \ldots, T$ such that $P(a_1, \ldots, a_t) = i$, let $X_i(h)$ be the sequence of a_{t+1} if $t < T$ and the information sets $I \in \mathcal{I}_i$ such that $(a_1, \ldots, a_t) \in I$. Perfect recall means that for any $h, h' \in I \in \mathcal{I}_i$ we have that $X_i(h) = X_i(h')$.

Given σ, μ, for $I \in \mathcal{I}_i$ we define the *outcome* conditional on I, $O(I)$, as a probability distribution over terminal histories defined as follows. If $h = (a_1, \ldots, a_T) \in Z$, $h' = (a_1, \ldots, a_t)$ for $t < T$, and $h' \in I$ then $O(I)(h) = \mu(I)(h') \prod_{s=t}^{T-1} \sigma_{P(a_1, \ldots, a_s)}(a_1, \ldots, a_s)(a_{s+1})$. If there is no sub-history h' of h in I then $O(I)(h) = 0$. We say that σ, μ are *sequentially rational* if, for any $I \in \mathcal{I}_i$, σ_i maximizes $\int_{h \in Z} u_i(h) O(I)(h)$. We say that σ, μ are *consistent* if there is a sequence σ^n, μ^n that converges to σ, μ such that σ^n is completely mixed (i.e., $\sigma_i^n(I)$ assigns positive density to any action) and μ^n is derived from σ^n using Bayes' rule.

Definition 2.2.1. A *sequential equilibrium* of a finite game with perfect recall is a sequentially rational and consistent pair σ, μ.

To prove existence, let us consider the following. A *trembling-hand equilibrium* of a finite static simultaneous-move game is a profile of strategies σ such that there is a sequence σ^k of completely mixed strategies such that $\sigma^k \to \sigma$ and σ_i is a best response to σ^k_{-i} for all $i \in \mathcal{N}, k \in \mathbb{N}$.

Proposition 2.2.2. In a finite static simultaneous-move game a trembling-hand equilibrium exists, and it is a Nash equilibrium.

Proof. For each $k \in \mathbb{N}$ let $\sigma^k \in \prod_{i\in\mathcal{N}} \Delta A_i$ be an equilibrium for the game with payoffs $u_i((\frac{1}{k}\frac{1}{|A_i|} + (1-\frac{1}{k})\sigma^k_i)_{i\in\mathcal{N}})$. By compactness we can pass to a convergent subsequence $\sigma^k \to \sigma$. Let k be large enough so that for each $i \in \mathcal{N}$ and $a \in A_i$ such that $\sigma_i(a) > 0$ we have that $\sigma^k_i(a) > 0$. This implies that if $\sigma_i(a) > 0$ then $u_i(a, \sigma^k_{-i})$ is maximum over $a \in A_i$. Hence σ_i is a best response to σ^k_{-i} for each k large enough, and σ is a trembling-hand equilibrium. ■

Now, consider a dynamic game of imperfect information. Define a static game of simultaneous moves as follows: for each information set $I \in \mathcal{I}_i$ define an agent with the same utility as i and actions $A(I)$. We define a *trembling-hand perfect equilibrium* for the dynamic game as a trembling-hand equilibrium of the derived static game. We have proved existence.

Proposition 2.2.3. A trembling-hand equilibrium is a sequential equilibrium.

Proof. Let σ be a trembling hand equilibrium, and σ^k completely mixed so that $\sigma^k \to \sigma$. Let μ^k be a belief system derived from σ^k by Bayes rule, pass to a convergent subsequence and take $\mu = \lim \mu^k$. It is easy to see that σ, μ are sequentially rational. ■

Now let us focus on games with observable actions, but in which there is a payoff-relevant characteristic of the agents that is not common knowledge. We allow for simultaneous moves, so, take a set of histories as before, but assume that every agent acts at each time, and histories are known. For each player $i \in \mathcal{N}$ there is a set of *types* Θ_i, and let $p_i \in \Delta\Theta_i$ be the prior distribution of types and $\Theta = \prod_{i\in\mathcal{N}} \Theta_i$. A (behavioral) strategy is a function $\sigma_i : (h, \theta_i) \in H \setminus Z \times \Theta_i \to \sigma_i(h, \theta_i) \in \Delta\pi_i(A(h))$. Let $\sigma = (\sigma_i)_{i\in\mathcal{N}}$ be a profile of strategies. Each $i \in \mathcal{N}$ has a utility $u_i : \Theta \times Z \to \mathbb{R}$ over terminal histories that depends on the types. For each agent $i \in \mathcal{N}$ let $\mu_i : H \to \Delta\Theta_i$ be the common belief of the other agents on her type after each history. Let $\mu = (\mu_i)_{i\in\mathcal{N}}$ be a profile of common beliefs. For each history $h = (a_1, \ldots, a_t)$ let the *outcome* $O_i(h, \theta_i)$ for $i \in \mathcal{N}$ be a probability distribution over terminal histories $h' = (a_1, \ldots, a_t, \ldots, a_T)$ such that $O_i(h, \theta_i)(h') = \int_{\theta_{-i}\in\Theta_{-i}} \prod_{s=t}^{T} \prod_{j\in\mathcal{N}} \sigma_j(a_1, \ldots, a_s, \theta_j)(a_{sj})\, d\mu(a_1, \ldots, a_{t-1})(\theta_{-i})$.

We say that σ, μ are *sequentially rational* if for each history $h \in H$ and each $\theta_i \in \Theta_i$ we have that $\sigma_i(\cdot, \theta_i)$ maximizes $\int_{h'\in Z} u_i(h', \theta_i) O_i(h, \theta_i)(h')$ for each $h \in H \setminus Z$.

Definition 2.2.4. A *perfect bayesian equilibrium* (PBE) is a pair σ, μ that is

(a) sequentially rational,

(b) $\mu_i(\varnothing) = p_i$ for each $i \in \mathcal{N}$,

(c) if $h \in H \setminus Z$, $a, a' \in A(h)$, $a_i = a'_i$ then $\mu_i(h, a) = \mu_i(h, a')$, and

(d) for $a \in A(h)$,

$$\mu_i(h, a)(\theta'_i) = \frac{\sigma_i(h, \theta'_i)(a_i)\mu_i(h)(\theta'_i)}{\sum_{\theta_i\in\Theta_i} \sigma_i(h, \theta_i)(a_i)\mu_i(h)(\theta_i)}$$

if the denominator is not zero.

We can interpret a game with observable actions as a game of imperfect information in which the first mover is Nature and chooses the types of each agent with the specified probabilities, so that only i knows her type after that move.

Proposition 2.2.5. A sequential equilibrium is a PBE.

Proof. Let σ, μ by a SE and σ^k, μ^k given by consistency. For each history h define $\tilde{\mu}_i(h)(\theta_i)$ as the limit of $\Pr(h|\theta_i)$ under σ^k, passing to a subsequence if necessary. Clearly $\tilde{\mu}$ satisfies the conditions of a PBE, and for $h \in I \in \mathcal{I}_i$, $h_1 = \theta_{-i} \in \Theta_{-i}$ and $\mu(I)(h) = \prod_{j \in \mathcal{N} \setminus \{i\}} \tilde{\mu}_j(h)(\theta_i)$. ■

A *signaling game* is a game with two players, the sender S and the receiver R. The sender knows her type $\theta \in \Theta$ and sends a message $m \in M$; the receiver then chooses an action $a \in A$; payoffs depend on θ, m, a. Let $p \in \Delta\Theta$ be the prior probabilities on types, $\sigma_S : \Theta \to \Delta M$ and $\sigma_R : M \to \Delta A$ be the strategies, and $\mu : M \to \Delta\Theta$ be the posterior probability on types based on message. We consider PBEs, so we assume that

1. $\sigma_S(\theta)(m) > 0$ only if $m \in \arg\max_{m \in M} u_S(\theta, m, \sigma_R(m))$,
2. if $\sigma_S(\tilde{\theta})(m) > 0$ for some $\tilde{\theta} \in \Theta$ then

$$\mu(m)(\theta) = \frac{\sigma_S(\theta)(m)p(\theta)}{\sum_{\tilde{\theta} \in \Theta} \sigma_S(\tilde{\theta})(m)p(\tilde{\theta})},$$

and

3. $\sigma_R(m)(a) > 0$ only if $a \in \arg\max_{a \in A} u_R(\mu(m), m, a)$.

We want to impose further restrictions on μ to rule out unreasonable off-equilibrium beliefs. Let $m \in M$ be an off-equilibrium message, i.e., such that there is no $\theta \in \Theta$ with $\sigma_S(\theta)(m) > 0$. Let $\theta \in \Theta$, $u_S^*(\theta) := \sum_{m' \in \sigma_S(\theta)} u_S(\theta, m', \sigma_R(m'))$ and $\mathrm{BR}(T, m) := \bigcup_{\tilde{\mu} \in \Delta T} \arg\max_{\sigma \in \Delta A} u_R(\tilde{\mu}, m, \sigma)$ for $T \subset \Theta$. Suppose that $u_S^*(\theta) > \sup_{\sigma \in \mathrm{BR}(\Theta, m)} u_S(\theta, m, \sigma)$, then by sending m the best that S can get is worse than what she gets if she follows σ_S. Hence if R observes m, she can infer that θ is not the true type. We can "prune" the pair (θ, m). This idea leads to the following equilibrium refinement.

Definition 2.2.6 (The intuitive criterion)**.** For each off-equilibrium $m \in M$, let $S(m)$ be the set of $\theta \in \Theta$ s.t. $u_S^*(\theta) > \sup_{\sigma \in \mathrm{BR}(\Theta, m)} u_S(\theta, m, \sigma)$. If there is $\theta \in \Theta$ such that $u_S^*(\theta) < \inf_{\sigma \in \mathrm{BR}(\Theta \setminus S(m), m)} u_S(\theta, m, \sigma)$ then the equilibrium fails the *intuitive criterion* (**?**).

A stronger refinement, called *universal divinity*, is defined as follows. For each off-equilibrium message m, we maintain a set $T(m) \subset \Theta$ of plausible types, which starts being $T(m) = \Theta$. For each $\theta \in T(m)$ we define $D_\theta := \{\sigma \in \mathrm{BR}(T(m), m) : u_S^*(\theta) < u_S(\theta, m, \sigma)\}$ and $D_\theta^0 := \{\sigma \in \mathrm{BR}(T(m), m) : u_S^*(\theta) = u_S(\theta, m, \sigma)\}$. If $\theta \in T(m)$ is such that $D_\theta \cup D_\theta^0 \subset \bigcup_{\theta' \in T(m) \setminus \{\theta\}} D_{\theta'}$ then we remove θ from $T(m)$. If $D_\theta = BR(T(m), m)$ for some $\theta \in T(m)$, the equilibrium fails the test. **?** show that a universally divine equilibrium always exists.

Application (Job market signaling) There is a job applicant S who has an privately known ability $\theta \in \Theta = \{\theta_1, \ldots, \theta_n\}$, with $\theta_1 < \cdots < \theta_n$ and $\Pr(\theta = \theta_i) = p_i$. She chooses a level of education $e \geqq 0$ at cost $c(e, \theta)$ which is public, where $c_e > 0$, $c_\theta < 0$, $c_{e\theta} < 0$. An employer R hires her and pays her a wage $w = \mathbb{E}_R(\theta|e)$; we can think that $u_R(\theta, e, w) = -(w - \theta)^2$. The applicant's utility is $u = w - c(e, \theta)$. It is easy to check that in any PBE we have that if $\sigma_S(\theta)(e), \sigma_S(\theta')(e') > 0$ for $\theta < \theta'$ then $e \leqq e'$.

Suppose that there are only two types. Let's assume that σ_S, μ is a *pooling equilibria*, i.e., there is e s.t. $\sigma_S(\theta_1)(e), \sigma_S(\theta_2)(e) > 0$. Let $\hat{e}$ be s.t. $w(e) - c(e, \theta_1) = \theta_2 - c(\hat{e}, \theta_1)$. If $e' > \hat{e}$, $u_S^*(\theta_1) > \theta_2 - c(e', \theta_1)$, which is the best possible outcome. Therefore θ_1 cannot choose e', so θ_2 can choose it and get $\theta_1 - c(e', \theta_1)$, which is greater than $w(e) - c(e, \theta_2) < \theta_2 - c(\hat{e}, \theta_2)$ if $e' > \hat{e}$ is small enough. So the equilibrium does not pass the intuitive criterion. Therefore the only equilibria that can satisfy it are *separating*, and in fact the only equilibrium is $\sigma(\theta_1) = 0$, $\sigma(\theta_2) = e$ s.t. $\theta_2 - c(e, \theta_1) = \theta_1 - c(0, \theta_1)$.

For n types, there could be pooling equilibria that pass the intuitive criterion. We can see that only the separating equilibrium with minimal education passes the universal divinity criterion. First, we show that each type, from θ_n to θ_1, will separate: if θ, θ' pool at e with $\theta < \theta'$, let $e' > e$ be off-equilibrium; then we can prune (θ, e'), so θ' will separate by choosing e', as desired. Second, by a similar argument we show that each type, from θ_2 to θ_n, will separate by the least amount, i.e., $\sigma(\theta_i) = e_i$ with $e_1 = 0$ and $\theta_{i+1} - c(e_{i+1}, \theta_i) = \theta_i - c(e_i, \theta_i)$.

2.3 Bargaining

Nash bargaining model There are two agents, a convex compact set $U \subset \mathbb{R}^2$, and a status quo $v \in \mathbb{R}^2$ s.t. $\forall u \in U. u \geqq v$. The agents have to choose $u \in U$; agent i gets utility u_i for $i = 1, 2$. A solution is a function $f : \mathcal{U} \to \mathbb{R}^2$ where $\mathcal{U}$ is the set of (U, v) such that, if $u = f(U, v)$ then (1) $u \in U$, (2) u is Pareto optimal in U, i.e., $\forall u' \in U. u' \geqq u \Rightarrow u' = u$, (3) if U is symmetric, i.e., $(x, y) \in U \Leftrightarrow (y, x) \in U$, then $u_1 = u_2$, (4) if $U' \subset U$ and $u \in U'$ then $f(U', v) = u$, and (5) if $T(x) = \operatorname{diag}(\alpha)x + \beta$ where $\alpha \in \mathbb{R}^2_{>0}$, $\beta \in \mathbb{R}^2$, then $f(TU, Tv) = Tu$.

Theorem 2.3.1. The unique solution to the Nash bargaining problem is

$$f(U, v) = \arg\max_{u \in U} \{(u_1 - v_1)(u_2 - v_2)\}.$$

Proof. Let u be the proposed solution. It clearly satisfies the axioms. Let f' be another solution. By (5) we can assume that $u = (1, 1)$ and $v = (0, 0)$. We have $U \subset U' = \{x \in \mathbb{R}^2_{\geqq 0} : x_1 + x_2 \leqq 2\}$ by the definition of u, and by (1–3) we have $f'(U', 0) = u$, therefore by (4) we have $f'(U, 0) = u$, as desired. ∎

Suppose A1 and A2 have to split a pie of measure 1, so each gets $(x, 1-x)$. Then A1 should get $\frac{1}{2} + \frac{1}{2}(v_1 - v_2)$. The outside option gives bargaining power.

Rubinstein-Ståhl model There are two agents that have to choose $x \in [0, 1]$; agent 1 gets x and agent 2 gets $1 - x$. They propose in turns until one accepts the other's proposal. Agent i discounts the future by $\delta_i \in [0, 1)$. Strategy: A1 proposes $x_1 = \frac{1-\delta_2}{1-\delta_1\delta_2}$ and A2 proposes $x_2 = \frac{\delta_1(1-\delta_2)}{1-\delta_1\delta_2}$; A1 accepts any offer $x \geqq x_2$ and rejects otherwise; A2 accepts any offer $x \leqq x_1$ and rejects otherwise.

Theorem 2.3.2. The strategy is the unique SPE of the game.

Proof. That it is a SPE is clear. To see that it is unique, let $\underline{v}, \overline{v}$ be the infimum and supremum payoffs that A1 can get in a SPE starting at some time when he proposes, and $\underline{w}, \overline{w}$ when A2 proposes. If A1 proposes x, A2 accepts it if $1 - x \geqq \delta_2(1 - \underline{w})$ and rejects it if $1 - x \leqq \delta_2(1 - \overline{w})$; thus $1 - \underline{v} \leqq \delta_2(1 - \underline{w})$ and $1 - \overline{v} \geqq \delta_2(1 - \overline{w})$. If A2 proposes x, A1 accepts if $x \geqq \delta_1\overline{v}$ and rejects it if $x \leqq \delta_1\underline{v}$; thus $\overline{w} \leqq \delta_1\overline{v}$ and $\underline{w} \geqq \delta_1\underline{v}$. This gives $\underline{v} = \overline{v} = x_1$ and $\underline{w} = \overline{w} = x_2$, and the strategy is the one proposed. ∎

Suppose the game stops with probability α, in which case they receive v_1 and v_2. By the same argument, A1 gets

$$\frac{1 - \delta_2(1 - \alpha) - \delta_2\alpha v_2 + \delta_1\delta_2\alpha(1 - \alpha)v_2}{1 - \delta_1\delta_2(1 - \alpha)^2}.$$

By letting $\delta_1, \delta_2 \to 1$ and then $\alpha \to 0$, A1 gets $\frac{1}{2} + \frac{1}{2}(v_1 - v_2)$, as in Nash bargaining.

Shapley value Let $N = \{1, \dots, n\}$ be a set of players, and V the set of functions $v : 2^N \to \mathbb{R}$ such that $v(\varnothing) = 0$. A *value* is a function $\phi : V \to \mathbb{R}^N$, i.e., a function that maps coalitional payoffs v to individual payoffs $\phi_i(v)$ (for $i \in N$).

A value ϕ satisfies *efficiency* iff $\sum_{i \in N} \phi_i(v) = v(N)$. We say that $i, j \in N$ are symmetric iff for each $S \subset N$ such that $i, j \notin S$, $v(S \cup i) = v(S \cup j)$; ϕ satisfies *symmetry* iff for every $i, j \in N$ symmetric we have $\phi_i(v) = \phi_j(v)$. We say that i is a dummy player iff $v(S \cup i) = v(S)$ for every $S \subset N$; ϕ satisfies *dummy* iff for every dummy $i \in N$ we have $\phi_i(v) = 0$. Finally, ϕ satisfies *additivity* iff $\phi(v + w) = \phi(v) + \phi(w)$.

Theorem 2.3.3. There exists a unique value ϕ that satisfies efficiency, symmetry, dummy and additivity, and it is given by

$$\phi_i(v) = \frac{1}{n!} \sum_{\sigma \in \Sigma} [v(p_i^\sigma \cup i) - v(p_i^\sigma)] = \frac{1}{n} \sum_{S \subset N \smallsetminus i} \frac{1}{\binom{n-1}{|S|}} [v(S \cup i) - v(S)],$$

where Σ is the set of permutations of N and $p_i^\sigma = \{j \in N : \sigma(j) < \sigma(i)\}$.

Proof. Given $R \subset N$ non-empty, the unanimity game v_R is given by $v_R(S) = \mathbb{1}(S \subset R)$. Clearly $i \notin R$ are dummies and $i, j \in R$ are symmetric, so $\phi_i(av_R) = \frac{a}{|R|}\mathbb{1}(i \in R)$ for any $a \in \mathbb{R}, i \in N$. The unanimity games clearly form a basis of the vector space V, so ϕ is indeed unique. It is easy to check that the formula satisfies the conditions. ■

Hart and Mas-Colell (1996) This is a game of perfect of information that implements the Shapley value. We assume monotonicity: $v(S \cup i) \geqq v(S) + v(i)$ if $i \notin S$.[2] At each stage, there is a non-empty set $S \subset N$ of active players. A member $i \in S$ is selected uniformly at random. That member makes a proposal $a_i \in \mathbb{R}^S$ such that $\sum_{j \in S} a_{ij} \leqq v(S)$, and every member votes yes or no; if "yes" is the unanimous answer, each one receives a_{ij}; otherwise, i is removed from S with probability $1 - \rho$, where $\rho \in [0, 1)$, and gets $v(i)$, and they move to the next stage. The solution concept is MPE, where the state is (S, i). We assume that players are risk neutral and when indifferent they make the decision which leads to earliest termination.

