FACULTADE DE MATEMÁTICAS

Traballo Fin de Grao

Elección Social: Teoremas de Imposibilidad

Jorge Vázquez Pérez

07/2022

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

GRAO DE MATEMÁTICAS

Trabajo Fin de Grao

Elección Social: Teoremas de Imposibilidad

Jorge Vázquez Pérez

08/2022

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Trabajo propuesto

Área de Conocimiento: Estadística e Investigación Operativa
Título: Elección Social: Teoremas de Imposibilidad
Breve descrición del contenido
El objetivo de este trabajo es realizar una introducción a las matemáticas del campo de la elección social, con especial énfasis en los conocidos como teoremas de imposibilidad iniciados por K. Arrow, quien fue galardonado con el Premio Nobel de Economía en 1972 por esta y otras contribuciones a la teoría económica. El Teorema de Arrow versa sobre la imposibilidad de definir funciones que, a la hora de escoger como sociedad entre tres o más alternativas, permitan agregar adecuadamente las preferencias de los individuos que componen dicha sociedad.
Recomendaciones
No hay recomendaciones específicas.
Otras observaciones
No hay observaciones específicas.

Índice

Resumen

Este trabajo trata de desarrollar una introducción a los conceptos y los resultados más importantes de la teoría de la elección social. Para ello, se unifica la notación de varias fuentes bibliográficas y se definen los conceptos básicos relacionados con la elección social. Después de esto, se presenta el primero de los resultados: el Teorema de Imposibilidad de Arrow. Después de éste, se presenta el segundo resultado: el Teorema de Imposibilidad de Gibbard-Satterthwaite. Una vez presentados los conceptos y a los resultados, con el fin de completar el trabajo, se desarrolla un estudio práctico para visualizar el resultado obtenido en el Teorema de Imposibilidad de Arrow.

Abstract

This paper attempts to develop an introduction to the most important concepts and results of social choice theory. For this purpose, the notation of several bibliographical sources is unified and the basic concepts related to social choice are defined. After this, the first result is presented: Arrow's Impossibility Theorem. This is followed by the second result: the Gibbard-Satterthwaite Impossibility Theorem. Once the concepts and results have been presented, in order to complete the work, a practical study is carried out to visualise the result obtained in Arrow's Impossibility Theorem.

Introducción

Este trabajo trata de realizar una introducción al campo de la elección social y un repaso a los resultados de imposibilidad más importantes que hay, entre los cuales encontramos el Teorema de Arrow y el Teorema de Gibbard-Satterthwaite.

En el primer capítulo de esta memoria trataremos de definir los conceptos más utilizados en el resto de capítulos aunando un cierto número de referencias bibliográficas y estandarizando la notación. Además, definimos una serie de reglas de clasificación, también conocidas como funciones de bienestar. Para concluir primer capítulo introductorio, completamos los conceptos restantes haciendo un breve repaso histórico a la teoría de la elección social utilizando como referencia el artículo [10].

En el segundo capítulo, entramos de lleno en los resultados de elección social. Aquí, tratamos de construir una función de bienestar social deseable empleando una serie de condiciones para, finalmente, observar que la existencia de esta función es imposible. Para este capítulo intentamos explicar cómo se pasa del planteamiento original de Arrow [1] a enunciados más modernos como son los de Mas-Colell [13].

En el tercer capítulo, tratamos otro teorema de imposibilidad muy importante: el de Gibbard-Satterthwaite. En este caso, en lugar de una función de bienestar, construimos una función de elección social con una serie de características deseables. De nuevo, combinamos el planteamiento original de Gibbard [11] y Satterthwaite [18] con planteamientos más modernos como es, otra vez, el de Mas-Colell [13].

Por último, en el cuarto capítulo, trataremos de llevar a cabo los experimentos realizados en el artículo [15] para poder observar, en la práctica, cuánto fallan realmente las reglas de clasificación definidas en el primer capítulo cuando las probamos frente a las condiciones deseables del segundo capítulo.

Capítulo 1

Introducción a la elección social

La teoría de la elección social, iniciada en su forma moderna por Arrow [2], se ocupa de la relación entre los individuos y la sociedad. En particular, se ocupa de la agregación del interés, juicio o bienestar individual en una noción agregada de bienestar social, juicio social o elección social [20].

El propio Arrow explicaba que en una democracia capitalista existen esencialmente dos métodos para tomar decisiones sociales: la votación, que suele utilizarse para tomar decisiones "políticas", y el mecanismo de mercado, que suele utilizarse para tomar decisiones "económicas" [2]. En esta monografía, Arrow decidió que, en su trabajo, la distinción entre la votación y el mecanismo de mercado sería ignorada, siendo ambos considerados como casos especiales de la categoría más general de elección social. Además, también decidió que el estudio se atendría a los aspectos puramente formales de la elección social, es decir, Arrow evita introducirse en los aspectos propios de la teoría de juegos que se encuentran asociados de forma intrínseca a la economía y a la política. En este trabajo seguiremos estas mismas directrices.

Además, en el modelo clásico de elección social, hay un número finito de agentes que tienen preferencias sobre un número finito de alternativas. Estas preferencias se agregan en una preferencia social según la llamada función de bienestar social o función de agregación social (la cual definiremos formalmente más adelante), o bien dan lugar a una alternativa común según la llamada función de elección social o función de votación.

El objetivo principal de este trabajo es realizar una introducción a las matemáticas del campo de la elección social centrándonos en el teorema de Arrow y en el teorema de Gibbard-Satterthwaite. El primer teorema se aplica a las funciones de bienestar social y dice que, si la preferencia social entre dos alternativas cualesquiera sólo debe depender de las preferencias individuales entre estas alternativas y, por tanto, no de las preferencias individuales que implican otras alternativas, entonces la función de bienestar social debe ser dictatorial. El segundo teorema se aplica a las funciones de elección social y dice que las únicas funciones de elección social que son invulnerables a la manipulación estratégica son las dictatoriales. Estos resultados suelen denominarse “teoremas de imposibilidad”, ya que las dictaduras se consideran, generalmente, indeseables. Además, con el fin de profundizar en el resultado de Arrow, analizaremos el rendimiento de una serie de funciones de bienestar con respecto a una serie de condiciones deseables con un caso práctico.

Así pues, comenzaremos esta introducción a la elección social describiendo los conceptos matemáticos que trataremos en este trabajo junto con la notación empleada para representarlos.

1.1. Definiciones y notación

Hemos comentado al principio que la elección social se ocupa de la agregación de los intereses, juicios o bienestar individuales en una noción agregada de bienestar social, juicio social o elección social. Esto se puede simplificar diciendo que esta disciplina se ocupa de los procesos que permiten a un conjunto de individuos decidir sobre una determinada materia considerando las preferencias de cada individuo del conjunto.

Estas preferencias se definen sobre un conjunto de alternativas o candidatos entre las cuales se escoge. De forma general, consideraremos a las alternativas mutuamente excluyentes y las denotaremos con letras minúsculas x, y, $z \ldots$ Así pues, si denominamos S al conjunto de alternativas, entonces el proceso de selección general será, y así lo consideraremos de aquí en adelante, de la siguiente forma [2]:

1. Considerando todos los pares posibles formados por las alternativas, tomemos x e y, las relaciones que puede haber entre este par de alternativas son: x es preferida a y, x es indiferente a y o y es preferida a x.

2. Las decisiones tomadas para cada par se consideran consistentes entre sí, es decir, si se prefiere x a y y se prefiere y a z, entonces x se prefiere a z. De forma similar, si x es indiferente a y e y a z, entonces x es indiferente a z.

3. En caso de existir una alternativa preferida a todas las demás se seleccionará tal alternativa.

Una vez explicado esto, podemos ver que la preferencia y la indiferencia son relaciones entre las alternativas. Sin embargo, no es demasiado conveniente trabajar con dos relaciones, por lo que emplearemos la relación “preferido o indiferente”. Denotaremos la relación “x es preferida o indiferente a y” como $x \succsim y$. Las relaciones de preferencia e indiferencia derivadas de $\succsim$ se denotan por $\succ$ y $\sim$, respectivamente [13].

Por lo tanto, es importante, para la claridad de la exposición, tener un símbolo para el conjunto de todas las posibles relaciones racionales de preferencia o indiferencia en S y para el conjunto de todas las posibles relaciones de preferencia en S que tienen la propiedad de que no hay dos alternativas distintas indiferentes, es decir, para las posibles relaciones de preferencia estricta. Denotamos estos conjuntos, respectivamente, por $\mathcal{R}$ y $\mathcal{P}$. Obsérvese que $\mathcal{P} \subset \mathcal{R}$.

Todo esto que acabamos de escribir se puede resumir en una serie de axiomas tal y como lo hace Arrow [2].

Axioma 1.1. *Para todo $x, y \in S$, o $x \succsim y$ o bien $y \succsim x$.*

Una relación $\succsim$ que satisface el axioma 1.1 se dirá que es **conexa** o, lo que es lo mismo, que satisface el axioma de conexidad. Nótese que se supone que el axioma 1.1 se cumple tanto cuando $x = y$, como cuando $x \neq y$. En el primer caso decimos que x es indiferente a sí mismo para cualquier x, y esto implica que $x \succsim x$, lo cual significará que la relación $\succsim$ es **reflexiva**. Obsérvese también que la palabra "o" en el enunciado del axioma 1.1 no excluye la posibilidad de que suceda tanto $x \succsim y$ como $y \succsim x$. Esa palabra simplemente afirma que al menos uno de los dos eventos debe ocurrir; ambos casos pueden ocurrir.

Axioma 1.2. *Para todo $x, y, z \in S$, si se cumple que $x \succsim y$ e $y \succsim z$, entonces $x \succsim z$.*

Una relación $\succsim$ que satisface el axioma 1.2 se dice que es **transitiva** o, dicho de otro modo, que satisface el axioma de transitividad. En el caso de que cumpla ambos axiomas, 1.1 y 1.2, se denomina **ordenación débil** o, a veces, simplemente **ordenación**. Una relación $\succsim$ que tenga estas dos propiedades tomadas conjuntamente crea una clasificación de las distintas alternativas. El adjetivo "débil" se refiere al hecho de que la ordenación no excluye la indiferencia, es decir, los axiomas 1.1 y 1.2 no excluyen la posibilidad de que para algunas x e y distintas, tanto $x \succsim y$ como $y \succsim x$; como ya sucedía cuando explicábamos el axioma 1.1. Una ordenación fuerte, por otro lado, es una ordenación en la que no son posibles los empates. Una ordenación débil es una generalización del concepto "mayor o igual que" aplicado a los números reales; una ordenación fuerte generaliza el concepto "mayor que" aplicado al mismo ámbito. En esta memoria cuando hablemos simplemente de orden o de ordenación nos referiremos siempre a la ordenación débil, mientras que para referenciar específicamente a la ordenación fuerte hablaremos de preferencia estricta.

Puede parecer que los dos axiomas en cuestión no caracterizan completamente el concepto de preferencia. Por ejemplo, es natural pensar que, no sólo la relación $\succsim$, sino también las relaciones de preferencia (estricta) $\succ$ y de indiferencia $\sim$ son transitivas. Mostraremos que, definiendo adecuadamente la preferencia y la indiferencia en términos de $\succsim$, se obtendrán todas las propiedades deseadas.

Definición 1.3. $x \succ y$ se define como la negación de $y \succsim x$. Es decir, es la relación de preferencia estricta.

Definición 1.4. $x \sim y$ significa que $x \succsim y$ e $y \succsim x$. Es decir, es la relación de indiferencia.

Una vez establecidas todas las definiciones básicas, establezcamos las diferentes implicaciones que existen entre las relaciones.

Lema 1.5. *Se cumplen las siguientes implicaciones:*

1. *Para todo x, $x \succsim x$.*
2. *Si $x \succ y$, entonces $x \succsim y$.*
3. *Si $x \succ y$ e $y \succ z$, entonces $x \succ z$.*
4. *Si $x \sim y$ e $y \sim z$, entonces $x \sim z$.*
5. *Para todo x e y, o bien se cumple que $x \succsim y$ o bien que $y \succ x$.*
6. *Si $x \succ y$ e $y \succsim z$, entonces se cumple que $x \succsim z$.*

Demostración. Vemos que:

1. En el axioma 1.1, sea $y = x$; entonces se cumple que, para todo x, o bien $x \succsim x$ o bien $x \succsim x$, lo que es lo mismo que decir que $x \succsim x$.
2. Directa de la definición 1.3 y del 1.1.
3. Supongamos que $z \succsim x$. Si demostramos que esto no puede ocurrir, entonces, por la definición 1.3 tendremos que $x \succ z$. Por hipótesis, tenemos que $x \succ y$. Esto implica que $x \succsim y$ por el punto 2. Como hemos supuesto que $z \succsim x$ y sabemos que $x \succsim y$, entonces, por el axioma 1.2, $z \succsim y$. Pero sabemos, por hipótesis, que $y \succ z$, lo que significa que, tomando la definición 1.3, no puede ocurrir $z \succsim y$. Por lo tanto, llegamos a una contradicción y podemos concluir que $x \succ z$.
4. Si tenemos $x \sim y$ e $y \sim z$, sabemos que, por la definición 1.4, $x \succsim y$ e $y \succsim z$. Siguiendo el axioma 1.2 sabemos que $x \succsim z$. De la misma forma, por la definición 1.4, sabemos que $z \succsim y$ e $y \succsim x$, por lo que, por el axioma 1.2, tenemos que $z \succsim x$. Es decir, tenemos que $x \succsim z$ y $z \succsim x$, lo que, por la definición 1.4, significa que $x \sim z$.
5. Directamente de la definición 1.3.

6. Al igual que hicimos en el punto 3, supongamos que $z \succsim x$. De $z \succsim x$ e $y \succsim z$ se tiene que $y \succsim x$, por el axioma 1.2. Pero, si tomamos la definición 1.3, se tiene que $x \succ y$ implica que no $y \succsim x$. Por lo tanto, suponer $z \succsim x$ lleva a una contradicción, lo que implica que $x \succ z$.