Theorem 2.3.4. There is a unique MPE: at (S, i), i proposes $a_{ij} = \rho\phi_j(v|_S) + (1 - \rho)\phi_j(v|_{S \smallsetminus i})$ for $j \neq i$ and $a_{ii} = v(S) - \sum_{j \in S \smallsetminus i} a_{ij}$; every player accepts. Each player gets $\phi_i(v)$ in the first round in expectation, and when $\rho \to 1$, they get $\phi_i(v)$ always.

Proof. By induction on $|S|$. For $|S| = 1$ it is obvious. Take S, and for $i, j \in S$ let a_{ij} be what i proposes to j in a MPE. Let $a_j = \frac{1}{|S|} \sum_{k \in S} a_{kj}$. If i gives $d_{ij} = \rho a_j + (1 - \rho)\phi_j(v|_{S \smallsetminus i})$ to each $j \neq i$ and accepts, she receives $d_{ii} = v(S) - \sum_{j \in S \smallsetminus i} d_{ij}$. Otherwise someone does not accept and she gets $\rho a_i + (1 - \rho)v(i)$. It is clear that d_{ii} is not lesser than this, so we get $a_{ij} = d_{ij}$ for any $i, j \in S$ and everybody accepts in the MPE. This is a linear system of equations with at most one solution, and the proposed strategies verify it. A crucial formula for proving this is

$$|S|\phi_i(v|_S) + \sum_{j \in S \smallsetminus i} \phi_i(v|_{S \smallsetminus j}) = v(S) - v(S \smallsetminus i).$$

Clearly the result is a MPE. ■

[2] Let $\mathcal{P}(S)$ be the set of partitions of S. We can define $\overline{v}(S) = \max_{P \in \mathcal{P}(S)} \sum_{R \in P} v(R)$, and assume that the agents can make that partition. Clearly $\overline{v}$ satisfies super-additivity: $\overline{v}(S \cup T) \geqq \overline{v}(S) + \overline{v}(T)$ for any $S \cap T = \varnothing$. We can just assume that. Of course, this depends on how we assume that agents can deal with negative externalities. If these are inevitable, the Shapley value will sanction a tax on individuals who do harm to society, which may not be compatible with a non-cooperative protocol unless there is an arbiter who can force individuals to pay.

Comment. To get Shapley values it is crucial that the person who proposes can credibly threaten to leave (with positive exogenous probability). Suppose that, again, a random member makes a proposal and it is implemented iff there is consensus; suppose that everybody discounts the future with factor δ. Then in the MPE the one who proposes gets $((1-\delta)n+\delta)\frac{v(N)}{n}$ and the rest get $\delta\frac{v(N)}{n}$; everybody gets $\frac{v(N)}{n}$ in expectation.

Baron-Ferejohn

Banks-Duggan model

Add Unions/strikes:

- Kennan, John. 1986. "Chapter 19 The Economics of Strikes." In Handbook of Labor Economics, Elsevier, 1091–1137.
- Shapley value
- Gul 1989
- Baron-Ferejohn

3 Economics

3.1 Competitive markets

Demand A consumer has preferences $\succsim$ over bundles of goods from the connected set $X \subset \mathbb{R}^L$; we define $x \sim y$ if $x \succsim y$ and $y \succsim x$, and $x \succ y$ if $x \succsim y$ but $y \not\succsim x$. We say that $\succsim$ is *rational* if it is transitive and complete; *continuous* if $\{y : y \succsim x\}$ and $\{y : x \succsim y\}$ are closed sets for each $x \in X$; *monotone* if $x > y$ implies $x \succ y$; *locally non-satiated* if for each $x \in X$ and $\epsilon > 0$ there is $y \in B_\epsilon(x) \cap X$ such that $y \succ x$; *convex* if $\{y : y \succsim x\}$ is convex; *strictly convex* if $y \succsim x$, $y \neq x$ and $\lambda \in (0,1)$ imply $\lambda y + (1-\lambda)x \succ x$. We say that the utility function $x : X \to \mathbb{R}$ *represents* $\succsim$ if for all $x, y \in X$, $x \succsim y$ iff $u(x) \geqq u(y)$. If $\succsim$ is rational and continuous then it can be represented by a continuous function; the converse holds; thus rationality and continuity imply that $\succsim$ is closed as a set of X^2. We have that $\succsim$ is (strictly) convex iff u is (strictly) quasi-concave. We will assume that $X = \mathbb{R}^L_{\geqq 0}$ and that $\succsim$ is rational, continuous and represented by a continuous u.

A Walrasian budget set is $B_{p,w} = \{x \in X : p^\top x \leqq w\}$, where $p \in \mathbb{R}^L_+$ are prices and $w \in \mathbb{R}_{\geqq 0}$ is wealth. The *utility maximization problem* (UMP) is to maximize $u(x)$ subject to $x \in B_{p,w}$. Since u is assumed to be continuous and $B_{p,w}$ is compact, there is always a solution; let $x^*(p,w) = \arg\max_{x \in B_{p,w}} u(x)$ be the Walrasian demand correspondence. If $\succsim$ is locally non-satiated then x^* satisfies Walras' law, i.e., that $p^\top x^*(p,w) = w$. If $\succsim$ is convex, $x^*(p,w)$ is a convex set; if $\succsim$ is strictly convex, $x^*(p,w)$ is a function. We define the *indirect utility function* as $v(p,w) = u(x^*(p,w))$; it is continuous in p and w by Berge's theorem; it is non-decreasing in p_i for each $i = 1, \ldots, L$; if $\succsim$ is locally non-satiated, $v(p,\cdot)$ is increasing.

The *expenditure minimization problem* (EMP) is to minimize $p^\top x$ s.t. $u(x) \geqq \tilde{u}$; the *Hicksian demand corresponde* is defined as $h(p,\tilde{u}) = \arg\min_{u(x)\geqq\tilde{u}} p^\top x$; we have $u(h(p,\tilde{u})) = \tilde{u}$ (if $u(x) > \tilde{u}$ for $x \in h(p,\tilde{u})$, $u(\alpha x) > \tilde{u}$ for $\alpha < 1$ close to 1, absurd); if $\succsim$ is convex, $h(p,\tilde{u})$ is convex; it $\succsim$ is strictly convex, $h(p,\tilde{u})$ is unique. The *expenditure function* is defined as $e(p,\tilde{u}) = p^\top h(p,\tilde{u})$; we have $h(p, v(p,w)) = x^*(p,w)$ and $h(p,\tilde{u}) = x^*(p, e(p,\tilde{u}))$; e is non-decreasing in p_i for each $i = 1, \ldots, L$; if $\succsim$ is locally non-satiated, $e(p,\cdot)$ is increasing; $e(\cdot,\tilde{u})$ is concave; e is continuous in p and $\tilde{u}$ by Berge's theorem. Compensated law of demand: if $\succsim$ is locally non-satiated and h is a function then $(p-p')^\top(h(p,u) - h(p',u)) \leqq 0$. If h is a function and $e(\cdot,u)$ is differentiable then $\nabla_p e(p,u) = h(p,u)$: $f(p) = p^\top h(p_0,u) - e(p,u) \geqq 0$ and $f(p_0) = 0$ therefore $\nabla f(p_0) = 0$ and we get the result. Slutsky's equation: we get $\partial_{p_i} h_j(p,u) = \partial_{p_i} x_j^*(p,w) + \partial_w x_j^*(p,w) x_j(p,w)$ by differentiating $h(p,u) = x^*(p, e(p,u))$. Roy's identity: $x^*(p,w) = -\frac{1}{\partial_w v(p,w)} \nabla_p v(p,w)$.

Welfare analysis. If prices change from p^0 to p^1 we define the *equivalent variation* (EV) as $EV = e(p^0,u^1) - e(p^0,u^0)$, where $u^1 = v(p^1,w)$ and $u^0 = v(p^0,w)$; it is the change in wealth that is equivalent (measured by utility) to the change in prices. We define the *compensating variation* (CV) as $CV = e(p^1,u^1) - e(p^1,u^0)$; it is the change of wealth that will leave the agent equally satisfied as before. If the change is just in price p_i, we have $EV = \int_{p_i^1}^{p_i^0} h(\cdot, p_{-i}, u^1)$ and $CV = \int_{p_i^1}^{p_i^0} h(\cdot, p_{-i}, u^0)$; we define the *consumer surplus* as $AV = \int_{p_i^1}^{p_i^0} x^*(\cdot, p_{-i}, w)$.

Von Neumann-Morgenstern utility. A *lottery* L over a finite set Z (noted $L \in \Delta Z$) is a probability distribution. A preference over lotteries $\succsim$ is *continuous* if the sets $\{\lambda \in [0,1] : \lambda L_1 + (1-\lambda)L_2 \succsim L_3\}$ and $\{\lambda \in [0,1] : \lambda L_1 + (1-\lambda)L_2 \precsim L_3\}$ are closed for all $L_1, L_2, L_3 \in \Delta Z$; it satisfies the *independence axiom* iff for all L_1, L_2, L_3 and $\lambda \in (0,1)$ then $L_1 \succsim L_2$ iff $\lambda L_1 + (1-\lambda)L_3 \succsim \lambda L_2 + (1-\lambda)L_3$. Expected utility theorem: if $\succsim$ is continuous and satisfies the independence axiom then there is $u : Z \to \mathbb{R}$ such that $U(L) = \sum_{z\in Z} p_z u(z)$ represents $\succsim$, and it is unique up to linear transformations with positive slope. *Proof.* Let $\underline{L}$ and $\overline{L}$ be the minimum and maximum lotteries; for every L there is $\lambda_L \in [0,1]$ such that $L \sim (1-\lambda_L)\underline{L} + \lambda_L \overline{L}$, clearly $\lambda_L = \sum_{z\in Z} p_z \lambda_{\delta_z}$, and $U(L) = \lambda_L$ represents $\succsim$, as claimed; if U' is other utility, $U'(L) = U'((1-\lambda_L)\underline{L} + \lambda_L\overline{L}) = (U'(\overline{L}) - U'(\underline{L}))\lambda_L + U'(\underline{L})$, as claimed.

Risk aversion. Let F be the cdf of a random variable X over $\mathbb{R}$, and $u : \mathbb{R} \to \mathbb{R}$ a utility function. We can think of F as a lottery, and define the utility $U(F) = \mathbb{E}[u(X)] = \int u(x)\,dF(x)$. We say that the agent is *risk averse* if $U(F) \leqq u(\mathbb{E}X)$ for every F; equivalently, if u is concave. The Arrow-Pratt coefficient of *absolute risk aversion* at x is $r_A(x) = -u''(x)/u'(x)$; the *relative risk aversion* is $r_R(x) = -xu''(x)/u'(x)$.

Production A *production vector* $y \in \mathbb{R}^L$ describes the net outputs from a production technology; $\{i : y_i > 0\}$ are interpreted as outputs and $\{i : y_i < 0\}$ as inputs. A *production set* $Y \subset \mathbb{R}^L$ describes the possible production vectors. When $Y = \{y : F(y) \leqq 0\}$ we call F the *transformation function*, and $\{y : F(y) = 0\}$ the transformation frontier. When only good L is produced, $Y = \{(y,q) \in \mathbb{R}^{L-1}_{\leqq 0} \times \mathbb{R} : q \leqq f(y)\}$, where f is called the *production function*.

Properties of production sets. No free lunch: if $y \in Y$ and $y \geqq 0$ then $y = 0$. Possibility of inaction: $0 \in Y$. Free disposal: $y \in Y$, $y' \leqq y$ then $y' \in Y$. Irreversibility: $y \in Y$ and $y \neq 0$ imply $-y \notin Y$. Nonincreasing returns to scale: if $y \in Y$ and $\alpha \in [0,1]$ then $\alpha Y \in Y$. Nondecreasing returns to scale: if $y \in Y$ and $\alpha \geqq 1$ then $\alpha y \in Y$. Constant returns to scale (or, Y is a cone): $y \in Y$ and $\alpha \geqq 0$ implies $\alpha y \in Y$. Additivity (or free entry): $y, y' \in Y$ then $y + y' \in Y$. Convexity: $y, y' \in Y$, $\alpha \in [0,1]$ then $\alpha y + (1-\alpha)y' \in Y$.

Profit maximization problem (PMP). *Profit function* $\pi(p) = \max\limits_{y\in Y} p^\top y$; *profit correspondence* $y^*(p) = \arg\max\limits_{y\in Y} p^\top y$. If Y is closed and has free disposal then π is convex; if Y is convex then $y^*(p)$ is convex; (Hotelling's lemma) if $y^*(p)$ consists of a single point, π is differentiable at p and $\nabla \pi(p) = y^*(p)$.

Cost minimization problem (CMP). Let $z \geqq 0$ be a vector of inputs, $f(z)$ the production function, q the amount of output, and $w > 0$ the vector of input prices. The CMP is to minimize $w^\top z$ s.t. $f(z) \geqq q$; $c(w,q) =_{\text{def}} \min\limits_{f(z)\geqq q} w^\top z$ is the *cost function* and $z(w,q) =_{\text{def}} \arg\max\limits_{f(z)\geqq q} w^\top z$ is the *conditional factor demand correspondence*. If Y is closed and has free disposal then $c(\cdot, q)$ is concave; if $|z(w,q)| = 1$ then $\nabla_w z(w,q) = z(w,q)$; if f is concave, $c(w,\cdot)$ is convex.

Equilibrium Let $\mathcal{I} = \{1, \ldots, I\}$ be a set of agents with utility functions u_i over bundles of L goods, endowments $\omega_i \in \mathbb{R}^L_+$, and property $\theta_{ij} \in [0,1]$ over firms $j \in \mathcal{J} = \{1, \ldots, J\}$, with production sets Y_j, such that for all j, $\sum_{i \in \mathcal{I}} \theta_{ij} = 1$. A market clearing allocation is $(x, y) \in \mathbb{R}^{I \times L}_+ \times \mathbb{R}^{J \times L}$ such that $y_j \in Y_j$ and $\sum_i x_i = \sum_i \omega_i + \sum_j y_j$. Let $C \subset \mathcal{I}$. We say the C blocks the allocation if there is an allocation (x', y') restricted to C and the firms completely owned by those agents such that $u_i(x'_i) \geqq u_i(x_i)$ for all $i \in C$ and the inequality is strict for some $i \in C$. The *core* is the set of allocations that can not be blocked by any subset of agents. An allocation is *Pareto optimal* if it can not be blocked by $\mathcal{I}$. A *Walrasian equilibrium* is an allocation plus a vector of prices $p \in \mathbb{R}^L_{>0}$ such that x_i maximizes $u_i(x)$ s.t. $p \cdot x \leqq p \cdot (\omega_i + \sum_j \theta_{ij} y_j)$ and y_j maximizes $p \cdot y_j$ for all i, j. A *price equilibrium with transfers* is $(x, y, p) \in \mathbb{R}^{I \times L}_+ \times \mathbb{R}^{J \times L} \times \mathbb{R}^L_+$ such that there exists wealth levels $w \in \mathbb{R}^I_+$ such that $\sum_i w_i = p \cdot \sum_i \omega_i + \sum_j p \cdot y_j$, $y_j \in Y_j$ maximizes $p \cdot y_j$, x_i maximizes $u_i(x)$ s.t. $p \cdot x \leqq w_i$, and $\sum_i x_i = \sum_i \omega_i + \sum_j y_j$.

First welfare theorem. If preferences are locally non-satiated then a Walrasian equilibrium is in the core and a price equilibrium with transfers is Pareto optimal. *Second welfare theorem.* If preferences are convex and locally non-satiated, and Y_j are convex, then for every Pareto optimal allocation (x, y) there is a price vector p such that y_j maximizes $p \cdot y_j$, and x_i minimizes $p \cdot x$ s.t. $u_i(x) \geqq u_i(x_i)$. *Proof.* Let $V_i = \{x'_i \mid x'_i \succ x_i\}$, $V = \sum_i V_i$, $Y = \sum_j Y_j$ and $\bar{\omega} = \sum_i \omega_i$; we have $V \cap (Y + \bar{\omega}) = \varnothing$ because (x, y) is Pareto optimal; therefore there is $p \in \mathbb{R}^L$ and $r \in \mathbb{R}$ s.t. $p \cdot x' \geqq r \geqq p \cdot (y' + \bar{\omega})$ for every $x' \in V$, $y' \in Y$; that is still valid taking $V_i = \{x'_i \mid x'_i \succsim_i x_i\}$, so $p \cdot x = r$, and we get the result. If preferences are continuous and for each i there is x'_i s.t. $p \cdot x'_i < p \cdot x_i$ then x_i solves the UMP with wealth $p \cdot x_i$. (If $p \cdot x''_i \leqq p \cdot x_i$ then $x'''_i = \alpha x''_i + (1 - \alpha) x'_i$ satisfies $x'''_i \prec_i x_i$ by the theorem and thus by continuity by taking $\alpha \to 1$ we get $x''_i \precsim_i x_i$.) Therefore if preferences are also monotone and $x_i > 0$, we get that (x, y, p) is a price equilibrium with transfers.

Existence of a Walrasian equilibrium. If preferences are continuous, strictly convex and strongly monotone, and Y_j are closed, strictly convex and bounded above, $0 \in Y_j$ and $\sum_i \omega_i \gg 0$, then there is a Walrasian equilibrium. In that case the *excess demand function* $z(p) = \sum_i x_i(p, p \cdot \omega_i + \sum_j \theta_{ij} \pi_j(p)) - \sum_i \omega_i - \sum_j y_j(p)$ defined for $p \gg 0$ satisfies the following conditions: it is continuous, homogeneous of degree zero, $p \cdot z(p) = 0$, it is bounded below, and if $p_n \to p$, $p \neq 0$ but $p_l = 0$ for some l then $\max\{z_1(p_n), \ldots, z_L(p_n)\} \to \infty$. And we have that under those conditions there is $p \gg 0$ such that $z(p) = 0$.

3.2 Contracts

There is a principal P and an agent A. The principal chooses a contract $w : \Phi \to \mathbb{R}$, where Φ is the set of observed outcomes. The agent observes a cost parameter $\theta \in \Theta$ with distribution G, and decides whether or not to accept the principal's offer, i.e., chooses $d \in \{0, 1\}$. If he accepts, then he chooses an action $e \in \mathcal{E} \subset \mathbb{R}_+$ incurring a cost $c(e, \theta)$ s.t. $c_e > 0$, $c_{ee} \geqq 0$ and $c_\theta < 0$. The agent's action generates a stochastic benefit $y \in \mathcal{Y} \subset \mathbb{R}$ with distribution $F(\cdot|e)$. Finally, the principal pays $w(\phi)$ to the agent; the observed outcome ϕ can be (e, y) under hidden information (*adverse selection*) or (θ, y) under hidden action (*moral hazard*). The principal is risk-neutral and earns $y - w$ if $d = 1$ and 0 otherwise; the agent's payoff is $u(w) - c$, where $u' > 0$, if $d = 1$, and u_0 otherwise.