□

Ahora que hemos definido formalmente el concepto de preferencia y su orden a través de la relación $\succsim$, es importante definir el concepto de elección, teniendo en cuenta que, en general, debemos considerar la elección entre un conjunto dado de alternativas como un conjunto en sí mismo [2]. Si S es el conjunto de alternativas disponibles, como ya habíamos denotado anteriormente, llamaremos $E(S)$ a la alternativa o alternativas escogidas del conjunto de alternativas S. $E(S)$ es, por supuesto, un subconjunto de S. Cada elemento de $E(S)$ es preferible a todos los elementos de S que no están en $E(S)$ e indiferente a todos los elementos de $E(S)$; y, por lo tanto, si $x \in E(S)$, se cumple que $x \succsim y$ para todo $y \in S$. Por otra parte, si tenemos que $x \succsim y$ para todo $y \in S$ y, además, $x \in S$, entonces, por la definición 1.3, no hay ningún elemento $z \in S$ tal que $z \succ x$. Por lo tanto, podemos definir $E(S)$ formalmente como sigue [1]:

Definición 1.6. $E(S)$ es el conjunto de todas las alternativas $x \in S$ tales que, para todo $y \in S$, se cumple que $x \succsim y$.

Hay que señalar que $E(S)$ describe una relación funcional, ya que asigna una elección a cada conjunto de alternativas S. Podemos llamarla función de elección. Arrow describía el siguiente lema para un conjunto de dos alternativas $S = \{x, y\}$ [2].

Lema 1.7. *$x \succ y$ es equivalente a decir que x es el único elemento de $E(\{x, y\})$.*

Demostración. En primer lugar, probemos que si $x \succ y$, entonces x es el único elemento de $E(\{x, y\})$. Sea $\{x, y\}$ el conjunto compuesto por las dos alternativas x e y. Supongamos que $x \succ y$. Entonces $x \succsim y$, por el punto 2 del lema 1.5, y $x \succsim x$, por el punto 1 del lema 1.5, de modo que $x \in E(\{x, y\})$; pero, de nuevo por la definición 1.3, ya que $x \succ y$, esto implica que no se cumple $y \succsim x$, de modo que $y \notin E(\{x, y\})$, es decir, $E(\{x, y\})$ contiene el único elemento x.

A la inversa, comprobemos que si x es el único elemento de $E(\{x, y\})$, entonces $x \succ y$. Supongamos que $E(\{x, y\})$ contiene el único elemento x. Como y no pertenece a $E(\{x, y\})$, entonces no se cumple $y \succsim x$ y, por la definición 1.3, $x \succ y$. □

En caso de que ni $x \succ y$ ni $y \succ x$, tenemos, claramente, que $x \sim y$, y esto equivale a decir que $E(\{x, y\})$ contiene tanto a x como a y. Si, entonces, conocemos $E(\{x, y\})$ para todos los conjuntos de dos elementos, hemos definido completamente las relaciones $\succ$ e $\sim$ y, por tanto, la relación $\succsim$; pero, por la definición 1.6, conocer la relación $\succsim$ determina completamente la elección

$E(S)$ para todos los conjuntos de alternativas S. Por lo tanto, tal y como entendemos la elección, ésta puede determinarse para cualquier conjunto de alternativas S mediante el conocimiento de las elecciones en los pares de alternativas.

Es importante indicar que la representación del mecanismo de elección a través de relaciones de orden no es la única existente. Por ejemplo, mucha bibliografía emplea las funciones de utilidad [7]. Estas se definen de la siguiente forma [16]:

Definición 1.8. Sea S un espacio métrico dotado de una relación de preferencia $\succsim$. Decimos que $\succsim$ es representable si existe alguna $u : S \longrightarrow \mathbb{R}$ tal que para todo x, $y \in S$,

$$x \succsim y \Leftrightarrow u(x) \geq u(y).$$

La aplicación u se denomina la función de utilidad que representa $\succsim$.

Esta representación se emplea en la literatura en la cual el uso de comparaciones interpersonales de bienestar o utilidad no se descarta. Las reglas clásicas de decisión social, como el utilitarismo o cualquier otra regla que permita compensaciones entre las utilidades experimentadas por diferentes individuos, simplemente no pueden expresarse en términos de una función de bienestar social de Arrow. Para que el procedimiento de elección social incorpore información sobre las comparaciones interpersonales de utilidad, hay que generalizar la noción de función de bienestar social. En lugar de determinar la preferencia social sobre la base de un perfil de ordenaciones de preferencias individuales, una función de bienestar social asignaría una preferencia social a cada perfil admisible de funciones de utilidad individual.

Sin embargo, Arrow indicaba que, en lo que respecta a las funciones de utilidad, existe la dificultad formal de que, si no se hacen suficientes suposiciones de continuidad sobre la ordenación, puede no existir ninguna manera de asignar números reales a las distintas alternativas de manera que se satisfagan los requisitos habituales de una función de utilidad. En cualquier caso, simplemente habría que sustituir la expresión $x \succsim y$ por la expresión $u(x) \geq u(y)$, y la estructura de las pruebas no cambiaría [2].

Una vez definidos los conceptos de ordenación, de preferencia y de elección, es importante aclarar otros conceptos relacionados con éstos: el de votación y función de bienestar social. Consideramos entornos en los que hay un conjunto de alternativas S, un conjunto de n votantes y una regla que describe cómo se utilizan las ordenaciones de los votantes para determinar un resultado. Consideramos dos tipos diferentes de reglas. Una regla de votación (*voting rule*) o función de elección social produce un único ganador, y una regla de agregación social (*ranking rule*) o función de bienestar produce una ordenación social sobre las alternativas. Las reglas de votación se utilizan, obviamente, en las elecciones o, en general, cuando un grupo tiene que elegir una de varias alternativas, es decir, tomar decisiones a nivel social. Una regla de clasificación podría utilizarse cuando un departamento universitario clasifica a las alternativas a la facultad basándose en las preferencias de los miembros actuales de la facultad. En ambos casos, suponemos que la clasificación de cada votante está representada por una relación de orden tal y como la hemos definido anteriormente, $\succsim$. Esto podría realizarse de manera más restrictiva utilizando relaciones de preferencia en lugar de las de orden [12], pero, por ser coherentes con el trabajo de Arrow, emplearemos las relaciones de orden.

De aquí en adelante denotaremos por $\succsim_i$ la relación de orden asociada al votante i y se denominará perfil de ordenación al conjunto $(\succsim_1, \ldots, \succsim_n) \in \mathcal{R}$ que describe las relaciones de orden de los n votantes. Además, diremos que V es un conjunto de votantes e $I = \{1, \ldots, n\}$ es el conjunto total de todos los votantes.

Definición 1.9. Una **regla de votación** o **función de elección social** definida sobre un subconjunto $\mathcal{A} \subset \mathcal{R}^n$ es una regla $f : \mathcal{A} \longrightarrow S$ que relaciona cada perfil de ordenación $(\succsim_1, \ldots, \succsim_n) \in \mathcal{R}^n$ con un elemento $f(\succsim_1, \ldots, \succsim_n) = x \in S$, al cual denominaremos ganador de la elección, es decir, $E(S) = \{x\}$.