Moral hazard Suppose that $\Theta = \{\theta\}$ and $\phi = y$, and let $c(e) := c(e,\theta)$. The optimal contract consists in choosing a function $w : \mathcal{Y} \to \mathbb{R}$ and an action $e \in \mathcal{E}$ to

$$\begin{aligned} \text{maximize} \quad & \mathbb{E}[y - w(y)|e] \\ \text{subject to} \quad & \mathbb{E}[u(w(y))|e] - c(e) \geqq \mathbb{E}[u(w(y))|e'] - c(e'), \text{ for all } e', && \text{(IC)} \\ & \mathbb{E}[u(w(y))|e] - c(e) \geqq 0. && \text{(IR)} \end{aligned}$$

The constraints are called the *incentive compatibility* (IC) and the *individual rationality* (IR) constraint.

If e is contractible, the principal can choose e so only (IR) applies, and it is clearly binding. If $u'' < 0$, the optimal $w(y)$ is constant, so the principal fully insures the agent and is the residual claimant. If $u(x) = x$, any $w(y)$ such that $\mathbb{E}[w(y)|e] = c(e)$ is optimal, and the optimal e maximizes $\mathbb{E}(y|e) - c(e)$.

Assume that e is private information. If $u(x) = x$, the principal can offer $w(y) = y - w_0$. The agent chooses e to maximize $\mathbb{E}(y|e) - c(e)$, so chooses the same e as if e was contractible. Hence if the agent is risk neutral, the principal makes him the residual claimant.

Now assume that $u'' < 0$. Assume that $\mathcal{E} = \{e_L, e_H\}$, with $e_L < e_H$ and $c' > 0$. Assume that $F(\cdot|e)$ has a density $f(\cdot|e)$. By taking $u(w(y))$ as the control variable we see that the problem has a concave objective with linear constraints, hence the Kuhn-Tucker conditions are necessary and sufficient. If $e = e_L$, clearly a constant w where (IR) binds is optimal, as if e was observable. If $e = e_H$, the first-order condition is

$$\frac{1}{u'(w(y))} = \lambda + \mu \left[1 - \frac{f(y|e_L)}{f(y|e_H)}\right].$$

We have $\lambda, \mu > 0$, since $f(y|e_H) < f(y|e_L)$ for some y, and if $w(y)$ is constant, (IC) is violated. Hence we see that the principal gives the agent incentives (i.e., $w(y)$ is not constant) and has to compensate him with rents ($\mathbb{E}[w(y)|e]$ is greater than if e was not observable). Therefore effort may be underprovided.[3] Note that $w(y)$ is weakly increasing iff $\frac{f(y|e_L)}{f(y|e_H)}$ is weakly decreasing (this is called the *monotone likelihood ratio property*). Finally, note that if z was a payoff-irrelevant observable variable, the first-order condition would have $\frac{f(y,z|e_L)}{f(y,z|e_H)}$ instead, so if z adds information above y (even if the agent cannot influence z directly), the principal will use it to provide incentives.

Linear contracts. Assume that $y = e + \epsilon$ with $\epsilon \sim N(0, \sigma^2)$, the agent's utility is $u(w, e) = -\exp\{-r(w - c(e)\}$ if he accepts and $-\exp\{-ru_0\}$ otherwise, and contracts are restricted to the form $w(y) = by + w_0$. We have $\mathbb{E}u = -\exp\{-r\left[be - \frac{1}{2}rb^2\sigma^2 + w_0 - c(e)\right]\}$, so IC is $b = c'(e)$ and IR is $be - \frac{1}{2}rb^2\sigma^2 + w_0 - c(e) \geqq u_0$. The principal thus maximizes $(1 - b)e - w_0$. In the case $c(e) = \frac{1}{2}ke^2$, we get $b = \frac{1}{1+r\sigma^2 k}$. The higher the uncertainty and risk-aversion of the agent, the lower the power b of incentives, and the lower the effort provided.

Multi-task.

Multiple principals.

Moral hazard in teams.

Screening Suppose that $\phi = e$, and y is a deterministic function of e s.t. $y_e > 0$ and $y_{ee} < 0$. By the revelation principle, the optimal contract consists in a function $w : \Theta \to \mathbb{R}$ and a

[3] With more than two effort levels, effort may be overprovided if e is not observable.

function $e : \Theta \to \mathcal{E}$ that, for all $\theta \in \Theta$,

$$\begin{aligned} &\text{maximize} && \mathbb{E}[y(e(\theta)) - w(\theta)] \\ &\text{subject to} && w(\theta) - c(e(\theta),\theta) \geqq w(\theta') - c(e(\theta'),\theta), \text{ for all } \theta', && \text{(IC)} \\ & && w(\theta) - c(e(\theta),\theta) \geqq 0. && \text{(IR)} \end{aligned}$$

The expectation in the maximand is over the principal's prior over θ. If θ were contractible, (IR) binds for all θ and $e(\theta)$ maximizes $y(e) - c(e,\theta)$.

Suppose that there are finite types $\theta_1 < \cdots < \theta_n$ with probabilities $g(\theta_i)$, and assume the *single-crossing* condition $c_{e\theta} < 0$. Note that given (IC) we can remove all (IR) except that for $\theta = \theta_1$. Summing (IC) we get $c(e(\theta),\theta) - c(e(\theta),\theta') \leqq c(e(\theta'),\theta) - c(e(\theta'),\theta')$, which implies $e(\theta) \geqq e(\theta')$ if $\theta \geqq \theta'$ since $c_{e\theta} < 0$. Using this, we can remove all (IC) except those that are local, i.e., those with $\{\theta,\theta'\} = \{\theta_i,\theta_{i+1}\}$. Now, consider the downward (IC), i.e., those with $\theta > \theta'$, plus the monotonicity constraints $e(\theta_{i+1}) \geqq e(\theta_i)$. If we solve the problem with these constraints plus (IR) for θ_1, in the solution all the (IC) bind, since if $w(\theta_{i+1}) - c(e(\theta_{i+1}),\theta_{i+1}) > w(\theta_i) - c(e(\theta_i),\theta_{i+1})$, we can substract $\epsilon > 0$ from all $w(\theta_j)$ for $j \geqq i+1$, improving the objective. The binding downward (IC) plus the monotonicity constraints imply the upward (IC) constraints at the optimum, so the solution must be the same. Replacing the equations for $w(\theta_i)$ in the objective, the first-order condition for $e(\theta_i)$ is

$$y'(e(\theta_i)) - c_e(e(\theta_i),\theta_i) = \frac{1 - G(\theta_i)}{g(\theta_i)}(c_e(e(\theta_i),\theta_i) - c_e(e(\theta_i),\theta_{i+1})).$$

If $n > 2$ we need to check that e is monotone for it to be the solution. In that case, we see that $y'(e(\theta_i)) - c_e(e(\theta_i),\theta_i) > 0$ for all i except $i = n$, for which we have $y'(e(\theta_i)) - c_e(e(\theta_i),\theta_i) = 0$. Thus, effort is underprovided for all types except the most efficient one, i.e., we have *no distortion at the top*. All types except the one at the bottom receive an informational rent $w(\theta) - c(e(\theta),\theta) > 0$, which increases with the type.

Suppose that $\Theta = [\underline{\theta},\overline{\theta}]$ and G has density g. Suppose that the solution $e(\theta)$ is differentiable. Then (IC) implies $w'(\theta) = c_e(e(\theta),\theta)e'(\theta)$. By applying integration by parts multiple times, the objective can be expressed as

$$\int_{\underline{\theta}}^{\overline{\theta}} \{[y(e(\theta)) - c(e(\theta),\theta)]\, g(\theta) + c_\theta(e(\theta),\theta)(1 - G(\theta))\}\; d\theta.$$

Assume that $c_{\theta ee} \leqq 0$. Then the optimal $e(\theta)$ satisfies

$$(y'(e(\theta)) - c_e(e(\theta),\theta))g(\theta) = -(1 - G(\theta))c_{e\theta}(e(\theta),\theta).$$

If $h(\theta) = \frac{g(\theta)}{1-G(\theta)}$, the *hazard rate*, is increasing and $c_{e\theta\theta} \geqq 0$ then $e(\theta)$ is monotone and it solves the problem. We see that $y'(e(\theta)) - c_e(e(\theta),\theta) > 0$ for $\theta < \overline{\theta}$, so effort is underprovided for all except the most efficient type. The rent $w(\theta) - c(e(\theta),\theta)$ increases with the type.

3.3 Growth

Modelo de Solow-Swan En una economía hay un stock de capital físico K (máquinas, edificios, etc), un stock de trabajo L, y un nivel de tecnología T. Capital y trabajo son *rival goods*, i.e., no se pueden usar al mismo tiempo por dos productores distintos; la tecnología sí. Tenemos una función de producción $Y = F(K,L,T)$ que produce un flujo de bienes. No hay comercio, así que la producción se usa para consumo C o ahorro/inversión S: $Y = C + S$. Notamos $s = S/Y$. Se asume que F cumple: *constant returns to scale*, i.e., $F(\lambda K,\lambda L,T) =$

$\lambda F(K,L,T)$, $\partial_K F, \partial_L F > 0$, $\partial_{KK} F, \partial_{LL} F < 0$, $\partial_K F(0^+) = \partial_L F(0^+) = +\infty$, $\lim_{K\to+\infty} \partial_K F = \lim_{L\to+\infty} \partial_L F = 0$. Sean k, l, y, c las variables per cápita; bajo pleno empleo son $K/L, 1, Y/L, C/L$, y $f = F/L$. Tenemos $y = f(k,1,T) = f(k)$. Asumimos que el crecimiento de la población $\dot{L}/L = n$ es exógeno y constante. Un ejemplo es la función de Cobb-Douglas: $F = A(T)K^\alpha L^{1-\alpha}$ con $0 < \alpha < 1$, que da $f(k) = Ak^\alpha$.

La empresa representativa tiene capital K; puede ponerlo a producir, y obtener un flujo $F(K,L,T) - wL - \delta K$, donde w es el costo del trabajo y $\delta \geqq 0$ es la tasa de depreciación del capital (constante), o alquilarlo a otra empresa y obtener rK. En el equilibrio es indiferente, por lo que $F(K,L,T) - wL - \delta K = rK$. Entonces $r = f'(k) - \delta$ y $w = f(k) - f'(k)k$. Tenemos $\dot{K} = Y - C - \delta K$, por lo que $\dot{k} = rk + wl - c - nk = f(k) - c - (n+\delta)k$. En el modelo de Solow-Swan s es constante, por lo que $c = (1-s)f(k)$ y $\dot{k} = sf(k) - (n+\delta)k$.

Tenemos $\gamma_k = \frac{\dot{k}}{k} = sf(k)/k - (n+\delta)$. Ahora $\partial_k \frac{f(k)}{k} = -(f(k) - f'(k)k)/k^2 = -w/k^2 < 0$. Sea k^* tal que $\dot{k} = 0$, es decir, $f(k^*) = s^{-1}(n+\delta)k^*$; existe y es único. Si $k < k^*$, $\gamma_k > 0$ y viceversa; entonces $k \to k^*$. En ese punto ninguna variable per cápita crece. El crecimiento debe darse, pues, por factores exógenos. Si s crece, k^* crece, pero tiene un límite. El crecimiento lo produce, pues, el cambio tecnológico. ¿Cómo elegir s? Bajo $k = k^*$ el consumo es constante; elegimos s, pues, para maximizarlo hoy: $c = (1-s)f(k^*) = f(k^*) - (n+\delta)k^*$ da $\partial_s c = (f'(k^*) - (n+\delta))\partial_s k^*$, por lo que c se maximiza con $f'(k^*) = n + \delta$. Esa es la "regla de oro".

Crecimiento endógeno Tenemos una producción per cápita $y = f(k)$ y un consumo $0 < c < y$; el individuo representativo busca maximizar

$$U = \int_0^\infty u(c)e^{-\rho t}\, dt,$$

donde $u(c) = \frac{c^{1-\sigma}-1}{1-\sigma}$, con $\sigma > 0$. Esto da $\frac{\dot{c}}{c} = \frac{1}{\sigma}(f'(k) - \rho)$. Asumimos $y = Ak$, con $A > 0$ (la idea es que k incluye capital humano), y la ecuación es $\gamma = \frac{\dot{c}}{c} = \frac{1}{\sigma}(A - \rho)$. Asumimos que $A > \rho$ de manera que $\gamma_c > 0$ (hay crecimiento endógeno, i.e., sin cambio tecnológico), pero $(1-\sigma)A < \rho$, de manera que $U < \infty$ (se ve derivando $u(c)e^{-\rho t}$). Resulta que $\frac{\dot{k}}{k} = \frac{\dot{y}}{y} = \frac{\dot{c}}{c} = \gamma$.

Gasto público La producción es $y = \Phi(k,g) = k\phi\left(\frac{g}{k}\right)$, donde g es el gasto público per cápita y $\phi' > 0$, $\phi'' < 0$. La idea es que el gasto del gobierno hace más productivo el capital, por externalidades positivas o porque cubre un costo fijo (si los privados proveyeran g habría undersupply). Tomamos $\phi\left(\frac{g}{k}\right) = A\left(\frac{g}{k}\right)^\alpha$ con $0 < \alpha < 1$. El gasto público se financia con un impuesto al ingreso: $g = \tau y$. Tenemos que $\dot{k} = (1-\tau)y - c$, por lo que

$$\gamma = \frac{\dot{c}}{c} = \frac{1}{\sigma}\left[(1-\tau)\frac{\partial y}{\partial k} - \rho\right] = \frac{1}{\sigma}\left[(1-\tau)(1-\alpha)\phi\left(\frac{g}{k}\right) - \rho\right].$$

Si τ es constante, tenemos $\frac{\dot{k}}{k} = \frac{\dot{y}}{y} = \frac{\dot{c}}{c} = \frac{\dot{g}}{g} = \gamma$, $c_0 = [(1-\tau)(\tau^\alpha A)^{\frac{1}{1-\alpha}} - \gamma]k_0 = \frac{\rho+\gamma(\sigma+\alpha-1)}{1-\alpha}k_0$, $c = e^{\gamma t}c_0$ y

$$U = \frac{c_0^{1-\sigma}}{(1-\sigma)[\rho - \gamma(1-\sigma)]} = \frac{[\rho + \gamma(\sigma+\alpha-1)]^{1-\sigma}k_0^{1-\sigma}}{(1-\alpha)^{1-\sigma}(1-\sigma)[\rho-\gamma(1-\sigma)]}.$$

Se ve que $\partial_\gamma U > 0$, que elegir τ para maximizar U es maximizar γ, y que γ se maximiza con $\tau = \alpha$.

Ahora, si el gobierno decidiera la inversión, $y = (\tau^\alpha A)^{\frac{1}{1-\alpha}}k$, por lo que

$$\gamma = \frac{\dot{c}}{c} = \frac{1}{\sigma}\left[(1-\tau)\frac{\partial y}{\partial k} - \rho\right] = \frac{1}{\sigma}\left[(1-\tau)(\tau^\alpha A)^{\frac{1}{1-\alpha}} - \rho\right].$$

4 Political economy

4.1 Collective preference

Let $N = \{1, \dots, n\}$ be a set of individuals and X a set of alternatives. Let $\mathcal{B}$ be the set of complete reflexive relations on X, and $\mathcal{R}$ be the set of weak orders on X; given $\succsim \in \mathcal{B}$ we define $\succ \in \mathcal{B}$ by $x \succ y$ iff $x \succsim y$ and not $y \succsim x$. We say that $\succsim \in \mathcal{B}$ is *transitive* if $x \succsim y$ and $y \succsim z$ imply $x \succsim z$; *quasi-transitive* if $x \succ y$ and $y \succ z$ imply $x \succ z$; *acyclic* if there are no cycles $x_1 \succ \cdots \succ x_n \succ x_1$. We assume that each individual $i \in N$ has a preference $\succsim_i \in \mathcal{R}$ and let $\rho = (\succsim_i)_{i \in N}$. A *preference aggregation rule* is a map $f : \mathcal{R}^n \to \mathcal{B}$. A set $L \subset N$ is *decisive* iff for all $x, y \in X$, $\forall i \in L.\, x \succ_i y$ implies $x \succ_{f(\rho)} y$; we call $\mathcal{L}(f)$ the set of decisive sets. Note that $\mathcal{L}(f)$ is monotonic (if $L \subset L' \subset N$ and $L \in \mathcal{L}(f)$ then $L' \in \mathcal{L}(f)$) and proper (if $L \in \mathcal{L}(f)$ then $N \smallsetminus L \notin \mathcal{L}(f)$). We say that f is *Paretian* if N is decisive, and satisfies *independence of irrelevant alternatives* (IIA) if for all $x, y \in X$, $\rho, \rho' \in \mathcal{R}^n$ such that $\rho|_{\{x,y\}} = \rho'|_{\{x,y\}}$, we have $x \succsim_{f(\rho)} y$ iff $x \succsim_{f(\rho')} y$. We say that $i \in N$ is a *dictator* if $\{i\}$ is decisive; i has a *veto* if for all $x, y \in X$, $x \succ_i y$ implies that not $y \succ_{f(\rho)} x$. We say that f is a *dictatorship* if there is a dictator, and *oligarchic* if there is $L \subset \mathcal{L}(f)$ such that every member has a veto. Note that if i has a veto then if everybody but i prefers y to x, she can oppose, and we can say that f does not satisfy minimal democracy.

Theorem 4.1.1 (Arrow)**.** If $|X| \geqq 3$, f is Paretian and IIA, then (1) if $f(\rho)$ is transitive for all $\rho \in \mathcal{R}^n$ then f is a dictatorship, and (2) if $f(\rho)$ is quasi-transitive for all $\rho \in \mathcal{R}^n$ then f is an oligarchy.

Proof. Assume that for any ρ, $f(\rho)$ is quasi-transitive. Let $L \subset N$. If $\exists \rho, x, y$ s.t. $x \succ_i y$ for $i \in L$, $y \succ_i x$ for $i \notin L$ and $x \succ_{f(\rho)} y$ then for any ρ, x, y s.t. $x \succ_i y$ for $i \in L$ and $y \succ_i x$ for $i \notin L$ we have $x \succ_{f(\rho)} y$. (First, we prove that for any $z \notin \{x, y\}$ there is ρ s.t. $x \succ_i z$ for $i \in L$, $z \succ_i x$ for $i \notin L$ and $x \succ_{f(\rho)} z$: take $x \succ_i y \succ_i z$ for $i \in L$ and $y \succ_i z \succ_i x$ for $i \notin L$; by IIA we have $x \succ_{f(\rho)} y$, by Pareto we have $y \succ_{f(\rho)} z$; by quasi-transitivity we have $x \succ_{f(\rho)} z$. Using the same method we get the result.) Moreover, L is decisive. (Using the same method.) We can also prove by the same method that if $\exists \rho, x, y$ s.t. $x \succ_i y$ for $i \in L$, $y \succ_i x$ for $i \notin L$ and $x \succsim_{f(\rho)} y$ then for any ρ, x, y s.t. $x \succ_i y$ for all $i \in L$ we have $x \succsim_{f(\rho)} y$.

Take $L \in \mathcal{L}(f)$ with minimum size. If $|L| = 1$ then f is dictatorial, and in particular oligarchic, so we are done. Suppose that $|L| > 1$. Take $i \in L$, $x, y, z \in X$ distinct, and ρ s.t. $x \succ_i y \succ_i z$, $z \succ_j x \succ_j y$ for $j \in L \smallsetminus \{i\}$ and $y \succ_j z \succ_j x$ for $j \notin L$. Since $L \in \mathcal{L}(f)$ we have $x \succ_{f(\rho)} y$. If $z \succ_{f(\rho)} y$, by what we proved we get that $L \smallsetminus \{i\}$ is decisive, absurd; if $x \succ_{f(\rho)} z$, $\{i\}$ is decisive, absurd. Therefore $x \succ_{f(\rho)} y \succsim_{f(\rho)} z \succsim_{f(\rho)} x$. If $f(\rho)$ is transitive, $y \succsim_{f(\rho)} x$, a contradiction, so we proved (1). By quasi-transitivity, $x \succsim_{f(\rho)} z$, so $x \succ_i z$, $z \succ_j x$ for all $j \neq i$ and $x \succsim_{f(\rho)} z$, therefore i has a veto by what we proved. Hence L is decisive and every member has a veto, which proves (2). ∎

We say that f is *collegial* if $K(f) = \bigcap_{L \in \mathcal{L}(f)} L$ is nonempty, and we call $K(f)$ the *collegium.*

Theorem 4.1.2. If $|X| \geqq n$, f is Paretian and $f(\rho)$ is acyclic for all $\rho \in \mathcal{R}^n$ then f is collegial.

Proof. If f is not collegial, for every $i \in N$ there is $L \in \mathcal{L}(f)$ s.t. $i \notin L$. Therefore $N \smallsetminus \{i\} \in \mathcal{L}(f)$ by monotonicity. Since $|X| \geqq n$, take $x_1, \dots, x_n \in X$ distinct and ρ s.t. $x_1 \succ_1 \cdots \succ_1 x_n$, $x_2 \succ_2 \cdots \succ_2 x_n \succ_2 x_1$, …, $x_n \succ_n x_1 \succ_n \cdots \succ_n x_{n-1}$. We have $x_1 \succ_{f(\rho)} x_2 \succ_{f(\rho)} \cdots \succ_{f(\rho)} x_n \succ_{f(\rho)} x_1$, so $f(\rho)$ is not acyclic. ∎

Given $\mathcal{L} \subset N$ proper, we can define $f_{\mathcal{L}}$ by $x \succ_{f_{\mathcal{L}}(\rho)} y$ iff $\exists L \in \mathcal{L}.\, \forall i \in L.\, x \succ_i y$. We say that f is a *simple rule* if $f = f_{\mathcal{L}(f)}$. *Plurality rule* is defined by $x \succ_{f_p(\rho)} y$ iff $|P_\rho(x, y)| > |P_\rho(y, x)|$, where $P_\rho(x, y) = \{i \in N : x \succ_i y\}$. *Majority rule* is defined by $x \succ_{f_m(\rho)} y$ iff $|P_\rho(x, y)| > n/2$.

We see that $\mathcal{L}(f_p) = \mathcal{L}(f_m) = \{L \subset N : |L| > n/2\}$, but $f_{\mathcal{L}(f_m)} = f_m$ and $f_m \neq f_p$, so majority rule is simple but plurality is not. As a partial converse to the previous Theorem, if f is simple and collegial then $f(\rho)$ is acyclic for all $\rho \in \mathcal{R}^n$, and in particular the *core* $\mathcal{C}_f(\rho) = \{x \in X : x \succsim_{f(\rho)} y \text{ for all } y \in X\}$ is not empty assuming that X is finite; in particular, if there is a veto player with strict preferences, the core is not empty. We say that f is *decisive* if given $\rho, \rho' \in \mathcal{R}^n$, $x, y \in X$, $P_\rho(x,y) = P_{\rho'}(x,y)$ and $x \succ_{f(\rho)} y$ then $x \succ_{f(\rho')} y$; *monotonic* if $P_\rho(x,y) \subset P_{\rho'}(x,y)$, $P_\rho(y,x) \supset P_{\rho'}(y,x)$ and $x \succ_{f(\rho)} y$ then $x \succ_{f(\rho')} y$; *neutral* if $P_\rho(x,y) = P_{\rho'}(x',y')$ and $P_\rho(y,x) = P_{\rho'}(y',x')$ imply that $x \succsim_{f(\rho)} y$ iff $x' \succsim_{f(\rho')} y'$. The following follows from the definitions.

Proposition 4.1.3. A rule f is simple iff it is decisive, monotonic and neutral.

The *Nakamura number* of a simple rule f is $s(f) = \infty$ if f is collegial and

$$s(f) = \min \left\{ |\mathcal{L}| : \mathcal{L} \subset \mathcal{L}(f), \bigcap_{L \in \mathcal{L}} L = \varnothing \right\}$$

otherwise. We have $s(f) \geqq 3$ since $\mathcal{L}(f)$ is proper, and $s(f) \leqq n$ if f is not collegial (taking $\mathcal{L} = \{N \setminus \{i\} : i \in N\}$). A simple rule f is *strong* if $L \notin \mathcal{L}(f)$ implies $N \setminus L \in \mathcal{L}(f)$. A q-rule is given by $\mathcal{L}(f) = \{L \subset N : |L| \geqq q\}$; a *weighted* q-rule is given by $\mathcal{L}(f) = \{L \subset N : \sum_{i \in L} w_i \geqq q\}$ for some $w \in \mathbb{R}^n$. A rule f is *anonymous* if for all permutations σ of N and $\rho \in \mathcal{R}^n$ we have $f(\rho) = f(\rho_\sigma)$. We have that a rule is simple and anonymous iff it is a q-rule. If f is simple, strong and non-collegial then $s(f) = 3$ (take A, B minimal collegial and distinct, and $C = N \setminus (A \cap B)$). If f is a q-rule with $q < n$ then $s(f) = \lceil \frac{n}{n-q} \rceil$ (to prove $s(f) \geqq \frac{n}{n-q}$, count twice the number of edges on the graph given by $\in$ on $V = N \cup \mathcal{L}$; for an example take $L_k = N \setminus \{(k-1)(n-q)+1, \ldots, k(n-q)\}$ for $k = 1, \ldots, \lceil \frac{n}{n-q} \rceil$). The following is immediate.

Proposition 4.1.4. A simple rule f is acyclic if and only if $|X| < s(f)$.

Let $\rho \in \mathcal{R}^n$. We say that ρ is *single-peaked* if there is a total order $\succ$ on X and for each $i \in N$ there is $x_i \in X$ such that for all $x \prec y \precsim x_i$ we have $x \prec_i y$ and for all $x_i \precsim x \prec y$ we have $x \succ_i y$; x_i is called the *ideal point* of i. We say that $x \in X$ is an f-median for ρ single-peaked if $L^+(x) \equiv \{i \in N : x_i \succ x\}$ and $L^-(x) \equiv \{i \in N : x_i \prec x\}$ are not decisive in f. The following is immediate.

Theorem 4.1.5. If f is simple and ρ is single-peaked then $\mathcal{C}_f(\rho)$ is the set of f-medians of ρ.

We say that $\rho \in \mathcal{R}^n$ is *order-restricted* is there is a permutation $\sigma : N \to N$ such that for every distinct $x, y \in X$ we have $P_\rho(x,y) = \sigma([1,i))$ or $P_\rho(x,y) = \sigma((i,n])$ for some $i \in [0, n+1]$. If f is simple then it is easy to verify that $f(\rho)$ is quasi-transitive and therefore $\mathcal{C}_f(\rho)$ is nonempty. Let $\ell = \min\{i : \sigma((i,n]) \notin \mathcal{L}(f)\}$, $r = \max\{i : \sigma([1,i)) \notin \mathcal{L}(f)\}$ and $M = \sigma([\ell, r])$. Every individual in M has a veto. The *Pareto set* $\mathcal{P}_L(\rho)$ of a coalition $L \subset N$ at $\rho \in \mathcal{R}^n$ is the set of $x \in X$ s.t. there is no $y \in X$ s.t. $y \succsim_i x$ for all $i \in L$ and $y \succ_i x$ for some $i \in L$. If X is finite, $\mathcal{P}_L(\rho)$ is nonempty. If f is simple and all individuals have strict preferences over X then if ρ is order-restricted then $\mathcal{C}_f(\rho) = \mathcal{P}_M(\rho)$. In particular, for the majority rule, $\mathcal{C}_f(\rho)$ is the preferred alternative by the median individual in the order given by σ.

The spatial model Let $X \subset \mathbb{R}^k$. We say that $\succsim \in \mathcal{R}(X)$ is *lower continuous* if for all $x \in X$ the set $\{y : y \succsim x\}$ is closed, *strictly convex* if $x \succsim y$ implies $\lambda x + (1-\lambda) y \succ y$ for any $\lambda \in (0,1)$, and *semi-convex* if $x \notin \mathrm{co}(\{y : y \succ x\})$. Let $m(\succsim)$ be the set of maximal elements in X.

Lemma 4.1.6. If X is compact and convex, $\succsim$ is lower continuous and semi-convex, then $m(\succsim) \neq \varnothing$.

Proof. First, we prove that $\succsim$ satisfies condition F: for all finite subset $Y \subset X$ there is $x \in X$ s.t. $x \succsim y$ for all $y \in Y$. Let $S(y) = \{x : x \succsim y\}$; by lower continuity it is closed; we want to show that $\bigcap_{y \in Y} S(y) \neq \varnothing$. We use the Knaster-Kuratowski-Mazurkiewicz Lemma: we just need that $\mathrm{co}(L) \subset \bigcup_{y \in L} S(y)$ for every $L \subset Y$, which clearly follows from semi-convexity. Now, we want to show that $\bigcap_{x \in X} S(x) \neq \varnothing$; by compactness, this follows from condition F. ∎

Theorem 4.1.7. Let $X \subset \mathbb{R}^k$ be compact and convex, f be a simple rule and $\rho \in \mathcal{R}^n$ such that $\succsim_i$ are lower continuous and semi-convex. If $k \leqq s(f) - 2$ then $\mathcal{C}_f(\rho) \neq \varnothing$.

Proof. By the lemma, we have to show that $f(\rho)$ is lower-continuous and semi-convex. The first is because $\{y : y \prec x\} = \bigcup_{L \in \mathcal{L}(f)} \bigcap_{i \in L} \{y : y \prec_i x\}$ is open. If $f(\rho)$ is not semi-convex, there is $x \in \mathrm{co}(\{y : y \succ x\})$, so by Caratheodoy there are $y_1, \ldots, y_{k+1} \succ x$ with $x \in \mathrm{co}(\{y_1, \ldots, y_{k+1}\})$. Since $k+1 < s(f)$, there is $i \in N$ with $y_1, \ldots, y_{k+1} \succ_i x$, which implies $x \in \mathrm{co}(\{y : y \succ_i x\})$, absurd. ∎

The bound is best possible: take f simple, $k = s(f) - 1$, $X = \Delta^k$, $L_1, \ldots, L_{k+1} \in \mathcal{L}(f)$ decisive with $\bigcap_{j=1}^{k+1} L_j = \varnothing$. Let $e_1, \ldots, e_{k+1}$ be the vertices of Δ^k, $F_i = \mathrm{co}(\{e_j : j \neq i\})$ the faces, and assign to each $i \in \bigcup_{j=1}^{n+1} L_j$ an ideal point $x_i \in \bigcap_{j : i \in L_j} F_j$ (i.e., $x_{ij} = 0$ if $i \in L_j$). Let $\succsim_i$ be given by $u_i(x) = -\|x - x_i\|^2$. Now, for every $x \in \Delta^k$, let j be s.t. $x_j \neq 0$, and y its projection on F_j; the members of L_j prefer y to x (by Pythagoras), so $y \succ_{f(\rho)} x$. Hence $\mathcal{C}_f(\rho) = \varnothing$, as desired.

Let $x \in X$ and γ_x a line $\{tx + (1-t)y : t \in \mathbb{R}\} \cap X$. The induced ideal point of $i \in N$ on γ_x is $b_i(\gamma_x) \in \gamma_x$ s.t. $b_i(\gamma_x) \succsim_i y$ for all $y \in \gamma_x$, which is well defined if $\succsim_i$ is strictly convex. Let $\mu_f(\rho|_{\gamma_x})$ be the set of f-medians of ρ on γ_x, defined as the points $z \in \gamma_x$ such that $\{i : b_i(\gamma_x) = ty + (1-t)z \text{ with } t > 0\} \notin \mathcal{L}(f)$ and $\{i : b_i(\gamma_x) = ty + (1-t)z \text{ with } t < 0\} \notin \mathcal{L}(f)$ where $y \in \gamma_x \setminus \{z\}$. In words, the f-medians on γ_x are the points that leave a non-decisive coalition on both sides. The following is immediate.

Theorem 4.1.8. Let f be a simple rule and $\rho \in \mathcal{R}^n$ such that $\succsim_i$ is strictly convex for each $i \in N$. Then $x \in \mathcal{C}_f(\rho)$ iff $x \in \mu_f(\rho|_{\gamma_x})$ for all lines γ_x.

Theorem 4.1.9. Let f be a simple rule and $\rho \in \mathcal{R}^n$ given by $u_i : X \to \mathbb{R}$ concave C^2. Then $x \in \mathcal{C}_f(\rho)$ iff $0 \in \bigcap_{L \in \mathcal{L}(f)} p_L(x)$, where $p_L(x) \equiv \{\sum_{i \in L} \lambda_i \nabla u_i(x) : \lambda \in \mathbb{R}^L_{\geqq 0} \setminus 0\}$.

Proof. If $0 \notin p_L(x)$ for some L, take $V = \mathrm{co}(\{\nabla u_i(x) : i \in L\})$ and $U = \{0\}$; we have $U \cap V = \varnothing$ so by Hahn-Banach there is $\varphi \in \mathbb{R}^k$ with $\nabla u_i(x) \cdot \varphi > 0$ for all $i \in L$. Take $y = x + t\varphi$ with $t > 0$ small and $y \succ_i x$ for all $i \in L$, so $x \notin \mathcal{C}_f(\rho)$. If $x \notin \mathcal{C}_f(\rho)$, there is y and L with $y \succ_i x$ for all $i \in L$; by Taylor around x, $\nabla u_i(x) \cdot (y - x) > 0$. If $0 \in p_L(x)$, there is $\lambda \in \mathbb{R}^L_{\geqq 0} \setminus 0$ with $\sum_{i \in L} \lambda_i \nabla u_i(x) = 0$, so $\sum_{i \in L} \lambda_i \nabla u_i(x) \cdot (y - x) = 0$, absurd. ∎

Now we look at majority rule f_m with preferences given by $u_i : X \to \mathbb{R}$ concave C^2. Let $x \in X^\circ$ and $L_x = \{i \in N : \nabla u_i(x) \neq 0\}$. We say that x satisfies the *Plott conditions* if there is a bijection $\pi : L_x \to L_x$ such that for all $i \in L_x$ there is $\lambda > 0$ with $\nabla u_i(x) = -\lambda \nabla u_{\pi(i)}(x)$. In words, all i s.t. x is not her ideal point can be matched in pairs such that each $\nabla u_i(x)$ lies on the same line passing through x with its pair, but in opposite sides. For $u_i(x) = -\|x - x_i\|^2$ this means that the ideal points are aligned with respect to x in pairs.

Theorem 4.1.10 (Plott)**.** If $x \in X^\circ$ then $x \in \mathcal{C}_{f_m}(\rho)$ if x satisfies the Plott conditions; the converse holds if the ideal points are distinct and n is odd.

Proof. If x satisfies the Plott conditions, let $L \in \mathcal{L}(f_m)$; if $\nabla u_i(x) = 0$ for some i, $0 \notin p_L(x)$; otherwise $L \subset L_x$ and there are $i, \pi(i) \in L$, and again $0 \notin p_L(x)$. Hence $x \in \mathcal{C}_{f_m}(\rho)$. If the Plott conditions fail, there is $v \in \mathbb{R}^k \setminus 0$ such that $A^+ = \{i \in N : \nabla u_i(x) = tv \text{ for some } t > 0\}$

and $A^- = \{i \in N : \nabla u_i(x) = tv \text{ for some } t < 0\}$ have different cardinalities, so assume wlog $|A^+| > |A^-|$. Take $\varphi \neq 0$ s.t. $v \cdot \varphi = 0$ and $\nabla u_i(x) \cdot \varphi \neq 0$ for all $i \in L_x \setminus A^+ \setminus A^-$, and define $B^+ = \{i \in N : \nabla u_i(x) \cdot \varphi > 0\}$, $B^- = \{i \in N : \nabla u_i(x) \cdot \varphi < 0\}$. Assume wlog $|B^+| \geqq |B^-|$. Let $C = \{i \in N : \nabla u_i(x) = 0\}$. We have $|C| \leqq 1$, hence $|A^+| + |B^+| > \frac{n-1}{2}$ and $L = A^+ \cup B^+$ is a majority. Let $y = x + t(\varphi + \lambda v)$ with $t, \lambda > 0$ small. Clearly $y \succ_i x$ for every $i \in L$, so $x \notin \mathcal{C}_{f_m}(\rho)$. ■

A conclusion is that "generically" majority rule does not have a core if the space has dimension at least 2 and n is odd. Moreover, we have McKelvey's "chaos" theorem.

Theorem 4.1.11 (McKelvey)**.** Let $X = \mathbb{R}^k$, $k \geqq 2$, $n \geqq 3$, $u_i(x) = -\|x - x_i\|^2$. If the core is empty, then for any $x, y \in X$ there are $x = a_1, \ldots, a_r = y \in X$ with $a_t \succ_{f_m(u)} a_{t+1}$ for all t.

Proof. Suppose that the core $\mathcal{C}$ is empty. Let M be the set of pairs $(\varphi, c) \in \mathbb{R}^k \times \mathbb{R}$ s.t. $\|\varphi\| = 1$ and $|\{i \in N : \varphi \cdot x_i < c\}| \leqq \frac{n}{2}$. Suppose that $x \in \mathbb{R}^k$ is s.t. for every $(\varphi, c) \in M$ we have $\varphi \cdot x \geqq c$; since $x \notin \mathcal{C}$, there is a majority L s.t. $0 \notin p_L(x)$, hence there is (φ, c) s.t $\varphi \cdot x < c < \varphi \cdot x_i$ for all $i \in L$, so $(\varphi, c) \in M$ but $\varphi \cdot x < c$, absurd. Hence $\bigcap_{(\varphi,c)\in M} \{x : \varphi \cdot x \geqq c\} = \varnothing$. By Helly, we can take $(\varphi_1, c_1), \ldots, (\varphi_t, c_t)$ with $t \leqq k + 1$ minimal such that $\bigcap_{i=1}^t \{x : \varphi_i \cdot x \geqq c_i\} = \varnothing$ [how to avoid the need for compactness?]. Let $z_i \in \bigcap_{j \neq i} \{x \in A : \varphi_j \cdot x = c_j\}$ [why does it exists?] and $z = \frac{1}{t} \sum_{i=1}^t z_i$. We have $\varphi_i \cdot z < c$ for all i.

Let $x \in \mathbb{R}^k$. There is i s.t. $\varphi_i \cdot (x - z) \leqq 0$, since otherwise $\varphi_i \cdot [\alpha(x - z)] > c_i$ for big enough α. Let $\tilde{x} = x + (c - \varphi \cdot (2x - z))\varphi$. It is easy to verify that $\|\tilde{x} - x_i\|^2 < \|x - x_i\|^2$ for i s.t. $\varphi \cdot x_i \geqq c_i$ (which form a majority), and $\|\tilde{x} - z\|^2 \geqq \|x - z\|^2 + (c_i - \varphi_i \cdot z)^2$. We can iterate and obtain a sequence $x = a_1, a_2, \ldots$ with $\|a_r - z\|$ arbitrarily large. We have $\|x_i - a_r\| \geqq \|a_r - z\| - \|x_i - z\|$, so taking $\|a_r - z\| > \|x_i - y\| + \|x_i - z\|$ we get that everybody prefers y to a_r. ■

4.2 Voting

Arranco con el teorema de Arrow. Sea $N \in \mathbb{N}_{\geqq 1}$, $\mathcal{N} = \{1, \ldots, N\}$ el conjunto de votantes, A el conjunto de opciones y $L_A = \{R \in A^2 \mid \forall a, b, c \in A.\ ((a, b) \in R \vee (b, a) \in R) \wedge ((a, b) \in R \wedge (b, c) \in R \Rightarrow (a, c) \in R)\}$ el conjunto de preferencias sobre las opciones. Si $R \in L_A$ llamemos $a\ R\ b$ a la relación $(a, b) \in R$ pero $(b, a) \notin R$. Buscamos agregar las preferencias de los individuos mediante una función $\phi : L_A^{\mathcal{N}} \to L_A$.