Ya hemos visto que el conjunto de alternativas escogidas puede ser un conjunto $E(S)$ cuando hablábamos de una relación individual de ordenación. En el caso de una función de elección social consideraremos únicamente los casos donde $E(S)$ posee un único elemento.

De esta idea de elección es de donde surge, naturalmente, la necesidad de tener, a parte de las reglas de votación, reglas de agregación social, es decir, funciones de bienestar social. Fue Bergson quien las denominó con este nombre [4] y, más tarde, Arrow propuso la siguiente definición [2]:

Definición 1.10. Una **función de bienestar social** o **función de agregación social** definida sobre un subconjunto $\mathcal{A} \subset \mathcal{R}^n$ es una regla $F : \mathcal{A} \longrightarrow \mathcal{R}$ que relaciona un perfil de ordenación $(\succsim_1, \ldots, \succsim_n) \in \mathcal{R}^n$ a una única relación de orden $\succsim = F(\succsim_1, \ldots, \succsim_n) \in \mathcal{R}$.

Es necesario decir que no hace falta que nos preguntemos si se puede definir tal función de bienestar social. Tomando el conjunto $\mathcal{A} \subset \mathcal{R}^n$, diremos que es un conjunto **admisible** de relaciones de ordenación individual si la función de bienestar social define una ordenación social, es decir, una relación que satisface los axiomas 1.1 y 1.2 [1]. Una función de bienestar social sería aquella para la que todo conjunto de ordenaciones individuales fuera admisible. Sin embargo, podemos considerar, con algún tipo de fundamento *a priori*, que ciertos tipos de ordenamientos individuales no tienen por qué ser admisibles. Por ejemplo, en la economía del bienestar se ha supuesto o implicado con frecuencia que cada individuo valora las distintas alternativas únicamente en función de su consumo de ellas. Si este es el caso, sólo deberíamos exigir que nuestra función de bienestar social se defina para aquellos conjuntos de ordenamientos individuales que sean del tipo descrito; sólo éstos deberían ser admisibles. Hablaremos más en profundidad sobre esto cuando expliquemos el teorema de Arrow.

Además, con el fin de facilitar la notación de las funciones de bienestar, para cualquier perfil de ordenación $(\succsim_1, \ldots, \succsim_n) \in \mathcal{R}^n$, denotaremos $\succ = F_p(\succsim_1, \ldots, \succsim_n) \in \mathcal{P}$ como la relación de preferencia estricta derivada de $\succsim = F(\succsim_1, \ldots, \succsim_n)$ [13].

La definición 1.10 de función de bienestar cobrará suma importancia cuando hablemos de los teoremas de imposibilidad. Esta definición, además, tiene una estrecha relación con la definición 1.9 de función de elección social. Por un lado, si tenemos una función de bienestar social siempre podemos obtener una regla de votación sin más que, una vez obtenida la relación de orden $\succsim$, eliminar la alternativa peor clasificada y repetir el proceso. Cuando sólo nos quede una alternativa habremos concluido y tendremos nuestra regla de votación. Una forma más evidente de obtener esto es generar un *ranking* basado en la ordenación $\succsim$ obtenida y determinar como ganador de la regla de votación al que se encuentre en el primer puesto del *ranking*. Por otro lado, en algunos casos también se cumple el recíproco, esto es, podemos calcular la función de bienestar a partir de la función de elección, pero no para todos los casos.

1.2. Funciones de bienestar

Ya hemos visto las definiciones de regla de votación y función de bienestar. Muy a menudo nos encontramos estos dos conceptos como si fuesen sinónimos en un claro abuso del lenguaje. Principalmente, nos encontramos con que muchas veces se emplea el término “regla de votación” en lugar del de “función de bienestar” debido a la fácil conversión que hemos visto de las segundas a las primeras. En esta sección analizaremos cinco funciones de bienestar bien conocidas en la literatura: la regla de Borda, la de veto, la de mayoría simple o pluralidad, la de mayoría por pares y la de Copeland. Dentro de éstas, vamos a distinguir dos tipos: las de puntuación (Borda, veto y mayoría simple) y las de comparación por pares (mayoría por pares y Copeland) [17].

1.2.1. Funciones de puntuación

Si tomamos el perfil de ordenación $(\succsim_1, \ldots, \succsim_n) \in \mathcal{R}^n$, cada una de las relaciones de ordenación que componen este perfil de ordenación, sea $\succsim_i$ la relación de ordenación del votante i, la denotaremos con un vector de puntuación [17] de la forma:

$$\boldsymbol{\alpha_i} = (\alpha_i^{(1)}, \ldots, \alpha_i^{(m)}),$$

siendo m el número de alternativas y cumpliéndose que $\alpha_i^{(1)} \geq \cdots \geq \alpha_i^{(m)}$, con $\alpha_i^{(j)} \in \mathbb{N}$ para todo $i \in I$ y $j \in \{1, \ldots, m\}$. Así, si tenemos una alternativa x en la j-ésima posición en la relación de ordenación del votante i-ésimo, entonces tendrá asociado el valor $\alpha_i^{(j)}$ del vector de puntuación. Este valor lo denotaremos como $\boldsymbol{\alpha}_i(x)$. Sumando todas las puntuaciones obtenidas para cada alternativa y ordenándolas obtenemos la relación de orden resultante de aplicar la función de bienestar. Para cada alternativa x, la puntuación total obtenida la denotaremos por $\boldsymbol{\alpha}(x)$.

Ejemplo 1.11. Vamos a tener en cuenta un escenario ficticio con el que ejemplificaremos el uso de cada una de las funciones de bienestar social [17]. Este escenario estará formado por tres votantes y por tres alternativas, $S = \{x, y, z\}$. Las relaciones de ordenación de cada votante son las siguientes:

1. Votante 1: $x \succsim_1 y \succsim_1 z$. Es decir, el orden del votante será xyz.
2. Votante 2: $y \succsim_2 z \succsim_2 x$. Es decir, el orden del votante será yzx.
3. Votante 3: $x \succsim_3 y \succsim_3 z$. Es decir, el orden del votante será xyz.

Mayoría simple o pluralidad

En la regla de mayoría simple, sólo la alternativa más votada obtiene un punto, todos los demás alternativas no obtienen ningún punto. Por lo tanto, el vector de puntuación de esta regla es el siguiente:

$$\boldsymbol{\alpha} = (1, 0, \dots, 0),$$

Ejemplo 1.12. Tomando el escenario definido en 1.11, vemos que las puntuaciones asociadas a cada alternativa son las siguientes:

Mayoría simple			
	x	y	z
Votante 1	1	0	0
Votante 2	0	1	0
Votante 3	1	0	0
Resultados	2	1	0

Por lo tanto, sumando las puntuaciones vemos que x tiene 2 puntos, y tiene 1 punto y z tiene 0 puntos. Por lo tanto, la relación de ordenación $\succsim$ obtenida por la función de bienestar cumple que $x \succsim y \succsim z$.