Theorem 4.2.1 (Arrow)**.** Si $|A| \geqq 3$, $\phi : L_A^{\mathcal{N}} \to L_A$ es (1) débilmente Pareto, i.e., para todos $R \in L_A^{\mathcal{N}}$ y $a, b \in A$ tales que para todo $i \in \mathcal{N}$ vale $a\ R_i\ b$ se tiene $a\ \phi(R)\ b$, y (2) hay independencia de opciones irrelevantes (IIA), i.e., para todos $R, S \in L_A^{\mathcal{N}}$ y $a, b \in A$ tales que para todo $i \in \mathcal{N}$ vale $a\ R_i\ b$ sii $a\ S_i\ b$ entonces $a\ \phi(R)\ b$ sii $a\ \phi(S)\ b$, entonces hay un dictador, i.e., existe un votante k tal que para todos $R \in L_A^{\mathcal{N}}$ y $a, b \in A$ vale que $a\ R_k\ b$ implica $a\ \phi(R)\ b$.

Proof. Sean $a, b \in A$, $a \neq b$, y sea $R^0 \in L_A^{\mathcal{N}}$ tal que $a\ R_i^0\ b$ para todo i. Sea $R^i = R^0$ excepto que $b\ R_j^i\ a$ para $j \leqq i$. Tenemos por Pareto que $a\ \phi(R^0)\ b$ y $b\ \phi(R^N)\ a$, luego hay $k \in \mathcal{N}$ tal que $b\ \phi(R^k)\ a$ pero no $b\ \phi(R^{k-1})\ a$, i.e., $(a, b) \in \phi(R^{k-1})$. Llamemos $k_{b/a}$ a un k que cumple esto.

Sea $c \in A \setminus \{a, b\}$. Sea S' con $b\ S_i'\ c\ S_i'\ a$ para $i < k$ y $a\ S_i'\ b\ S_i'\ c$ para $i \geqq k$. Por IIA con R^{k-1} tenemos $(a, b) \in \phi(S')$, y por Pareto tenemos $b\ S'\ c$. Luego $a\ S'\ c$. Sea S'' con $b, c\ S_i''\ a$ para $i < k$, $b\ S_k''\ c\ S_k''\ a$ y $a\ S_i''\ b, c$ para $i > k$. Por IIA con R^k tenemos $b\ S''\ a$ y por IIA con S' tenemos $a\ S''\ c$. Luego $b\ S''\ c$. Entonces si S cumple $b\ S_k\ c$ tenemos $b\ \phi(S)\ c$ por IIA con S''. Ahora $k_{b/a} \leqq k_{c/b}$, ya que si $k_{b/a} > k_{c/b}$ armamos S con $c\ S_i\ b$ para $i \leqq k_{c/b}$ y $b\ S_i\ c$ para $i > k_{c/b}$ y tenemos $c\ \phi(S)\ b$, pero $b\ S_{k_{b/a}}\ c$, luego $b\ \phi(S)\ c$, absurdo. Entonces $k_{b/a} \leqq k_{c/b} \leqq k_{a/c} \leqq k_{b/a}$, y cambiando a y b tengo $k_{a/b} \leqq k_{b/a}$, por lo que $k_{a/b} = k_{b/a}$ y en general todos los k son iguales, por lo que hay un dictador. ■

Tenemos un conjunto finito de votantes $\mathcal{N} = \{1, \dots, N\}$ con $N \in \mathbb{N}_{\geq 1}$ y un conjunto de políticas $\mathcal{P}$. Decimos que $p^* \in \mathcal{P}$ es un *ganador de Condorcet* si para toda $p \in \mathcal{P} \setminus \{p^*\}$ por lo menos la mitad de los votantes prefiere p^* a p.

Definition 4.2.2. Si $\mathcal{P}$ tiene un orden total, decimos que $u : \mathcal{P} \to \mathbb{R}$ es *single-peaked* si existe $p^* \in \mathcal{P}$ tal que para todos p, p' con $p < p' \leqq p^*$ o $p^* \leqq p' < p$ tenemos $u(p) < u(p')$.

Sea p_m la política preferida por un votante mediano. El siguiente teorema es obvio.

Theorem 4.2.3. Si los votantes tienen utilidades single-peaked entonces p_m es un ganador de Condorcet.

Lo anterior asume democracia directa. Supongamos ahora que tenemos dos partidos, A y B, que ofrecen políticas $p_A, p_B \in \mathcal{P}$ y se comprometen a implementarlas. Lo único que les interesa es ganar: la utilidad de ambos es $Q > 0$ si ganan y 0 si no. Los votantes eligen el partido que ofrece la política que más prefieren; si son indiferentes, votan a A con probabilidad 1/2. El siguiente teorema es obvio.

Theorem 4.2.4 (Downs)**.** Si los votantes tienen utilidades single-peaked entonces $p_A = p_B = p_m$ es el único equilibrio.

Supongamos que los votantes deciden en función de $\alpha_i \in \mathbb{R}$.

Definition 4.2.5. Decimos que $u_i : \mathcal{P} \to \mathbb{R}$ es una *preferencia intermedia* para el votante i si $u_i(p) = f(p) + g(\alpha_i)h(p)$ para todo $p \in \mathcal{P}$, con $f, h : \mathcal{P} \to \mathbb{R}$ y $g : \mathbb{R} \to \mathbb{R}$ creciente.

Theorem 4.2.6. Si los votantes tienen preferencias intermedias, m es el votante con α_m mediano, y p_m es su política preferida, entonces p_m es un ganador de Condorcet.

Proof. Sea $p \in \mathcal{P}$. El votante i acepta votar p_m sobre p si $u_i(p_m) \geqq u_i(p)$. Ahora $u_i(p_m) - u_i(p) = g(\alpha_i)(h(p_m) - h(p))$. Como g es creciente, para todo i con $\alpha_i \geqq \alpha_m$ (o $\alpha_i \leqq \alpha_m$, dependiendo del signo de $h(p_m) - h(p)$) tengo que $u_i(p_m) - u_i(p) \geqq u_m(p_m) - u_m(p) \geqq 0$, luego esos i aceptan votar por p_m. ■

Definition 4.2.7. Si $\mathcal{P}$ tiene un orden total, decimos que $u_i(p) = u(p, \alpha_i)$ es una preferencia con la propiedad de *single-crossing* para el votante i si para $p > p'$ y $\alpha_i' > \alpha_i$, o $p < p'$ y $\alpha_i' < \alpha_i$, vale que $u(p, \alpha_i) > u(p', \alpha_i)$ implica que $u(p, \alpha_i') > u(p', \alpha_i')$.

El siguiente teorema es obvio.

Theorem 4.2.8. Si los votantes tienen preferencias single-crossing, m es el votante con α_m mediano, y p_m es su política preferida, entonces p_m es un ganador de Condorcet.

Supongamos que los votantes se dividen en grupos $g \in G$, de manera que la utilidad del votante i del grupo g es $u_i(p, P) = u_g(p) + \nu_{gi}(P)$, donde P es un partido que ofrece una política $p \in \mathcal{P} \subset \mathbb{R}^k$, con $\mathcal{P}$ convexo, y $\nu_{gi}(P)$ es una variable aleatoria que sigue una distribución con densidad f_g^P. Definimos $F_g^P(x) = \int_{-\infty}^x f_g^P$. Si los partidos son A y B, podemos fijar $\nu_{gi}^A = 0$ y poner $\nu_{gi} = \nu_{gi}(B)$. El votante i vota a A si $u_i(p_A, A) > u_i(p_B, B)$ (la probabilidad de que sean iguales es 0), es decir, con probabilidad $F_g(u_g(p_A) - u_g(p_B))$. Si el tamaño relativo del grupo g es λ_g, tenemos que la proporción esperada de votos de A es $\pi_A = \sum_{g \in G} \lambda_g F_g(u_g(p_A) - u_g(p_B))$, y la de B es $\pi_B = 1 - \pi_B$. Asumimos que los partidos buscan maximizar π_A y π_B (y no la probabilidad de ganar, que es más complicado; aunque si pensamos a los votantes como unidades de masa, o a $N \to +\infty$, es lo mismo). Sea

$$H_g(p, x) = \lambda_g f_g(x) \nabla\nabla u_g(p) + |f_g'(x)| \nabla u_g(p) \nabla u_g(p)^\top.$$

Sea $D_g = \max u_g - \min u_g$. Lo siguiente es obvio.

Theorem 4.2.9. Si f_g es derivable sobre $[-D_g, D_g]$, y $H_g(p,x)$ semidefinida negativa para todos $p \in \mathcal{P}$, $x \in [-D_g, D_g]$, entonces $\pi_A(\cdot, p_B)$ y $\pi_B(p_A, \cdot)$ son cóncavas y existe un equilibrio de Nash en estrategias puras.

En particular, si u_g son cóncavas, $\nu_{gi} \sim \mathcal{U}(-\frac{1}{2\phi_g}, \frac{1}{2\phi_g})$ con $0 < \phi_g \leqq \frac{2}{D_g}$, entonces existe un equilibrio de Nash en estrategias puras. En ese caso se ve que $p_A = p_B = p^*$ que maximiza $\sum_{g \in G} \lambda_g \phi_g u_g(p)$.

4.3 Electoral accountability

4.4 Lobbying

Common agency (Dixit et al., 1997) There is a set of principals L and an agent who has to choose an action $a \in A$. First, the principals choose compensations to the agent $C_i : A \to \mathbb{R}_+$ (subject to $C_i \leqq \overline{C}_i$ for some $\overline{C}_i$). Then the agent chooses $a \in A$ to maximize her utility $G(a, C(a))$, and each principal gets $W_i(a, C_i(a))$. We assume that G and W_i are continuous, $G(a, \cdot)$ is increasing and $W_i(a, \cdot)$ is decreasing. A *truthful strategy* relative to $W_i^* \in \mathbb{R}_+$ is C_i such that $W_i(a, C_i(a)) = W_i^*$ if $W_i^* \in [W_i(a, 0), W_i(a, \overline{C}_i(a))]$, $C_i(a) = 0$ if $W_i^* < W_i(a, 0)$, and $C_i(a) = \overline{C}_i(a)$ if $W_i^* > W_i(a, \overline{C}_i(a))$.

Theorem 4.4.1. If there exists an equilibrium in truthful strategies, then (1) it is Pareto efficient in the following sense: there does not exists $\tilde{a} \in A$, $\tilde{C}$ such that $G(\tilde{a}, \tilde{C}) \geqq G(a^*, C^*(a))$, $W_i(\tilde{a}, \tilde{C}_i(\tilde{a})) \geqq W_i(a^*, C_i^*(a^*))$ for all $i \in L$ and at least one inequality is strict. (2) For every $i \in L$ we have $G(a^*, C^*) = \max_{a \in A} G(a, (C_{-i}^*, 0))$.

Proof. To see (1), $W_i(\tilde{a}, \tilde{C}_i(\tilde{a})) \geqq W_i(a^*, C_i^*(a^*))$ implies $W_i(\tilde{a}, \tilde{C}_i(\tilde{a})) \geqq W_i(\tilde{a}, C^*(\tilde{a}))$ and $\tilde{C}(\tilde{a}) \leqq C^*(\tilde{a})$, but then $G(\tilde{a}, \tilde{C}) \leqq G(\tilde{a}, C^*) \leqq G(a^*, C^*) \leqq G(\tilde{a}, \tilde{C})$, so all are equalities, absurd. To see (2), if $G(a^*, C^*) > \max_{a \in A} G(a, (C_{-i}^*, 0))$, consider $\tilde{C}_i$ given by $\tilde{C}_i(a^*) = C^*(a^*) - \epsilon$ and $\tilde{C}_i(a) = 0$ for $a \neq a^*$; by continuity, the agent still chooses a^* for some $\epsilon > 0$, but i is better off, so C^* cannot be an equilibrium. ■

With quasi-linear preferences $W_i(a, c) = w_i(a) - c$ and $G(a, c) = g(a) + \sum_{i \in L} c_i$, if C^*, a^* is a truthful equilibrium and $0 < C_i^*(a^*) < \overline{C}_i(a^*)$ for all i, the Theorem tells us that a^* must maximize $g(a) + \sum_{i \in L} w_i(a)$.

4.5 Influence

Basic setup The basic model of influence is

1. Nature chooses $\omega \in \Omega$ following $\omega \sim \mu_\omega$.
2. A observes a signal $\sigma \in \Sigma \sim \mu_\sigma(\cdot|\omega)$.
3. A chooses an action $a \in A$.
4. P observes $\tau \in T \sim \mu_\tau(\cdot|a, \sigma)$.
5. P chooses $x \in X$.
6. Player $i = A, P$ receives $u_i(x, a, \omega)$.

An equilibrium is a pair of strategies $a : \Sigma \to A$ and $x : A \times T \to X$ and a belief $\beta : A \times T \to \Delta(\Omega)$ that follows Bayes rule such that $x(a, \tau)$ maximizes $\mathbb{E}u_P$, etc.

Cheap talk is a special case when $\tau = a$ and the payoffs do not depend on a, which is interpreted as a cost-less message. If a entails a cost to A but is payoff-irrelevant to P, we get a model of money burning.

Cheap talk when $T = \{0, 1\}$. A has to choose $p : \Sigma \to [0, 1]$, so P infers $\mu(\omega) = \int p(a|\sigma) d\mu_\sigma(\sigma|\omega) \mu_\omega(\omega)$ and chooses x to max $\int \int u_P(x, a, \omega) p(a|\sigma) d\mu_\sigma(\sigma|\omega) d\mu_\omega(\omega)$

4.6 Redistribution

Static redistribution The government has a linear tax $\tau \in [0,1]$ and redistributes its revenue equally among the citizens $i \in [0,1]$. Each citizen has a productivity $\omega_i > 0$ and access to a production function $y_i = \frac{1}{\alpha}\omega_i^{1-\alpha}e_i^\alpha$, where $e_i \geqq 0$ is effort and $\alpha \in (0,1)$. The citizen's utility function is $u_i = c_i - e_i$, where c_i is consumption. Each citizen receives a transfer $\tau \int_0^1 y_i\, di = \tau\overline{y}$, so $u_i = (1-\tau)y_i + \tau\overline{y} - e_i$. Given τ, the citizen chooses $e_i = (1-\tau)^{\frac{1}{1-\alpha}}\omega_i$, and $u_i = \frac{1}{\alpha}(1-\tau)^{\frac{1}{1-\alpha}}\left[(1-\alpha)\omega_i + \tau\int_0^1 \omega_j\, dj\right]$. We have $\partial_\tau\partial_{\omega_i}u_i < 0$, so preferences over τ are single-crossing, which implies that if $\partial_i\omega_i > 0$ (voters are ordered by their productivity) then the median voter's most preferred τ is the Condorcet winner. This also implies that the smaller ω_i is, and also the smaller $\omega_i / \int_0^1 \omega_j\, dj$ is, the higher the most preferred τ is. The pre-tax income $y_i - e_i$ is $\frac{1-\alpha}{\alpha}\omega_i$. Hence if inequality increases, i.e., the median over mean pre-tax income decreases, τ increases. More inequality leads to more redistribution. This is consistent with Meltzer and Richard (1981).

Insurance The government has a linear tax $\tau \in [0,1]$ the applies to the employed among the citizens $i \in [0,1]$. Each citizen has a productivity $\omega_i \leqq 0$, and a probability $q_i \in [0,1]$ of being employed at each period. If employed, i produces $y_i = \frac{1}{\alpha}\omega_i^{1-\alpha}e_i^\alpha$ and gets $c_i = (1-\tau)y_i$. If not, it receives an insurance payment from the government b, which is the tax revenue divided by the measure of unemployed citizens. The citizens are risk averse, with CRRA $\rho > 1$. Citizens decide τ, then the employment status is revealed, and the employed citizens decide their efforts. Employed citizens choose $e_i = (1-\tau)^{\frac{1}{1-\alpha}}\omega_i$. Hence at the time of deciding τ, $u_i = q_i u(c_i - e_i) + (1-q_i)u(b)$, where $u(c) = \frac{c^{1-\rho}}{1-\rho}$. We have

$$u_i = q_i u\left(\frac{1-\alpha}{\alpha}(1-\tau)^{\frac{1}{1-\alpha}}\omega_i\right) + (1-q_i)u\left(\frac{1}{\alpha}\tau(1-\tau)^{\frac{1}{1-\alpha}}\frac{\int_0^1 q_j\omega_j\, dj}{1-\int_0^1 q_j\, dj}\right).$$

It is easy to see that if $\rho \to 1$ then $\partial_\tau\partial_i u_i < 0$, assuming that $\partial_i q, \partial_i\omega_i > 0$, so u_i is single-crossing and the median voter decides τ. Now $\partial_{\omega_i}\partial_\tau u_i > 0$, so, if inequality increases, keeping q and the average expected pre-tax income $\frac{1-\alpha}{\alpha}\int_0^1 q_j\omega_j\, dj$ fixed, the demand for insurance τ decreases. Conditional on the risk distribution, more income inequality leads to less insurance. This is consistent with Moene and Wallerstein (2001).

Now, $\partial_{q_i}\partial_\tau u_i < 0$, so if the distribution of q_i is right-skewed (the median is greater than the mean) then if inequality increases with $\int_0^1 q_i\, d_i$ fixed, the median voter has higher q_i, and demands less insurance. So, conditional on the unemployment rate and the income distribution, more inequality in the risk-pool leads to less insurance. This is consistent with Rehm (2011).

4.7 Institutions

Regime change (Acemoglu & Robinson, 2006) First, there are $1-\delta > \frac{1}{2}$ poor (p) and δ elite (r), with income y^p and y^r. Let $\overline{y} = (1-\delta)y^p + \delta y^r$ and $\theta = \frac{\delta y^r}{\overline{y}}$, so $y^p = \frac{1-\theta}{1-\delta}\overline{y}$ and $y^r = \frac{\theta}{\delta}\overline{y}$; θ measures inequality. Assume that $0 < y^p < \overline{y} < y^r$. There is a constant tax τ so that i gets $c^i(\tau) = (1-\tau)y^i + (\tau - C(\tau))\overline{y}$, where $C(0) = 0$, $C(1) = 1$, $C' > 0$, $C'' > 0$ is the distortion. With probability q nature sets $s = H$ (otherwise $s = L$), the poor can revolt (choose $\rho = 1$) and get $\frac{1-\mu}{1-\delta}\overline{y}$ in the present and in the future, while the rich get nothing. In each period it is revealed whether the poor can revolt, the elite choose τ, and the poor choose ρ. The payoff is $U^i = \sum_{t\geqq 0}\beta^t\mathbb{E}[u^i]$. If $\theta \leqq \mu$ the poor never revolt. If not, in an MPE the elite choose $\tau = 0$ when there is no risk of revolt and τ^* such that $c^p(\tau^*) + \frac{\beta}{1-\beta}(qc^p(\tau^*) + (1-q)y^p) = \frac{1-\mu}{(1-\beta)(1-\delta)}\overline{y}$ if it exists, i.e., if $\mu \geqq \mu^*$, in which case there is no revolt; if $\mu < \mu^*$, the poor choose to revolt, where μ^* is defined by $c^p(\tau^p) + \frac{\beta}{1-\beta}(qc^p(\tau^p) + (1-q)y^p) = \frac{1-\mu^*}{(1-\beta)(1-\delta)}\overline{y}$, where $\tau^p = \arg\max c^p$.