Veto

La función de bienestar anterior, es decir, mayoría simple, si observamos el resultado, premia a la alternativa que se encuentra más veces en la primera posición en cada relación de ordenación de cada votante. Otro argumento diferente a este puede ser el decir que debería ganar la alternativa que menos veces se encuentra en la última posición. Así es como se define la regla de veto (también conocida como antipluralidad) y, por lo tanto, su vector de puntuaciones es el siguiente:

$$\boldsymbol{\alpha} = (1, \dots, 1, 0),$$

Ejemplo 1.13. Tomando de nuevo el escenario definido en 1.11, vemos que las puntuaciones asociadas a cada alternativa son las siguientes:

Veto			
	x	y	z
Votante 1	1	1	0
Votante 2	0	1	1
Votante 3	1	1	0
Resultados	2	3	1

Por lo tanto, sumando las puntuaciones vemos que x tiene 2 puntos, y tiene 3 puntos y z tiene 1 punto. Por lo tanto, en este caso, la relación de ordenación $\succsim$ obtenida por la función de bienestar cumple que $y \succsim x \succsim z$.

Borda

Otro protocolo de puntuación muy conocido es la regla de Borda, debido a Jean-Charles de Borda [6]. Cuando hay m alternativas, la alternativa en la primera posición de una votación obtiene $m-1$ puntos de este votante, la alternativa en la segunda posición obtiene $m-2$ puntos, y así sucesivamente hasta que la alternativa en la última posición obtiene 0 puntos. El vector de puntuación para Borda con m alternativas es, según lo explicado, el siguiente:

$$\boldsymbol{\alpha} = (m-1, m-2, \dots, 0),$$

Ejemplo 1.14. Tomando de nuevo el escenario definido en 1.11, vemos que las puntuaciones asociadas a cada alternativa son las siguientes:

Borda			
	x	y	z
Votante 1	2	1	0
Votante 2	0	2	1
Votante 3	2	1	0
Resultados	4	5	1

Por lo tanto, sumando las puntuaciones vemos que x tiene 4 puntos, y tiene 4 puntos y z tiene 1 puntos. Por lo tanto, en este caso, la relación de ordenación $\succsim$ obtenida por la función de bienestar cumple que $x \succsim y \succsim z$ y que $y \succsim x \succsim z$. En este caso, a diferencia de los anteriores, tenemos que el conjunto $E(S) = \{x, y\}$ está formado por dos alternativas, en lugar de por una, como en el resto de casos vistos hasta ahora.

1.2.2. Funciones de comparación de pares

A diferencia de los protocolos de puntuación, aquí los candidatos no reciben puntos inmediatamente de los votantes según su posición en las votaciones, sino que todos los candidatos se comparan entre sí en enfrentamientos uno contra uno. Es de esta forma como se comprueba quién de los dos candidatos es preferido (es decir, quién está por delante del otro en las listas de preferencias de los votantes) por la mayoría de los votantes. Si hay un número par de votantes, pueden producirse empates. Veremos una de las funciones de bienestar más conocidas y que satisface estas características.

Mayoría por pares

Este método consiste en enfrentar a cada candidato con todos los demás en una serie de comparaciones de uno contra uno. El ganador de cada emparejamiento es el candidato preferido por la mayoría de los votantes. A menos que haya un empate, siempre hay una mayoría cuando sólo hay dos opciones. Esta regla de comparación entre pares se denomina mayoría por pares, la cual explicaremos formalmente en secciones posteriores.

Se considera que el candidato preferido por cada votante es el de la pareja que el votante clasifica (o valora) más alto en su relación de orden particular. Por ejemplo, si tomamos las alternativas $x, y \in S$ y realizamos un enfrentamiento por pares entre ellas, es necesario contar tanto el número de votantes que han preferido a x antes que a y, como el número de votantes que han preferido a y antes que a x. Si x es preferida por más votantes, entonces es la ganadora de ese emparejamiento. Veamos ahora un ejemplo de cómo ejecutar esta función de bienestar.

Ejemplo 1.15. Tendremos ahora para este caso otro escenario ficticio. Supongamos que tenemos ahora el conjunto de alternativas $S = \{a, b, c, d\}$. Además, tendremos 6 votantes, cuyas relaciones de ordenación son las siguientes [17]:

1. Votante 1: $B \succsim_1 d \succsim_1 c \succsim_1 a$. Es decir, el orden del votante será *bdca*
2. Votante 2: $d \succsim_2 b \succsim_2 c \succsim_2 a$. Es decir, el orden del votante será *dbca*.
3. Votante 3: $c \succsim_3 a \succsim_3 b \succsim_3 d$. Es decir, el orden del votante será *cabd*.
4. Votante 4: $a \succsim_4 c \succsim_4 d \succsim_4 b$. Es decir, el orden del votante será *acdb*.

5. Votante 5: $a \succsim_5 c \succsim_5 b \succsim_5 d$. Es decir, el orden del votante será $acbd$.

Representaremos los enfrentamientos uno contra uno de la forma $a?b$, siendo en este caso a y b los candidatos comparados entre sí. Veamos en la siguiente tabla todos los enfrentamientos posibles y su resultado:

Condorcet						
	$a?b$	$a?c$	$a?d$	$b?c$	$b?d$	$c?d$
Votante 1	b	c	d	b	b	d
Votante 2	b	c	d	b	d	d
Votante 3	a	c	a	c	b	c
Votante 4	a	a	a	c	d	c
Votante 5	a	a	a	c	b	c
Resultados	a	c	a	c	b	c

Por lo tanto, viendo los resultados tenemos que c gana 3 enfrentamientos por pares, a gana 2, b gana 1 y d no gana ninguno. Por lo tanto, el orden social establecido será $c \succsim a \succsim b \succsim d$.

Este método sirve como base para las funciones de bienestar basadas en comparación por pares. De hecho, muchas veces se utiliza como sinónimo de método de Condorcet. En esta memoria utilizaremos ese término para englobar a un conjunto de funciones de bienestar concretas. Veremos más adelante cuáles cuando hablemos del trabajo de Condorcet.

Hay que decir que este método no tiene en cuenta los empates en esos enfrentamientos por pares, por eso decimos que es una "regla base" para el resto de métodos de comparación por pares. Para este caso, veremos el siguiente método.

Copeland

Como sucede anteriormente, para determinar los puntos de los candidatos, todos ellos entran en una competición uno contra uno con todos los demás. Si uno de los dos candidatos en dicha contienda es preferido por la mayoría de los votantes, recibe un punto y el otro ningún punto. Si ninguno de ellos tiene la mayoría (es decir, si empatan, lo que sólo es posible para un número par de votantes), entonces ambos reciben medio punto. La suma de los puntos de un candidato da la puntuación Copeland de este candidato. El que tiene más puntos (es decir, la mayor puntuación Copeland) es el ganador de Copeland. Ramón Llull propuso un sistema de votación que también se basa en comparaciones por pares [8]. Los sistemas de Llull y el mencionado sistema de Copeland son muy parecidos. La única diferencia crucial es que los empates en las comparaciones por pares (que pueden producirse si hay un número par de votantes) se recompensan con un punto en el sistema de Llull, no con medio punto.

Intentando establecer una relación entre el trabajo de Llull y Copeland, se puede construir toda una familia de sistemas de votación que generaliza la puntuación dada por un empate: Para cada $\alpha \in \mathbb{Q}$ tal que $\alpha \in [0, 1]$, Copeland$^{\alpha}$ denota el mismo sistema que el descrito anteriormente, excepto que un empate en una comparación por parejas se recompensa con α puntos. Así, si denotamos la puntuación obtenida por una alternativa x con Copeland$^{\alpha}$ como $\mathrm{C}^{\alpha}(x)$, este valor se corresponderá con el número de victorias de x más α veces el número de empates en las comparaciones por pares de x. Así pues, la ordenación $\succsim$ resultante de la función de bienestar se definirá según el *ranking* resultado de ordenar (en este caso hablamos del orden usual en $\mathbb{R}$) los valores de $\mathrm{C}^{\alpha}(x)$ para cada alternativa x. En esta notación, Copeland$^{\frac{1}{2}}$ es el sistema Copeland común, y el sistema de Llull es Copeland^{1}. Si hay un número impar de votantes, no es posible el empate en ninguna comparación por mayoría; los sistemas Copeland$^{\alpha}$ son entonces idénticos para todos los α.

Ejemplo 1.16. Tendremos ahora para este caso otro escenario ficticio. Supongamos que tenemos ahora el conjunto de alternativas $S = \{a, b, c, d\}$. Además, tendremos 6 votantes, cuyas relaciones de ordenación son las siguientes [17]:

1. Votante 1: $a \succsim_1 d \succsim_1 c \succsim_1 b$. Es decir, el orden del votante será *adcb*

2. Votante 2: $c \succsim_2 d \succsim_2 b \succsim_2 a$. Es decir, el orden del votante será *cdba*.

3. Votante 3: $c \succsim_3 d \succsim_3 b \succsim_3 a$. Es decir, el orden del votante será *cdba*.

4. Votante 4: $b \succsim_4 d \succsim_4 a \succsim_4 c$. Es decir, el orden del votante será *bdac*.

5. Votante 5: $a \succsim_5 c \succsim_5 d \succsim_5 b$. Es decir, el orden del votante será *acdb*.

6. Votante 6: $a \succsim_6 c \succsim_6 b \succsim_6 d$. Es decir, el orden del votante será *acbd*.

Como ya hicimos en el ejemplo de la regla de mayoría por pares, representaremos los enfrentamientos uno contra uno de la forma $a?b$, siendo en este caso a y b los candidatos comparados entre sí. Veamos en la siguiente tabla todos los enfrentamientos posibles y su resultado:

Copeland						
	$a?b$	$a?c$	$a?d$	$b?c$	$b?d$	$c?d$
Votante 1	a	a	a	c	d	d
Votante 2	b	c	d	c	d	c
Votante 3	b	c	d	c	d	c
Votante 4	b	a	d	b	b	d
Votante 5	a	a	a	c	d	c
Votante 6	a	a	a	c	b	c
Resultados	?	a	?	c	d	c

Por lo tanto, viendo los resultados (donde denotamos el empate como ?), obtenemos los siguientes resultados:

$$
\begin{aligned}
\mathrm{C}^{\alpha}(a) &= 1 + 2\alpha \\
\mathrm{C}^{\alpha}(b) &= \alpha \\
\mathrm{C}^{\alpha}(c) &= 2 \\
\mathrm{C}^{\alpha}(d) &= 1 + \alpha
\end{aligned}
$$

Así, en caso de tener $\alpha = \frac{1}{2}$ como en el Copeland común, entonces habría dos ganadores, $E(S) = \{a, c\}$ y el orden social establecido es $c \succsim a \succsim d \succsim b$ o $a \succsim c \succsim d \succsim b$, lo cual se puede escribir de la forma $c \sim a \succsim d \succsim b$. Sin embargo, si utilizásemos la variante desarrollada por Llull, entonces el ganador sería únicamente a, $E(S) = \{a\}$ y el orden establecido sería $a \succsim c \sim d \succsim b$.

1.3. Contextualización histórica

En esta introducción hemos explicado ya conceptos muy importantes que conforman la estructura formal de la teoría de la elección social, esto es: función de bienestar, función de elección, relación de ordenación y preferencia estricta. Además, hemos explicado también una serie de funciones de bienestar ampliamente conocidas y que serán nombradas a lo largo de este trabajo.

Sin embargo, aún nos queda por definir una serie de conceptos importantes y, sobre todo, nos falta por exponer los problemas que existen en torno a la elección social y que, en gran medida, explican la necesidad de que una teoría sea desarrollada. Para esto, realizaremos una contextualización histórica sobre varias de las figuras más importantes de la teoría de la elección social junto con sus aportaciones [10]. Muchas de ellas ya las hemos visto como parte de la Sección 1.1 de definiciones, pero sin contexto (las funciones de bienestar social de Arrow, por ejemplo). Por lo tanto, en esta sección trataremos de contextualizar estos conceptos y completar con otros que nos faltaban.

Para ello, dividiremos la época histórica de la elección social en dos etapas. Una primera etapa de nacimiento y de desarrollo; y una segunda etapa centrada en el bienestar social y donde irrumpe la teoría desarrollada por Arrow.

1.3.1. Primera etapa

Comenzaremos esta etapa hablando de uno de los autores más importantes en la teoría de la eleción social: Condorcet. Hemos explicado en la Sección 1.2.2 la función de bienestar de mayoría por pares, que es una función de bienestar muy importante en la teoría desarrollada por Condorcet, como veremos a continuación. En esta sección profundizaremos sobre otros conceptos de suma importancia y que, como resulta lógico, están muy ligados a esta función de bienestar.

En primer lugar, cabe decir que su principal contribución a la disciplina es el *Essai* de 1785 [9]. Esta obra está esencialmente dedicada a la utilización de la teoría de la probabilidad para resolver cuestiones concretas, como lo que hoy se denomina teorema del jurado de Condorcet. Pero los desarrollos sobre este tema son bastante recientes y Condorcet es quizás más famoso por lo que propone en la parte preliminar de su *Essai*, en la que se pueden descubrir las nociones de ganador de Condorcet y la paradoja de Condorcet.

Definición 1.17. Sea un conjunto de alternativas S y un conjunto de relaciones de preferencia $\{\succsim_1, \dots, \succsim_n\}$. Si existe $x \in S$ tal que para todo $y \in S$ se cumple que x es el ganador del método de mayoría por pares, entonces llamamos a x el **ganador de Condorcet** o **alternativa de Condorcet**. De la misma forma, si existe $x \in S$ tal que para todo $y \in S$ se cumple que y es el ganador del método de mayoría por pares, entonces llamamos a x el **perdedor de Condorcet**.

Esta definición de ganador de Condorcet se incluye dentro de lo que se conoce como criterio de Condorcet. Dada cualquier función de elección social, diremos que tal función satisface el criterio de Condorcet si, de existir un ganador de Condorcet, esta función lo elige como ganador. Se dice que estos sistemas de votación son métodos de Condorcet. La regla de votación asociada a la mayoría por pares, definida en 1.2.2, satisface este criterio y se considera método de Condorcet, como es lógico, pues es la empleada para definir el concepto. Es por esto por lo que decíamos que la función de mayoría por pares tiene cierta relevancia. También cumple este criterio la regla de mayoría absoluta que hemos definido en 1.2.2.