In a SPE, the rich can set $\tau = \tau^s$ in each state $s \in \{L, H\}$, and the poor set $\rho = 1$ when the rich set $\tau < \tau^s$ in any past state, under the following conditions. Note first that the elite set $\tau = 0$ if they deviated in the past, because they know they will be punished. When $s = L$, the rich have to prefer setting $\tau = \tau^L$ rather than deviate and be punished in the future: $c^r(\tau^L) + \frac{\beta}{1-\beta}(qc^r(\tau^H) + (1-q)c^r(\tau^L)) \geqq \frac{1}{1-\beta(1-q)}y^r$. When $s = H$ under no deviation, so $\tau = \tau^H$, the poor have to prefer to set $\rho = 0$: $c^p(\tau^H) + \frac{\beta}{1-\beta}(qc^p(\tau^H) + (1-q)c^p(\tau^L)) \geqq y^R$, where $y^R = \frac{1-\mu}{(1-\beta)(1-\delta)}\overline{y}$. If there is a deviation, the poor have to prefer to revolt: $y^R \geqq c^p(\tau) + \frac{\beta}{1-\beta(1-q)}(qy^R + (1-q)y^p)$ for any τ. This expands the set of μ for which in equilibrium the poor do not revolt to $\mu \geqq \mu^{**}$ for some $\mu^{**} > 0$.

Second, we add the possibility of repression and democratization. In each period $s \in \{H, L\}$ is revealed; the elite choose whether to use repression $\omega \in \{0, 1\}$; if they use it, the stage ends; if not, they choose whether to democratize $\phi \in \{0, 1\}$; if they decide to democratize, the poor decide τ forever; if not, the elite set τ, and the citizens decide whether to revolt $\rho \in \{0, 1\}$. If $\rho = 1$ the poor get $\frac{1-\mu}{1-\delta}\overline{y}$ forever and the elite get nothing; if $\rho = 0$, $u_i = \omega(1-\kappa)y^i + (1-\omega)c^i(\tau)$. If $s = L$ or $\mu \geqq \theta$, in equilibrium $\phi = \omega = \tau = \rho = 0$. Otherwise, if the elite set $\phi = 1$, they get $\frac{1}{1-\beta}c^r(\tau^p)$; if they set $\omega = 1$, they get $(1 - \kappa + \frac{\beta(1-\kappa q)}{1-\beta})y^r$. Let μ^* be given by $c^p(\tau^p) + \frac{\beta}{1-\beta}(qc^p(\tau^p) + (1-q)y^p) = \frac{1-\mu^*}{(1-\beta)(1-\delta)}\overline{y}$; if $\mu < \mu^*$, the elite use democratization or repression (the former if repression is sufficiently costly, $\kappa > \overline{\kappa}$); if $\mu \geqq \mu^*$, the elite choose repression or getting $c^r(\tau^*) + \frac{\beta}{1-\beta}(qc^r(\tau^*) + (1-q)y^r)$, where τ^* is defined by $c^p(\tau^*) + \frac{\beta}{1-\beta}(qc^p(\tau^*) + (1-q)y^p) = \frac{1-\mu}{(1-\beta)(1-\delta)}\overline{y}$, again depending whether repression is costly enough. Therefore democratization occurs iff the cost of repression is high ($\kappa > \overline{\kappa}$) and the cost of revolution is low ($\mu < \mu^* < \theta$). We see that $\frac{\partial \tau^p}{\partial \theta} > 0$, $\frac{\partial c^p(\tau^p)}{\partial \theta} < 0$, $\frac{\partial \mu^*}{\partial \theta} > 0$ and $\frac{\partial \overline{\kappa}}{\partial \theta} > 0$. Democracy emerges with intermediate levels of inequality.

Third, we add the possibility of a coup. In each period $s \in \{H, L\}$ is revealed; is the regime is non-democratic (N), the model is as before; if it is democratic (D), the poor set τ and if $s = H$, with probability s (sorry for the repetition) the elite choose whether to mount a coup $\zeta \in \{0, 1\}$; if $\zeta = 1$, a fraction φ of income is lost in that period, the elite sets τ, and the regime becomes non-democratic. I will not describe the MPE, but according to μ, κ and φ you get N forever without repression, or with repression, or a transition to a fully consolidated D when $s = H$ (in which $\tau = \tau^p$ always), or a semi-consolidated D (in which $\tau < \tau^p$ when $s = H$), or an unconsolidated D in which the elite mount a coup whenever $s = H$ (and thus there is a continuous switch between regimes). In D, less inequality makes a switch back to N less likely. There is a non-monotonic relationship between inequality and redistribution, since in D, θ has a positive effect on τ^p, but if it is sufficiently large, it destabilizes the regime, and τ can decrease. Hence societies with intermediate levels of inequality redistribute the most. Higher inequality breeds regime instability, which leads to fiscal volatility (changes in τ). If C decreases, τ^p will be larger, and democracy will be costlier to the elite, and hence less likely to be consolidated; when taxation is less distortionary (for example when a large fraction of GDP is generated from natural resources), democracies may be harder to consolidate.

Persistence of elite power (Acemoglu & Robinson, 2008) There are L citizens C and M elites E; there are two possible political institutions s_t at time t: non-democracy N and democracy D. An agent of type $h \in \{C, E\}$ derives utility $\sum_{t=0}^{\infty} \beta^t(c_t^i + G_t^h)$, where c_t^i is consumption and G_t^h is utility from a public good; $G_t^C = \gamma^C \mathbb{1}[s_t = D]$ and $G_t^E = \gamma^E \mathbb{1}[s_t = N]$. There are two possible economic institutions τ_t at t: labor repressive e and competitive c; if $\tau_t = c$, wage $c_t^C = A > 0$ and rents $c_t^E = 0$; if $\tau_t = e$, $c_t^C = \lambda(1-\delta)A$ and $c_t^E = R = (1-\lambda)(1-\delta)AL/M$. We have $\tau_t = e$ iff the elite are more powerful than the citizens; elite member i invests $\theta_i \geqq 0$ in power; citizens have $\omega_t + \eta\mathbb{1}[s_t = D]$ power, where $\omega_t \sim F$ is random and $\eta > 0$;

hence $\tau_t = e$ iff $\phi^{s_t} \sum_{i \in \mathcal{E}} \theta_i \geqq \omega_t + \eta \mathbb{1}[s_t = D]$, where $\phi^s \geqq 0$. If $\phi^{s_t} \sum_{i \in \mathcal{E}} \theta_i \geqq \omega_t + \xi \mathbb{1}[s_t = D]$, with $\xi \geqq \eta$, the elite choose s_{t+1}; if the opposite holds, the citizens choose s_{t+1}. We consider symmetric MPEs, in which θ_i is the same for any member of the elite. Suppose $\phi^N = \phi^D$; if $\eta = \xi$, for interior equilibria the elite choose θ_i so that the probability that $s_{t+1} = N$ is the same in democracy and in non-democracy; if $\eta < \xi$, $\Pr(s_{t+1} = N)$ is higher under $s_t = N$, but in this situation the probability that the economic institutions are in line with elite's interests is higher under democracy. If $\phi^N = \phi^C$, in interior equilibria θ_i and $\Pr(s_{t+1} = N)$ are increasing in R, β, η, and decreasing in M; $\Pr(s_{t+1} = N)$ is increasing in ϕ^E.

4.8 Reading list

On institutions:

1. Acemoglu, Modelling Ineffcient Institutions
2. Sonin, Why the rich may favor poor protection of property rights
3. Robinson, When is the state predatory?
4. Acemoglu, Golosov, Tsyvinski, Power fluctuations and political economy
5. Dixit, Predatory States and Failing States
6. Acemoglu, Wolitzky, The Economics of Labor Coercion
7. Acemoglu (2005) Politics and Economics in Weak and Strong States
8. Besley, Persson (2009) The Origins of State Capacity
9. Acemoglu, Robinson (2017) The Emergence of Weak, Despotic and Inclusive States
10. Albertus, Menaldo (2013) Gaming Democracy: Elite Dominance During Transition and the Prospects for Redistribution
11. Acemoglu, Ticchi, Vindigni (2010) Emergence and Persistence of Inefficient States
12. Svolik (2012) The Politics of Authoritarian Rule

Accountability:

- Barro
- Ferejohn (1986)
- Fearon
- Persson, Tabellini
- Svolik (2013) Learning to love democracy
- Espinosa (JMP) Corruption and Political Turnover over the Business Cycle: Theory and Evidence
- Nannicini, T., Stella, A., Tabellini, G., & Troiano, U. (2013). Social Capital and Political Accountability. American Economic Journal: Economic Policy, 5(2), 222–250.

Unions:

- Oswald (1985) The Economic Theory of Trade Unions: An Introductory Survey

Collective action:

- Olson
- Simple models
- Information models (Lohmann)
- Kuran (1989) Sparks and Prairie Fires: A Theory of Unanticipated Political Revolution
- Roemer, Kantian optimization

Parties
Beliefs/information

5 Statistics

Antes de empezar hago un repaso lo más rápido posible de teoría de probabilidad básica. La teoría de la medida que uso (convergencia dominada, Radon-Nikodym, teorema de diferenciación de Lebesgue, Fubini) está en otro apunte.

5.1 Probabilidad

Un *espacio de probabilidad* es un espacio de medida $(\Omega, \mathcal{B}, P)$, donde Ω es el conjunto de *outcomes*, $\mathcal{B}$ es la σ-álgebra sobre Ω de los *eventos*, y P es una medida positiva con $P(\Omega) = 1$. Una *variable aleatoria* es una función medible $X : \Omega \to \mathcal{X}$, donde $\mathcal{X}$ es un espacio topológico, es decir, $X^{-1}(\mathcal{U}) \in \mathcal{B}$ para todo abierto $\mathcal{U} \subset \mathcal{X}$. Define una probabilidad sobre $\mathcal{X}$ dada por $\mu_X(E) = P(X \in E)$, su *distribución*. Define una σ-álgebra $\sigma(X) \subset \mathcal{B}$ dada por $X^{-1}(\mathcal{B}_\mathcal{X})$, donde $\mathcal{B}_\mathcal{X}$ es la σ-álgebra de Borel de $\mathcal{X}$. Decimos que dos variables $X, Y : \Omega \to \mathcal{X}$ tienen la misma distribución si $\mu_X = \mu_Y$, y lo notamos $X \sim Y$. Decimos que X es una *variable discreta* si $\mathcal{X}$ es discreto y numerable, y *absolutamente continua* si $\mathcal{X}$ tiene una medida σ-finita $\mu_\mathcal{X}$ y $\mu_X \ll \mu_\mathcal{X}$ (i.e., $\mu_X(E) = 0$ si $\mu_\mathcal{X}(E) = 0$), es decir, por Radon-Nikodym, sii hay una función medible $f : \mathcal{X} \to \mathbb{R}_{\geqq 0}$, la *función de densidad* de X, con $\mu_X(E) = \int_E f \, d\mu_\mathcal{X}$. Si $\mathcal{X} = \mathbb{R}^d$ definimos la *función de distribución* $F_X : \mathbb{R}^d \to [0, 1]$ dada por $F_X(a) = P(X \leqq a)$.

Si $X : \Omega \to \mathbb{R}$ es variable aleatoria integrable, definimos la *esperanza* como $\mathbb{E}X = \int X \, dP = \int x \, d\mu_X$; es lineal y, si f es medible, $\mathbb{E}[f(X)] = \int f \, d\mu_X$. La *varianza* de X se define como $V(X) = \mathbb{E}[(X-\mathbb{E}X)^2] = \mathbb{E}X^2-(\mathbb{E}X)^2$. La *covarianza* es $\operatorname{Cov}(X, Y) = \mathbb{E}[(X-\mathbb{E}X)(Y-\mathbb{E}Y)] = \mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y$ y la *correlación* es $\rho(X, Y) = \frac{\operatorname{Cov}(X,Y)}{\sqrt{V(X)V(Y)}} \in [-1, 1]$. Tenemos (Markov) $P(|X| \geqq t) \leqq \frac{1}{t}\mathbb{E}|X|$ y (Chebyshev) $P(|X - \mathbb{E}X| \geqq t) \leqq \frac{1}{t^2}V(X)$. Si $X : \Omega \to \mathbb{R}^n$, definimos $\mathbb{E}X = (\mathbb{E}X_1, \dots, \mathbb{E}X_n)$ y $V(X) = (\operatorname{Cov}(X_i, X_j))_{i,j=1}^n$.

La *probabilidad condicional* de A dado B es $P(A|B) = \frac{P(A\cap B)}{P(B)}$. Dos eventos A y B se dicen *independientes* si $P(A \cap B) = P(A)P(B)$; $\{A_i\}_{i\in I}$ si dicen independientes si $P(\bigcap_{j\in J} A_j) = \prod_{j\in J} P(A_j)$ para todo $J \subset I$ finito. Dos variables aleatorias X e Y se dicen independientes si $P(X \in A \wedge Y \in B) = P(X \in A)P(Y \in B)$ para todos A, B; esto sucede sii $\mu_{(X,Y)} = \mu_{X\times Y}$, la medida producto. En ese caso $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ por Fubini, y $V(X + Y) = V(X) + V(Y)$. Si X, Y son absolutamente continuas entonces son independientes sii (X, Y) es absolutamente continua y $f_{(X,Y)}(x, y) = f_X(x)f_Y(y)$ en casi todo punto (por el teorema de diferenciación de Lebesgue).

Si $X : \Omega \to \mathbb{R}$ es una variable aleatoria y $Y : \Omega \to \mathcal{X}$ es otra, definimos la *esperanza condicional* de X respecto a Y como una variable aleatoria $\mathbb{E}(X|Y) : (\mathcal{X}, \mathcal{B}_\mathcal{X}, \mu_Y) \to \mathbb{R}$ que satisface $\int_E \mathbb{E}(X|Y) \, d\mu_Y = \int_{Y\in E} X \, dP$ para todo $E \in \mathcal{B}_\mathcal{X}$. Existe: ν definida por $\nu(E) = \int_{Y\in E} X \, dP$ es una medida sobre $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ y satisface $\nu \ll \mu_Y$, por lo que Radon-Nikodym hace el resto. Es lineal, $\mathbb{E}(X|Y) \geqq 0$ si $X \geqq 0$ y $\mathbb{E}[\mathbb{E}(X|Y)] = \mathbb{E}(X)$. Si $f : \mathcal{X} \to \mathbb{R}$ es $\mathcal{B}_\mathcal{X}$-medible, $\mathbb{E}[f(Y)X|Y] = f(Y)\mathbb{E}[X|Y]$. Si X, Y son independientes, $\mathbb{E}(X|Y) = \mathbb{E}(X)$. Definimos la *varianza condicional* como $V(X|Y) = \mathbb{E}[(X - \mathbb{E}(X|Y))^2|Y] = \mathbb{E}(X^2|Y) - \mathbb{E}(X|Y)^2$. Ley de varianza total: $V(X) = \mathbb{E}[V(X|Y)] + V(\mathbb{E}(X|Y))$.

Supongamos que $X \in L^2(\Omega)$, i.e., $V(X) < \infty$. Sea $S = \{f(Y) \mid f \in L^2(\mathcal{X}, \mathcal{B}_\mathcal{X}, \mu_Y)\}$; es un subespacio cerrado de $L^2(\Omega)$. Como $L^2(\Omega)$ es un espacio de Hilbert, existe $\operatorname{pr}_S(X) = f(Y)$ con $\mathbb{E}[(X - f(Y))^2]$ mínimo, y cumple $\mathbb{E}[f(Y)g(Y)] = \mathbb{E}[Xg(Y)]$ para toda $g \in L^2(\mathcal{X})$. En particular, para $g = \mathbb{1}_{Y\in E}$ con $E \in \mathcal{B}_\mathcal{X}$ tenemos $\int_E f \, d\mu_Y = \int_{Y\in E} X \, dP$, por lo que $f = \mathbb{E}(X|Y)$ en casi todo punto. Esto nos dice que $\mathbb{E}(X|Y)$ minimiza $\mathbb{E}[(X - f(Y))^2]$ sobre $f : \mathcal{X} \to \mathbb{R}$ μ_Y-medibles y que es única (i.e., si hay otra, son iguales en casi todo punto).

Sean X_n y X variables aleatorias $\Omega \to \mathcal{X}$. Decimos que X_n converge a X *casi seguramente*, escrito $X_n \overset{as}{\to} X$, si $P(\lim_{n\to\infty} X_n = X) = 1$, es decir, si $X_n \to X$ en casi todo punto. Notar

Table 1: Distribuciones

X	Parámetros	Soporte	Densidad
$\mathcal{U}(a, b)$	$a < b$	$x \in [a, b]$	$\dfrac{1}{b-a}$
$N(\mu, \Sigma)$	$\mu \in \mathbb{R}^n$, $\Sigma \in \mathbb{R}^{n \times n}$, Σ simétrica definida positiva	$x \in \mathbb{R}^n$	$(2\pi)^{-\frac{n}{2}} \det \Sigma^{-\frac{1}{2}} \times$ $\exp(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu))$
$\mathrm{Exp}(\lambda)$	$\lambda > 0$	$x \geqq 0$	$\lambda e^{-\lambda x}$
$\mathrm{Gamma}(a, b)$	$a, b > 0$	$x > 0$	$\dfrac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}$
$\mathrm{InvGaussian}(\mu, \lambda)$	$\mu, \lambda > 0$	$x > 0$	$\sqrt{\dfrac{\lambda}{2\pi}} \dfrac{1}{x^{3/2}} \exp\left(-\dfrac{\lambda(x-\mu)^2}{2\mu^2 x}\right)$
Generalized Gamma(a, b, p)	$a, b, p > 0$	$x > 0$	$\dfrac{b^a p}{\Gamma(a/p)} x^{a-1} e^{-(bx)^p}$
$\mathrm{Beta}(a, b)$	$a, b > 0$	$x \in (0, 1)$	$\dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} x^{a-1}(1-x)^{b-1}$
$\mathrm{Dirichlet}(a)$	$a \in \mathbb{R}^n_{>0}$	$x \in (0, 1)^n$, $\sum_{i=1}^n x_i = 1$	$\dfrac{\prod_{i=1}^n \Gamma(a_i)}{\Gamma(\sum_{i=1}^n a_i)} \displaystyle\prod_{i=1}^n x_i^{a_i - 1}$
$\mathrm{Categorical}(p)$	$p \in [0, 1]^n$, $\sum_{i=1}^n p_i = 1$	$x \in \{0, 1\}^n$, $\sum_{i=1}^n x_i = 1$	$\prod_{i=1}^n p_i^{x_i}$
$\mathrm{Binomial}(n, p)$	$n \in \mathbb{N}$, $p \in [0, 1]$	$x \in \{0, \dots, n\}$	$\binom{n}{x} p^x (1-p)^{n-x}$
$\mathrm{Multinomial}(n, p)$	$n \in \mathbb{N}$, $p \in [0, 1]^k$	$x \in \{0, \dots, n\}^k$, $\sum_{i=1}^k x_i = n$	$\frac{n!}{x_1! \dots x_k!} \prod_{i=1}^k p_i^{x_i}$
$\mathrm{Poisson}(\lambda)$	$\lambda > 0$	$x \in \mathbb{N}_{\geqq 0}$	$\frac{\lambda^x e^{-\lambda}}{x!}$
Negative Binomial(r, p)	$r > 0$, $p \in (0, 1)$	$x \in \mathbb{N}_{\geqq 0}$	$\binom{x+r-1}{x} p^x (1-p)^r$