En cuanto a la **paradoja de Condorcet** (o paradoja de la votación), ésta puede presentarse de forma muy sencilla y expone los problemas asociados a las funciones de bienestar de los que hablábamos al comienzo de esta sección. Supongamos que hay tres votantes y tres opciones x, y

y z. Supongamos que tenemos las siguientes relaciones de ordenación para cada votante:

1. Votante 1: $x \succsim_1 y \succsim_1 z$. Es decir, el orden del votante será xyz.
2. Votante 2: $y \succsim_2 z \succsim_2 x$. Es decir, el orden del votante será yzx.
3. Votante 3: $z \succsim_1 x \succsim_1 y$. Es decir, el orden del votante será zxy.

Según la regla de mayoría por pares, explicada en la Sección 1.2.2, tenemos la siguiente tabla de puntuaciones:

Mayoría por pares			
	$x?y$	$x?z$	$y?z$
Votante 1	x	x	y
Votante 2	y	z	y
Votante 3	x	z	z
Resultados	x	z	y

Así pues, analizando la tabla de resultados vemos que x se declara socialmente preferida a y, y socialmente preferida a z y z socialmente preferida a x. Por lo tanto, no hay ningún ganador de Condorcet y, no sólo eso, si no que se quebranta el axioma de transitividad a través del cual definíamos el orden, es decir, el resultado de la función de bienestar social no es un orden. Esta paradoja muestra que las reglas de elección social no son tan simples como podría parecer en un primer momento, pues con ejemplo sencillo de una regla de elección muy popular hemos conseguido quebrar la transitividad que ya exponíamos cuando hablábamos de orden en el axioma 1.2. Otro ejemplo proporcionado por Condorcet consiste en mostrar que la regla de Borda, la cual hemos explicado en 1.2.1, puede fallar en la selección de un ganador de Condorcet existente.

Junto con Nicolas de Condorcet, Jean-Charles de Borda es considerado también uno de los padres de la teoría de la elección social y del voto. El trabajo de Borda sobre la votación es bastante limitado, apenas unas páginas en un documento matemático en el que introduce lo que ahora se conoce como la regla de Borda, descrita en la Sección 1.2.1; pero en el que también demuestra que la regla de la pluralidad es defectuosa, ya que puede seleccionar una opción que es derrotada en parejas (utilizando la regla de la mayoría por pares) tomando todas las demás opciones, es decir, tal y como vemos en la definición 1.17 sería entonces un perdedor de Condorcet. Sin embargo, sorprendentemente, no existe ninguna prueba de que un ganador de Borda no pueda ser un perdedor de Condorcet [6].

Entre las contribuciones importantes, también hay que mencionar a Charles Dodgson (conocido bajo su seudónimo de Lewis Carroll) y Edward Nanson. En particular, Dodgson propone varias reglas basadas en la regla de la mayoría y la regla de Borda, pero ahora es más conocido por una regla que ha sido estudiada por los informáticos [3]. Cuando no hay un ganador de Condorcet, Dodgson sugiere que seleccionemos la opción que se convierta en ganadora de Condorcet tras el menor número de transposiciones en las preferencias individuales. Por ejemplo, en el caso de la paradoja de Condorcet presentada anteriormente, si el votante 2 en lugar de tener la clasificación yzx tiene una clasificación yxz, se puede observar que x se convierte en ganador de Condorcet tras una única transposición, es decir, si zx se convierte en xz. Sin embargo, este ejemplo en concreto, por sus propiedad de simetría, cumple que transposiciones similares llevarían a y o z a ser ganadores de Condorcet tras una única transposición.

1.3.2. Segunda etapa

Esta segunda etapa está protagonizada por la monografía de Kenneth Arrow de 1951 [2], de la cual ya hemos hablado, y que pertenece a la ascendencia de una literatura que trata de rescatar la economía de bienestar. Esta afirmación se ejemplifica con algunos hechos, como el de que Arrow utiliza el mismo concepto de "función de bienestar social". En la segunda edición de la monografía [1], la función de bienestar social se supone que satisface cuatro condiciones. En primer lugar, las ordenaciones individuales de preferencias no están restringidas. En segundo lugar, en caso de preferencias unánimes de una alternativa x sobre otra alternativa y, entonces la alternativa x debe ser socialmente preferida a la y (criterio de Pareto). En tercer lugar, no existe ningún individuo cuya preferencia estricta se convierta, sean cuales sean las preferencias de los demás individuos, en la preferencia social estricta (dicho individuo se llamaría "dictador"). En cuarto lugar, si en dos perfiles de ordenaciones de preferencias individuales dos alternativas x e y son idénticas en cuanto a preferencias (por ejemplo, el votante 1 prefiere x a y en ambos perfiles, el votante 2 es indiferente entre x e y en ambos perfiles, etc.), entonces los dos valores tomados por la función de bienestar social deben ser idénticos en cuanto a la relación de orden del par x e y. Por ejemplo, x se prefiere socialmente a y en ambos perfiles. Esta cuarta condición es denominada por Arrow "independencia de las alternativas irrelevantes". Arrow demuestra que si hay al menos dos votantes (el número de votantes es finito) y tres estados sociales, no existe ninguna función de bienestar social que satisfaga las cuatro condiciones. Éste se suele denominar el Teorema de Imposibilidad de Arrow y será explicado en profundidad en este trabajo.

Capítulo 2

Teorema de Imposibilidad de Arrow

En esta sección profundizaremos en lo explicado al final de la Sección 1.3, es decir, cómo las cuatro condiciones que Arrow impone a las funciones de bienestar social son imposibles de cumplir, lo que se denomina Teorema de Imposibilidad de Arrow.

2.1. Propiedades de la función de bienestar

2.1.1. Caso con dos alternativas

Comenzamos nuestro análisis considerando el caso más sencillo posible: aquel en el que sólo hay dos alternativas sobre las que decidir. Las llamamos alternativa x y alternativa y. La alternativa x, por ejemplo, podría ser el "statu-quo", y la alternativa y podría ser un proyecto público concreto cuya ejecución se está contemplando. Los datos de nuestro problema son las preferencias individuales de los miembros de la sociedad sobre las dos alternativas. Suponemos que hay 4 individuos, o votantes. La familia de preferencias individuales entre las dos alternativas puede describirse mediante el perfil siguiente:

$$\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n) \in \mathbb{R}^n.$$

donde β_i toma el valor 1, 0, o -1 dependiendo de si el votante i prefiere la alternativa x a la alternativa y, es indiferente entre ellas, o prefiere la alternativa y a la alternativa x, respectivamente. Es importante indicar que esta notación será empleada únicamente para el caso de dos alternativas y, aunque se asemeje a la notación que empleábamos en la Sección 1.2.1, nada tiene que ver con ella. En este caso, cada elemento del perfil se corresponde con la preferencia de un individuo, mientras que en el otro caso, cada elemento del perfil se correspondía con cada una de las alternativas y establecía la relación de orden entre ellas. Así pues, empleando esta notación, podemos redefinir la función de bienestar de la siguiente forma [13].

Definición 2.1. Una función de bienestar social (o agregador de bienestar social) es una regla F que asigna una preferencia social, es decir, $F(\boldsymbol{\beta}) \in \{-1, 0, 1\}$, a cada posible perfil de preferencias individuales $\boldsymbol{\beta} \in \{-1, 0, 1\}^n$.