Table 2: Distribuciones

X	$\mathbb{E}(X)$	$V(X)$	Moda
$\mathcal{U}(a,b)$	$\dfrac{b+a}{2}$	$\dfrac{(b-a)^2}{12}$	todo $x \in [a,b]$
$N(\mu,\Sigma)$	μ	Σ	μ
$\mathrm{Exp}(\lambda)$	$1/\lambda$	$1/\lambda^2$	0
$\mathrm{Gamma}(a,b)$	a/b	a/b^2	$\max\left\{\dfrac{a-1}{b}, 0\right\}$
$\mathrm{InvGaussian}(\mu,\lambda)$	μ	μ^3/λ	$\mu\left(\sqrt{1+\dfrac{9\mu^2}{4\lambda^2}} - \dfrac{3\mu}{2\lambda}\right)$
Generalized Gamma(a,b,p)	$\dfrac{\Gamma((a+1)/p)}{b\Gamma(a/p)}$	$\frac{1}{b}\mathbb{E}(X) - \mathbb{E}(X)^2$	$\dfrac{\max\{a-1,0\}^{1/p}}{bp^{1/p}}$
$\mathrm{Beta}(a,b)$	$\dfrac{a}{a+b}$	$\dfrac{ab}{(a+b)^2(a+b+1)}$	$\begin{cases} \frac{a-1}{a+b-2} & \text{si } a,b>1, \\ 0 \text{ ó } 1 & \text{si no} \end{cases}$
$\mathrm{Dirichlet}(a)$	$\dfrac{a}{\sum_{i=1}^n a_i}$	$\dfrac{\mathrm{diag}(a_0 a) - aa^\top}{a_0^2(a_0+1)}$, con $a_0 = \sum_{i=1}^n a_i$	$\dfrac{a-\iota}{a_0-n}$ si $a > \iota$, con $\iota = (1,\dots,1)$
$\mathrm{Categorical}(p)$	p	$\mathrm{diag}(p) - pp^\top$	
$\mathrm{Binomial}(n,p)$	np	$np(1-p)$	$\lfloor (n+1)p \rfloor$ o $\lceil (n+1)p \rceil - 1$
$\mathrm{Multinomial}(n,p)$	np	$n(\mathrm{diag}(p) - pp^\top)$	
$\mathrm{Poisson}(\lambda)$	λ	λ	$\lfloor \lambda \rfloor$ o $\lceil \lambda \rceil - 1$
Negative Binomial(r,p)	$\dfrac{pr}{1-p}$	$\dfrac{pr}{(1-p)^2}$	$\max\left\{\left\lfloor \dfrac{p(r-1)}{1-p} \right\rfloor, 0\right\}$

que esto es equivalente a que X_n converge a X casi uniformemente, esto es, dado $\epsilon > 0$ existe $E \subset \Omega$ medible con $P(\Omega \smallsetminus E) < \epsilon$ tal que $X_n \to X$ uniformemente en E. Decimos que X_n converge a X *en probabilidad*, escrito $X_n \xrightarrow{p} X$, si para todos $\epsilon, \delta > 0$ existe $n_0 \in \mathbb{N}$ tal que $P(|X_n - X| \geqq \delta) < \epsilon$ para todo $n \geqq n_0$. Claramente $X_n \xrightarrow{as} X$ implica $X_n \xrightarrow{p} X$. Por otro lado, $X_n \xrightarrow{p} X$ sii toda subsecuencia X_{n_k} tiene una sub-subsecuencia $X_{n_{k_r}} \xrightarrow{as} X$. Vale que si $X_n \xrightarrow{p} X$ y $Y_n \xrightarrow{p} Y$ entonces $(X_n, Y_n) \xrightarrow{p} (X, Y)$ y que si f es continua (salvo en un conjunto de medida nula) entonces $f(X_n) \xrightarrow{p} f(X)$. Decimos que X_n es *acotada en probabilidad*, o $X_n = O_P(1)$, si para todo $\epsilon > 0$ existe C tal que $P(\|X_n\| \geqq C) < \epsilon$ para todo n. Escribimos $X_n = O_P(a_n)$ para decir $a_n^{-1} X_n = O_P(1)$ y $X_n = o_P(a_n)$ para decir $a_n^{-1} X_n \xrightarrow{p} 0$.

Decimos que X_n converge a X *en distribución*, escrito $X_n \xrightarrow{d} X$, si se da alguna de las siguientes condiciones equivalentes.

Theorem 5.1.1 (Portmanteau)**.** Sean $X_n, X : \Omega \to \mathcal{X}$ variables aleatorias, con $\mathcal{X}$ un espacio métrico. Equivalen: (1) $\mathbb{E} f(X_n) \to \mathbb{E} f(X)$ para toda función continua y acotada $f : \mathcal{X} \to \mathbb{R}$, (2) $\mathbb{E} f(X_n) \to \mathbb{E} f(X)$ para toda función acotada y Lipschitz $f : \mathcal{X} \to \mathbb{R}$ (i.e., tal que existe $M > 0$ tal que $|f(x) - f(y)| \leqq M d(x, y)$), (3) $\underline{\lim}\, P(X_n \in G) \geqq P(X \in G)$ para todo abierto G, (4) $\overline{\lim}\, P(X_n \in F) \leqq P(X \in F)$ para todo cerrado F, (5) $P(X_n \in E) \to P(X \in E)$ para todo boreliano E tal que $P(X \in \partial E) = 0$, donde $\partial E = \overline{E} \smallsetminus E^\circ$, y, si $\mathcal{X} = \mathbb{R}^d$, (6) $P(X_n \leqq x) \to P(X \leqq x)$ para todos los $x \in \mathbb{R}^d$ tales que F_X es continua en x.

Proof. (1) $\Rightarrow$ (2) es obvio. Veamos (2) $\Rightarrow$ (3). Sea G abierto y $f_m(x) = \max\{m d(x, G^c), 1\}$ para $m \in \mathbb{N}$; f_m son Lipschitz, acotadas y $0 \leqq f_m \nearrow \mathbb{1}_G$; luego $P(X_n \in G) \geqq \mathbb{E} f_m(X_n) \to \mathbb{E} f_m(X)$ y $\underline{\lim}\, P(X_n \in G) \geqq \mathbb{E} f_m(X) \to P(X \in G)$. Que (3) $\Leftrightarrow$ (4) es obvio. Que (4) $\Rightarrow$ (5) sale porque $P(X \in E^\circ) \leqq \underline{\lim}\, P(X_n \in E^\circ) \leqq \overline{\lim}\, P(X_n \in \overline{E}) \leqq P(X \in \overline{E})$.

Veamos que (5) $\Rightarrow$ (1). Sea $f : \mathcal{X} \to \mathbb{R}$ continua y acotada; podemos asumir $f \geqq 0$. El conjunto $Q = \{a \in \mathbb{R} \mid P(f(X) = a) > 0\}$ es a lo sumo numerable (si no $\{a \in \mathbb{R} \mid P(f(X) = a) > \frac{1}{n}\}$ sería infinito para agún $n \in \mathbb{N}$, imposible); entonces $A = \mathbb{R} \smallsetminus Q$ es denso; como f es acotada, $|f| < M$ (con $M \in A$), y, para cada $m \in \mathbb{N}$, hay particiones $\pi_m = \{0 = a_0 < a_1 < \ldots < a_m = M\} \subset A$ tales que $\|\pi_m\| = \max_{i=1}^{m}\{a_i - a_{i-1}\} \to 0$; sea $f_m(x) = \sum_{i=1}^m a_{i-1} \mathbb{1}_{[a_{i-1}, a_i)}(f(x))$; tenemos $\mathbb{E} f(X_n) - \mathbb{E} f_m(X_n) = \mathbb{E}(\sum_{i=1}^m (f(X_n) - a_{i-1}) \mathbb{1}_{[a_{i-1}, a_i)}(f(X_n))) \leqq \|\pi_m\|$. Ahora $\mathbb{E} f_m(X_n) = \sum_{i=1}^n a_{i-1} P(f(X_n) \in [a_{i-1}, a_i)) \to \sum_{i=1}^n a_{i-1} P(f(X) \in [a_{i-1}, a_i)) = \mathbb{E} f_m(X)$ por (5), ya que $\partial f^{-1}[a_{i-1}, a_i) \subset \{f = a_{i-1}\} \cup \{f = a_i\}$ y $P(X \in \partial f^{-1}[a_{i-1}, a_i)) = 0$. Con esto tenemos (1).

Ahora (5) $\Rightarrow$ (6) es obvio. Veamos que (6) $\Rightarrow$ (1). Notar que F_X es continua en a sii $P(X \in \partial\{x \mid x \leqq a\}) = 0$, y esto se da si $P(X_i = a_i) = 0$ para todo $i = 1, \ldots, d$. Hay pues densos $A_1, \ldots, A_d \subset \mathbb{R}$ tales que si $a < b \in \mathbb{R}^d$ con $a_i, b_i \in A_i$ entonces $P(X_n \in I) \to P(X \in I)$ donde $I = \prod_{i=1}^d (a_i, b_i]$; sea $\mathcal{I}$ el conjunto de dichos intervalos. Sea $f : \mathbb{R}^d \to \mathbb{R}$ continua y acotada, $|f(x)| \leqq M$, y $\epsilon > 0$. Sea $I \in \mathcal{I}$ grande de manera que $P(X \in I) \geqq 1 - \epsilon$, y n_0 de manera que $P(X_n \in I) \geqq 1 - 2\epsilon$ si $n \geqq n_0$; como f es absolutamente continua en $\overline{I}$, sea $\delta > 0$ tal que si $x, y \in I$, $\|x - y\| < \delta$, entonces $|f(x) - f(y)| < \epsilon$. Parto I en intervalitos $I_1, \ldots, I_m \in \mathcal{I}$ con $\max_i \operatorname{diam}(I_i) \leqq \delta$; elijo $x_i^* \in I_i$ y pongo $f_m(x) = \sum_{i=1}^m f(x_i^*) \mathbb{1}_{I_i}(x)$; entonces $|\mathbb{E} f(X_n) - \mathbb{E} f_m(X_n)| \leqq \sum_{i=1}^m \sup_{x \in I_i} |f(x) - f(x_i^*)| P(X_n \in I_i) + M P(X_n \notin I) \leqq (2M + 1)\epsilon$. Ahora $\mathbb{E} f_m(X_n) = \sum_{i=1}^m f(x_i^*) P(X_n \in I_i) \to \sum_{i=1}^m f(x_i^*) P(X \in I) = \mathbb{E} f_m(X)$. Es fácil ver ahora que $\mathbb{E} f(X_n) \to \mathbb{E} f(X)$. ∎

Se ve que $X_n \xrightarrow{p} X$ implica $X_n \xrightarrow{d} X$. Si $X = c$, una constante, vale el recíproco. Si $X_n \xrightarrow{d} X$ y $Y_n \xrightarrow{d} c$, $(X_n, Y_n) \xrightarrow{d} (X, c)$. Además, si $X_n \xrightarrow{d} X$ y $f : \mathcal{X} \to \mathcal{Y}$ es continua, $f(X_n) \xrightarrow{d} f(X)$. Por lo tanto tenemos (teorema de Slutzky) que si $X_n \xrightarrow{d} X$ y $Y_n \xrightarrow{d} c$ entonces $X_n + Y_n \xrightarrow{d} X + c$, $X_n Y_n \xrightarrow{d} cX$ y (si $c \neq 0$) $X_n / Y_n \xrightarrow{d} X_n / c$. Si $X_n \xrightarrow{d} X$ y $\|X_n - Y_n\| \xrightarrow{d} 0$ entonces $Y_n \xrightarrow{d} X$. Si $X_n \xrightarrow{d} X$ entonces $X_n = O_P(1)$, Dada toda esta teoría llegamos finalmente a lo interesante, que son los teoremas básicos de convergencia. Los doy sin demostración.

Theorem 5.1.2 (Ley fuerte de los grandes números)**.** Sean $X_1, X_2, \ldots$ variables aleatorias iid (independientes e igualmente distribuidas) con $\mathbb{E}|X_1| < \infty$. Entonces

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{as} \mathbb{E}(X_1).$$

Theorem 5.1.3 (Teorema central del límite)**.** Sean $X_i \overset{\text{iid}}{\sim} X$ para cada $i \in \mathbb{N}$ con $\mathbb{E}(X) = \mu$ y $V(X) = \Sigma$ inversible. Entonces

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right) \xrightarrow{d} N(0, \Sigma).$$

De la ley fuerte de los grandes números podemos deducir la siguiente ley de los grandes números uniforme.

Theorem 5.1.4. Sean, para cada $i \in \mathbb{N}$, $z_i \sim z \in \mathbb{R}^d$ iid, $\Theta \subset \mathbb{R}^p$ compacto, $a : \mathbb{R}^d \times \Theta \to \mathbb{R}^k$ una función tal que, para todo $\theta \in \Theta$, $a(z, \theta)$ es medible, $P\Big(a(z,\theta) = \lim_{\theta' \to \theta} a(z, \theta')\Big) = 1$, y $\|a(z, \theta)\| \leqq d(z)$, con $\mathbb{E}[d(z)] < \infty$. Entonces $\mathbb{E}[a(z, \cdot)]$ es continua y

$$\sup_{\theta \in \Theta} \left\| \frac{1}{n}\sum_{i=1}^{n} a(z_i, \theta) - \mathbb{E}[a(z, \theta)] \right\| \xrightarrow{p} 0.$$

Proof. La continuidad de $\mathbb{E}[a(z, \cdot)]$ es evidente en virtud del teorema de convergencia dominada y las hipótesis. Sea $\epsilon > 0$. Definimos $u : \mathbb{R}^d \times \Theta \times \mathbb{R}_+ \to \mathbb{R}$ dada por $u(z, \theta, \delta) = \sup_{\|\theta' - \theta\| \leqq \delta} \|a(z, \theta') - a(z, \theta)\|$. Tenemos $\lim_{\delta \to 0} u(z, \theta, \delta) = 0$ con probabilidad 1 para todo $\theta \in \Theta$ por hipótesis. Luego por convergencia dominada tenemos $\mathbb{E}[u(z, \theta, \delta)] \leqq \epsilon$ si $\delta \leqq \delta(\theta)$. Tenemos $\Theta \subset \bigcup_{\theta \in \Theta} B_{\delta(\theta)}(\theta)$. Luego, por compacidad, hay $\theta_1, \ldots, \theta_m$ tales que $\Theta \subset \bigcup_{i=1}^{m} B_{\delta(\theta_i)}(\theta_i)$. Ahora sea $\theta \in \Theta$; tomo $k \in \{1, \ldots, m\}$ tal que $\theta \in B_{\delta(\theta_k)}(\theta_k)$; tenemos

$$\begin{aligned} \left\| \frac{1}{n}\sum_{i=1}^{n} a(z_i, \theta) - \mathbb{E}[a(z, \theta)] \right\| &\leqq \left\| \frac{1}{n}\sum_{i=1}^{n} a(z_i, \theta) - \frac{1}{n}\sum_{i=1}^{n} a(z_i, \theta_k) \right\| + \\ &+ \left\| \frac{1}{n}\sum_{i=1}^{n} a(z_i, \theta_k) - \mathbb{E}[a(z, \theta_k)] \right\| + \|\mathbb{E}[a(z, \theta_k)] - \mathbb{E}[a(z, \theta)]\|. \end{aligned}$$

Ahora $\left\|\frac{1}{n}\sum_{i=1}^{n} a(z_i, \theta) - \frac{1}{n}\sum_{i=1}^{n} a(z_i, \theta_k)\right\| \leqq \frac{1}{n}\sum_{i=1}^{n} u(z_i, \theta_k, \delta(\theta_k)) \leqq \mathbb{E}[u(z, \theta_k, \delta(\theta_k)] + \epsilon \leqq 2\epsilon$ con probabilidad 1 si $n \geqq n(k)$ por la ley fuerte de los grandes números. También $\|\frac{1}{n}\sum_{i=1}^{n} a(z_i, \theta_k) - \mathbb{E}[a(z, \theta_k)]\| \leqq \epsilon$ con probabilidad 1 si $n \geqq n(k)'$ por la ley fuerte de los grandes números. Y $\|\mathbb{E}[a(z, \theta_k)] - \mathbb{E}[a(z, \theta)]\| \leqq \mathbb{E}\|a(z, \theta_k) - a(z, \theta)\| \leqq \mathbb{E}[u(z, \theta, \delta(\theta_k)] \leqq \epsilon$. Entonces tenemos que, para todo $\theta \in \Theta$,

$$\left\| \frac{1}{n}\sum_{i=1}^{n} a(z_i, \theta) - \mathbb{E}[a(z, \theta)] \right\| \leqq 4\epsilon$$

con probabilidad 1 si $n \geqq \max_{k \in \{1, \ldots, m\}} \{n(k), n(k)'\}$. Esto prueba la aserción. ∎

5.2 Extremum estimators

Empezamos a hacer estadística. El setup básico es el siguiente. Tenemos un *modelo paramétrico*, esto es, una familia de distribuciones $\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$ sobre $\mathbb{R}^d$ determinadas por un parámetro $\theta \in \Theta \subset \mathbb{R}^p$. La naturaleza elige $\theta_0 \in \Theta$, el parámetro verdadero, que desconocemos, y produce variables $X_i \overset{\text{iid}}{\sim} P_{\theta_0}$ para cada $i \in \mathbb{N}$. Nuestro objetivo es inferir θ_0 en base a la observación de las realizaciones de $X_1, \dots, X_n$. Para eso construimos un *estimador* $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$. Decimos que el estimador es *insesgado* si $\mathbb{E}[\hat{\theta}_n] = \theta_0$, *consistente* si $\hat{\theta}_n \xrightarrow{p} \theta_0$, y *asintóticamente normal* si $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \Sigma)$.

Nos vamos a concentrar en estimadores de un determinado tipo, llamados *extremum estimators*, que se obtienen como sigue. Para cada $n \in \mathbb{N}$ y cada $\theta \in \Theta$ tenemos una variable aleatoria $\hat{Q}_n(\theta) = \hat{Q}_n(X_1, \dots, X_n, \theta)$. Decimos que $\hat{\theta}_n$ es un estimator extremal de $\hat{Q}_n$ si $\hat{\theta}_n \in \arg\max_{\theta \in \Theta} \hat{Q}_n(\theta)$. Tenemos los siguientes teoremas de consistencia para estos estimadores.

Theorem 5.2.1. Sea $\hat{\theta}_n$ un estimador extremal de $\hat{Q}_n$, y $Q_0 : \Theta \to \mathbb{R}$. Si (1) Q_0 tiene un máximo único en θ_0, (2) Θ es compacto, (3) Q_0 es continua, y (4) $\sup_{\theta \in \Theta} |\hat{Q}_n(\theta) - Q_0(\theta)| \xrightarrow{p} 0$ entonces $\hat{\theta}_n$ es consistente.

Proof. Sean $\epsilon, \delta > 0$. Queremos ver que $P(\|\hat{\theta}_n - \theta_0\| \geqq \delta) < \epsilon$ si n es suficientemente grande. Sean $M = Q_0(\theta_0)$ y $m < M$ el máximo de Q_0 en $\Theta \setminus B_\delta(\theta_0)$, que existe porque Q_0 es continua y Θ es compacto. Tenemos que $P\left(\sup_{\theta \in \Theta} |\hat{Q}_n(\theta) - Q_0(\theta)| < \frac{M-m}{2}\right) > 1 - \epsilon$ si $n \geqq n_0$ por hipótesis. Ahora $\sup_{\theta \in \Theta} |\hat{Q}_n(\theta) - Q_0(\theta)| < \frac{M-m}{2}$ implica $\hat{\theta}_n \in B_\delta(\theta_0)$; en efecto, supongamos que no; entonces $\hat{Q}_n(\hat{\theta}_n) < Q_0(\hat{\theta}_n) + \frac{M-m}{2} \leqq m - M + Q_0(\theta_0) + \frac{M-m}{2} < m - M + \hat{Q}_n(\theta_0) + \frac{M-m}{2} + \frac{M-m}{2} = \hat{Q}_n(\theta_0)$, y $\hat{\theta}_n$ no maximiza $\hat{Q}_n$, absurdo. Entonces $P\left(\sup_{\theta \in \Theta} |\hat{Q}_n(\theta) - Q_0(\theta)| < \frac{M-m}{2}\right) \leqq P\left(\hat{\theta}_n \in B_\delta(\theta_0)\right)$, y obtenemos $P(\|\hat{\theta}_n - \theta_0\| \geqq \delta) < \epsilon$ si $n \geqq n_0$, como queríamos. ∎

Theorem 5.2.2. Sea $\hat{\theta}_n$ un estimador extremal de $\hat{Q}_n$, y $Q_0 : \Theta \to \mathbb{R}$. Si (1) Q_0 tiene un máximo único en θ_0, (2) θ_0 está en el interior del convexo Θ y $\hat{Q}_n$ es cóncava, y (3) $\hat{Q}_n(\theta) \xrightarrow{p} Q_0(\theta)$ para todo $\theta \in \Theta$, entonces $\hat{\theta}_n$ es consistente.

Proof. Como $\hat{Q}_n$ son cóncavas y $\hat{Q}_n(\theta) \xrightarrow{p} Q_0(\theta)$ para todo $\theta \in \Theta$, se ve que Q_0 es cóncava. Sea $K = \overline{B_\epsilon(\theta_0)}$, con $\epsilon > 0$ tal que $B_{2\epsilon}(x_0) \subset \Theta$. Por un argumento muy similar al del teorema 1.2.3, tenemos que $\sup_{\theta \in K} |\hat{Q}_n(\theta) - Q_0(\theta)| \xrightarrow{p} 0$. Podemos aplicar el teorema anterior sobre K (en lugar de Θ), lo que nos da que si $\tilde{\theta}_n \in \{\theta \in \arg\max_{\theta \in K} \hat{Q}_n(\theta) \mid \|\theta - \theta_0\| \text{ es máximo}\}$, $\tilde{\theta}_n \xrightarrow{p} \theta_0$. Faltaría ver que $\|\hat{\theta}_n - \tilde{\theta}_n\| \xrightarrow{p} 0$. Sea $\delta > 0$. Si $\|\hat{\theta}_n - \tilde{\theta}_n\| \geqq \delta$, esto implica que, si $\tilde{\theta}_n \in B_{\delta/2}(\theta_0)$, $\|\hat{\theta}_n - \theta_0\| > \frac{\delta}{2}$. Ahora $\tilde{\theta}_n \in B_{\delta/2}(\theta_0)$ implica que $\hat{Q}_n(\tilde{\theta}_n) > \max_{\theta \in \partial B_{\delta/2}(\theta_0)} \hat{Q}_n(\theta)$ (por cómo está elegido $\tilde{\theta}_n$). Dado que $\|\tilde{\theta}_n - \theta_0\| < \frac{\delta}{2}$ y $\|\hat{\theta}_n - \theta_0\| > \frac{\delta}{2}$, existe $\lambda \in (0,1)$ tal que $\bar{\theta}_n = \lambda \tilde{\theta}_n + (1-\lambda)\hat{\theta}_n \in \partial B_{\delta/2}(\theta_0)$. Entonces $\hat{Q}_n(\bar{\theta}_n) \geqq \lambda \hat{Q}_n(\tilde{\theta}_n) + (1-\lambda)\hat{Q}_n(\hat{\theta}_n) > \lambda Q_n(\bar{\theta}_n) + (1-\lambda)\hat{Q}_n(\hat{\theta}_n)$, por lo que $\hat{Q}_n(\bar{\theta}_n) > Q_n(\hat{\theta}_n)$, absurdo. Luego $P(\|\hat{\theta}_n - \tilde{\theta}_n\| \geqq \delta) \leqq P(\tilde{\theta}_n \notin B_{\delta/2}(\theta_0)) \to 0$, listo. ∎

Tenemos el siguiente teorema de normalidad asintótica.

Theorem 5.2.3. Sea $\hat{\theta}_n$ un estimador extremal de $\hat{Q}_n$ con $\hat{\theta}_n \xrightarrow{p} \theta_0$, y (1) $\theta_0 \in \Theta^\circ$, (2) $\hat{Q}_n$ es dos veces derivable en un entorno de $\mathcal{U}$ de θ_0, (3) $\sqrt{n}\nabla \hat{Q}_n(\theta_0) \xrightarrow{d} N(0, \Sigma)$, (4) hay $H : \mathcal{U} \to \mathbb{R}^{p \times p}$ continua en θ_0 tal que $\sup_{\theta \in \mathcal{U}} \|\nabla^2 \hat{Q}_n(\theta) - H(\theta)\| \xrightarrow{p} 0$, y (5) $H = H(\theta_0)$ es inversible. Entonces

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, H^{-1}\Sigma H^{-1}).$$

Proof. Sea $\hat{\imath}_n = \mathbb{1}[\hat{\theta}_n \in \mathcal{U}]$; como $\hat{\theta}_n \xrightarrow{p} \theta_0$, $\hat{\imath}_n \xrightarrow{p} 1$. Si $\hat{\theta}_n \in \mathcal{U}$, $\nabla \hat{Q}_n(\hat{\theta}_n) = 0$, por lo que $\hat{\imath}_n \nabla \hat{Q}_n(\hat{\theta}_n) = 0$. Además, para cada $i = 1, \ldots, p$, hay $\bar{\theta}_{ni} \in \mathcal{U}$ tal que $\nabla \hat{Q}_n(\hat{\theta}_n)_i = \nabla \hat{Q}_n(\theta_0)_i + \nabla^2 \hat{Q}_n(\bar{\theta}_{ni})_i^\top (\hat{\theta}_n - \theta_0)$. Si defino $\bar{\theta}_{ni} = \theta_0$ cuando $\hat{\theta}_n \notin \mathcal{U}$ obtengo $\hat{\imath}_n \nabla^2 \hat{Q}_n(\bar{\theta}_{ni})_i^\top (\hat{\theta}_n - \theta_0) = -\hat{\imath}_n \nabla \hat{Q}_n(\theta_0)_i$ incondicionalmente. Sea $\bar{H}_n \in \mathbb{R}^{p \times p}$ tal que su fila i es el vector $\nabla^2 \hat{Q}_n(\bar{\theta}_{ni})_i^\top$. Por (4) tenemos $\bar{H}_n \xrightarrow{p} H$. Sea $\bar{\imath}_n = \mathbb{1}[\hat{\theta}_n \in \mathcal{U}$ y $\bar{H}_n$ es inversible]. Por (5) vale que $\bar{\imath}_n \xrightarrow{p} 1$ y $\bar{\imath}_n \bar{H}_n^{-1} \xrightarrow{p} H^{-1}$. Obtengo $\bar{\imath}_n(\hat{\theta}_n - \theta_0) = -\bar{\imath}_n \bar{H}_n^{-1} \nabla \hat{Q}_n(\theta_0)$ y

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -\bar{\imath}_n \bar{H}_n^{-1} \sqrt{n} \nabla \hat{Q}_n(\theta_0) + (1 - \bar{\imath}_n)\sqrt{n}(\hat{\theta}_n - \theta_0),$$

con lo que por (3) y Slutzky se obtiene el resutado. ■

Aplicación (estimadores de máxima verosimilitud) Asumimos que las distribuciones P_θ son absolutamente continuas sobre una medida μ de $\mathbb{R}^d$, con densidad $p(\cdot, \theta)$. Definimos $\hat{Q}_n(\theta) = \frac{1}{n}\sum_{i=1}^n \ln p(X_i, \theta)$, y $Q_0(\theta) = \mathbb{E}[\ln p(X, \theta)]$. Tenemos que Q_0 se maximiza en θ_0: $\mathbb{E}[\ln p(X, \theta)] - \mathbb{E}[\ln(X, \theta_0)] = \mathbb{E}[\ln \frac{p(X,\theta)}{p(X,\theta_0)}] \leqq \ln \mathbb{E}[\frac{p(X,\theta)}{p(X,\theta_0)}] = \ln(\int \frac{p(X,\theta)}{p(X,\theta_0)} p(X, \theta_0)\, d\mu) = 0$ por Jensen, con igualdad sii $p(X, \theta) = p(X, \theta_0)$ en casi todo punto, i.e., sii $P_\theta = P_{\theta_0}$. Para probar la condición (4) del teorema 5.2.1 podemos usar el teorema 5.1.4. Tenemos, pues, el siguiente teorema de consistencia.

Theorem 5.2.4. Si (1) $\mathcal{P}(\theta) \neq \mathcal{P}(\theta_0)$ para todo $\theta \neq \theta_0$, y (2) Θ es compacto, $p(X, \cdot)$ es continua con probabilidad 1 y $\mathbb{E}[\sup_{\theta \in \Theta} |\ln p(X, \theta)|] < \infty$, o (2′) $\theta_0 \in \Theta^\circ$, Θ es convexo, $p(X, \cdot)$ es cóncava y $\mathbb{E}[|\ln p(X, \theta)|] < \infty$ para todo $\theta \in \Theta$, entonces $\hat{\theta}_n \xrightarrow{p} \theta_0$, donde $\hat{\theta}_n$ es el estimador de máxima verosimilitud.

Asumimos que $p(X, \cdot)$ es dos veces derivable. Tenemos $\nabla \hat{Q}_n(\theta_0) = \frac{1}{n}\sum_{i=1}^n \frac{\nabla_\theta p(X_i, \theta_0)}{p(X_i, \theta_0)}$. Si existe $\epsilon > 0$ tal que $\int \sup_{\theta \in B_\epsilon(\theta_0)} \|\nabla_\theta p(\cdot, \theta)\|\, d\mu < \infty$ entonces[4] $\mathbb{E}[\frac{\nabla_\theta p(X,\theta_0)}{p(X,\theta_0)}] = \int \nabla_\theta p(-, \theta_0)\, d\mu = \nabla_\theta \int p(-, \theta_0)\, d\mu = 0$, y $\Sigma = V(\frac{\nabla_\theta p(X,\theta_0)}{p(X,\theta_0)}) = \mathbb{E}[\frac{\nabla_\theta p(X,\theta_0) \nabla_\theta p(X,\theta_0)^\top}{p(X,\theta_0)^2}]$. Si Σ es definida positiva, por el teorema central del límite tenemos $\sqrt{n}\nabla \hat{Q}_n(\theta_0) \xrightarrow{d} N(0, \Sigma)$. Sea $H(\theta) = \mathbb{E}[\nabla_{\theta\theta} \ln p(X, \theta)]$; si $\nabla_{\theta\theta} p(X, \theta)$ es continua en un entorno $\mathcal{U}$ de θ_0 con probabilidad 1 y $\mathbb{E}[\sup_{\theta \in \mathcal{U}} \|\nabla_{\theta\theta} \ln p(X, \theta)\|] < \infty$ entonces por el teorema 5.1.4 $\sup_{\theta \in \mathcal{U}} \|\nabla^2 \hat{Q}_n(\theta) - H(\theta)\| \xrightarrow{p} 0$. Ahora $H = H(\theta_0) = -\Sigma + \mathbb{E}[\frac{\nabla_{\theta\theta} p(X,\theta)}{p(X,\theta)}] = -\Sigma$ si $\int \sup_{\theta \in \mathcal{U}} \|\nabla_{\theta\theta} p(\cdot, \theta)\|\, d\mu < \infty$. Tenemos, pues, el siguiente teorema de normalidad asintótica.

Theorem 5.2.5. Si $\hat{\theta}_n$ es un estimador de máxima verosimilitud consistente, (1) $\theta_0 \in \Theta^\circ$, (2) $p(X, \cdot)$ es C^2 en un entorno $\mathcal{U}$ de θ_0 con probabilidad 1, (3) $\mathcal{I} = \mathbb{E}[\frac{\nabla_\theta p(X,\theta_0) \nabla_\theta p(X,\theta_0)^\top}{p(X,\theta_0)^2}]$ es inversible, (4) $\int \sup_{\theta \in \mathcal{U}} \|\nabla_\theta p(\cdot, \theta)\|\, d\mu < \infty$, $\int \sup_{\theta \in \mathcal{U}} \|\nabla_{\theta\theta} p(\cdot, \theta)\|\, d\mu < \infty$ y $\mathbb{E}[\sup_{\theta \in \mathcal{U}} \|\nabla_{\theta\theta} \ln p(X, \theta)\|] < \infty$, entonces

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \mathcal{I}^{-1}).$$

Vamos a probar la cota de Cramér–Rao, lo que nos va a permitir afirmar que el estimador de máxima verosimilitud es asintóticamente eficiente bajo condiciones de regularidad.

Proposition 5.2.6. Sean x, y dos variables aleatorias, $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, con $V(y)$ inversible. Entonces $V(x) \geqq \mathrm{Cov}(x, y) V(y)^{-1} \mathrm{Cov}(y, x)$.

[4] Usamos el siguiente resultado, que se prueba fácilmente. Sean $U \subset \mathbb{R}^n$ abierto, $V \subset \mathbb{R}^m$ medible, $f : U \times V \to \mathbb{R}$ medible y $x_0 \in U$. Si $f(x, -) \in L^1(V)$ para todo $x \in B_\epsilon(x_0)$, $f(-, y)$ es diferenciable en $B_\epsilon(x_0)$ para casi todo $y \in V$ y existe $g \in L^1(V)$ tal que $|\partial_j f(x, y)| \leqq g(y)$ para todo $x \in B_\epsilon(x_0)$ y casi todo $y \in V$, entonces si $F(x) = \int_V f(x, -)$, $\partial_j F(x) = \int_V \partial_j (x, -)$.

Proof. Restando las medias lo que hay que probar es que $\mathbb{E}(xx^t) \geqq \mathbb{E}(xy^t)\mathbb{E}(yy^t)^{-1}\mathbb{E}(yx^t)$. Sean $a \in \mathbb{R}^n$ y $b \in \mathbb{R}^m$. Por Cauchy-Schwarz tenemos $\mathbb{E}(a^t xy^t b)^2 \leqq \mathbb{E}(a^t xx^t a)\mathbb{E}(b^t yy^t b)$, es decir, $(a^t\mathbb{E}(xy^t)b)^2 \leqq a^t\mathbb{E}(xx^t)ab^t\mathbb{E}(yy^t)b$. Pongo $b = \mathbb{E}(yy^t)^{-1}\mathbb{E}(yx^t)a$ y estamos. ∎

Theorem 5.2.7 (Cramér–Rao)**.** Sea $\hat{\theta}_n$ un estimador insesgado para todo $\theta_0 \in \mathcal{U}$, con $\mathcal{U} \subset \Theta$ abierto. Sea $\theta_0 \in \mathcal{U}$ y $\mathcal{I} = \mathbb{E}[\frac{\nabla_\theta p(X,\theta_0)\nabla_\theta p(X,\theta_0)^\top}{p(X,\theta_0)^2}]$. Si $\mathcal{I}$ es inversible y $\int \sup_{\theta\in\mathcal{U}} \|\hat{\theta}_n(\cdot)\nabla_\theta(\cdot,\theta)^\top\| < \infty$ entonces $V(\hat{\theta}_n) \geqq \frac{1}{n}\mathcal{I}^{-1}$.

Proof. Llamemos X^n a $(X_1,\ldots,X_n)$. Sea $s_n = \frac{1}{n}\nabla_\theta \ln p(X^n,\theta_0) = \frac{1}{n}\sum_{i=1}^n \frac{\nabla_\theta p(X_i,\theta_0)}{p(X_i,\theta_0)}$. Tenemos $\mathbb{E}(s_n) = 0$ y $V(s_n) = \frac{1}{n}\mathcal{I}$. Ahora $\mathbb{E}(\hat{\theta}_n s_n^\top) = \frac{1}{n}\mathbb{E}[\hat{\theta}_n \frac{\nabla_\theta p(X^n,\theta_0)^\top}{p(X^n,\theta_0)}] = \frac{1}{n}\int \hat{\theta}_n(\cdot)\nabla_\theta p(\cdot,\theta_0)^\top$ $= \frac{1}{n}\nabla_\theta|_{\theta=\theta_0}\int \hat{\theta}_n(\cdot)p(\cdot,\theta) = \frac{1}{n}\nabla_\theta|_{\theta=\theta_0}\mathbb{E}_\theta[\hat{\theta}_n] = \frac{1}{n}\nabla_\theta|_{\theta=\theta_0}\theta = \frac{1}{n}I_n$. Esto más la proposición anterior dan el resultado. ∎

Por último, presento el *algoritmo EM*, que es útil para encontrar estimadores de máxima verosimilitud. El setting es: tenemos un modelo $\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$, $(x,y) \sim P_\theta$ tienen densidad $p(x,y|\theta)$ y observamos x. Buscamos pues $\hat{\theta} \in \arg\max_{\theta\in\Theta} p(x|\theta)$, donde $p(x|\theta) = \int p(x,y|\theta)\,dy$. La idea es encontrar iterativamente $\hat{\theta}_1, \hat{\theta}_2, \ldots$ con $p(x|\hat{\theta}_n) \leqq p(x|\hat{\theta}_{n+1})$ para todo $n \in \mathbb{N}$, e idealmente $\hat{\theta}_n \to \hat{\theta}$. Para ello dado $\hat{\theta}_n$ calculamos (paso E) $Q(\theta) = \mathbb{E}[\ln p(x,y|\theta)|x,\hat{\theta}_n]$ y obtenemos (paso M) $\hat{\theta}_{n+1} \in \arg\max_{\theta\in\Theta} Q(\theta)$. Veamos que $p(x|\hat{\theta}_n) \leqq p(x|\hat{\theta}_{n+1})$. En efecto,

$$\begin{aligned}
\ln p(x|\hat{\theta}_{n+1}) - \ln p(x|\hat{\theta}_n) &= \ln\left(\frac{p(x|\hat{\theta}_{n+1})}{p(x|\hat{\theta}_n)}\right) = \ln\left(\int \frac{p(x,y|\hat{\theta}_{n+1})}{p(x|\hat{\theta}_n)}dy\right) = \\
&= \ln\left(\int \frac{p(x,y|\hat{\theta}_{n+1})}{p(x,y|\hat{\theta}_n)}p(y|x,\hat{\theta}_n)dy\right) \geqq \int \ln\left(\frac{p(x,y|\hat{\theta}_{n+1})}{p(x,y|\hat{\theta}_n)}\right)p(y|x,\hat{\theta}_n)dy = \\
&= Q(\hat{\theta}_{n+1}) - Q(\hat{\theta}_n) \geqq 0.
\end{aligned}$$

5.3 Add

- Causal inference / 2SLS
- Bayesian estimators
- Exponential distributions / sufficient statistics
- Bootstrap
- Lasso
- Semiparametric models
- Identification tests (?)
- Model selection
- GMM
- Time series
- Duration models

References

Dixit, Avinash, Gene M. Grossman, and Elhanan Helpman, "Common Agency and Coordination: General Theory and Application to Government Policy Making," *Journal of Political Economy*, August 1997, *105* (4), 752–769.

Hart, Sergiu and Andreu Mas-Colell, "Bargaining and Value," *Econometrica*, 1996, *64* (2), 357–380.

Meltzer, Allan H. and Scott F. Richard, "A Rational Theory of the Size of Government," *Journal of Political Economy*, October 1981, *89* (5), 914–927.

Moene, Karl Ove and Michael Wallerstein, "Inequality, Social Insurance, and Redistribution," *American Political Science Review*, 2001, *95* (4), 16.

Rehm, Philipp, "Social Policy by Popular Demand," *World Politics*, April 2011, *63* (2), 271–299.

Topkis, Donald M., *Supermodularity and Complementarity* Frontiers of Economic Research, Princeton, N.J: Princeton University Press, 1998.