Si recordamos la definición general de función de bienestar social 1.10, vemos que esta función lo que hace es, tomando todas las relaciones de orden $\succsim_i$ de cada votante i, asociarles una relación de orden $\succsim$. En este caso, las únicas relaciones de orden posibles son:

1. La relación $\succsim_1$ que cumple que $x \succsim_1 y$. Asociamos esta relación con el número 1.
2. La relación $\succsim_2$ que cumple que $x \succsim_2 y$ y que $y \succsim_2 x$, es decir, $x \sim y$. Asociamos esta relación con el número 0.
3. La relación $\succsim_3$ que cumple que $y \succsim_1 x$. Asociamos esta relación con el número -1.

Por lo tanto, vemos que esta definición es, como era de esperar, un caso particular de la definición 1.10. Una vez visto esto, es momento de avanzar en el análisis de las funciones de bienestar social de la mano de este caso particular de dos alternativas. Y, para continuar, es necesario dar la siguiente definición.

Definición 2.2. La función de bienestar social $F(\boldsymbol{\beta})$ es **de Pareto**, o cumple el **criterio de Pareto**, si respeta la unanimidad de preferencia estricta por parte de los votantes, es decir, si $F(1, \ldots, 1) = 1$ y $F(-1, \ldots, -1) = -1$.

Todas las funciones de bienestar social que consideraremos serán de Pareto. Esto significa que las funciones respetarán, aunque de forma débil, las preferencias individuales.

Ejemplo 2.3. Abundan las funciones de Pareto de bienestar social entre dos alternativas [13]. Sea $(\gamma_1, \ldots, \gamma_n) \in \mathbb{R}^n_+$ un vector de números no negativos, no todos nulos. Entonces, se podría definir:

$$F(\boldsymbol{\beta}) = \operatorname{sign}\left(\sum_i \gamma_i \beta_i\right),$$

donde, para cualquier $a \in \mathbb{R}$, se cumple que

$$\operatorname{sign}(a) = \begin{cases} 1 & \text{si } a > 0, \\ 0 & \text{si } a = 0, \\ -1 & \text{si } a < 0. \end{cases}$$

Esta función cumpliría que es de Pareto (por el hecho de que $\gamma_i \geq 0$, para todo $i \in \{1, \ldots, n\}$).

Un caso particular importante es la votación por mayoría. Denotaremos ahora la función por $M(\cdot)$ y la definiremos como $F(\cdot)$ con $\gamma_i = 1$ para todo $i \in I$. Entonces $M(\boldsymbol{\beta}) = 1$ si y sólo si el número de votantes que prefieren la alternativa x a la alternativa y es mayor que el número de votantes que prefieren y a x. Del mismo modo, $M(\boldsymbol{\beta}) = -1$ si y sólo si los que prefieren y a x son más numerosos que los que prefieren x a y. Finalmente, en caso de igualdad de estos dos números, tenemos $M(\boldsymbol{\beta}) = 0$, es decir, la relación resultante es la indiferencia.

Naturalmente, esta regla que acabamos de explicar es exactamente la misma que la regla de la mayoría por pares que explicamos la Sección 1.2.2, pero con la particularidad de que, en lugar de haber una cantidad arbitraria de alternativas, tenemos sólo dos. Por lo demás, exceptuando el cambio de notación que empleamos para el caso particular de dos alternativas, todo es igual. Teniendo en cuenta esto, podemos definir el siguiente tipo de funciones de bienestar social.

Definición 2.4. Decimos que una función de bienestar social es **dictatorial** si existe un votante d, llamado dictador, tal que, para cualquier perfil $\boldsymbol{\beta}$, si se cumple que $\beta_d = 1$, entonces $F(\boldsymbol{\beta}) = 1$ y, análogamente, si se cumple que $\beta_d = -1$, entonces $F(\boldsymbol{\beta}) = -1$. Es decir, la relación de orden del dictador prevalece como preferencia social.

Una función de bienestar social dictatorial es de Pareto en el sentido de la definición 2.2. Para las funciones de bienestar social del ejemplo 2.3, tenemos dictadura siempre que $\gamma_d > 0$ para algún individuo d y $\gamma_i = 0$ para todo $i \neq d$, ya que en este caso $F(\boldsymbol{\beta}) = \beta_d$.

La función de bienestar social de la mayoría desempeña un papel de referencia en la teoría de la elección social. Además de ser de Pareto, tiene tres propiedades importantes, que pasamos a enunciar formalmente. La primera (simetría entre votantes) dice que la función de bienestar social trata a todos los votantes en igualdad de condiciones. La segunda (neutralidad entre alternativas) dice que, igualmente, la función de bienestar social no distingue, *a priori*, ninguna de las dos alternativas. La tercera (monotonía) dice, con más fuerza que la propiedad de Pareto de la definición 2.2, que la función de bienestar social es sensible a las preferencias individuales.

Definición 2.5. La función de bienestar social F es **simétrica entre votantes** (o **anónima**) si los nombres de los votantes no importan, es decir, si una permutación de preferencias entre votantes no altera la preferencia social. Para expresar esta idea formalmente, tomemos una función biyectiva $\pi : \{1, \ldots, n\} \longrightarrow \{1, \ldots, n\}$ (una permutación); entonces, para cualquier perfil $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)$ tenemos que $F(\beta_1, \ldots, \beta_n) = F(\beta_{\pi(1)}, \ldots, \beta_{\pi(n)})$.

Definición 2.6. La función de bienestar social F es **neutral entre alternativas** si $F(\boldsymbol{\beta}) = F(\beta_1, \ldots, \beta_n) = -F(-\beta_1, \ldots, -\beta_n) = -F(-\boldsymbol{\beta})$ para cada perfil $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)$, es decir, si la preferencia social se invierte cuando invertimos las preferencias de todos los individuos.

Definición 2.7. La función de bienestar social F es **monótona**, siempre que $\boldsymbol{\beta}' \geq \boldsymbol{\beta}$, $\boldsymbol{\beta}' \neq \boldsymbol{\beta}$ y $F(\boldsymbol{\beta}) \geq 0$, entonces tenemos que $F(\boldsymbol{\beta}') \geq 0$[1]. Es decir, si x es socialmente preferido o indiferente

[1]Cuando decimos que $\boldsymbol{\beta}' \geq \boldsymbol{\beta}$, lo que estamos diciendo es que para cada $\beta_i \in \boldsymbol{\beta}$ y $\beta'_i \in \boldsymbol{\beta}'$, se cumple que $\beta'_i \geq \beta_i$.

a y y algunos votantes elevan su consideración de x, entonces x seguirá siendo socialmente preferido.

Veamos que, como anticipamos antes, la votación por mayoría satisface las tres propiedades recién definidas.

Proposición 2.8. *La regla de la mayoría M satisface las propiedades de simetría entre votantes (anonimato), neutralidad entre alternativas y monotonía.*

Demostración. Demostremos que la regla de la mayoría satisface cada una de las definiciones. Tomemos el perfil $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)$.

1. **Anonimato**. Sea $\pi : \{1, \ldots, n\} \longrightarrow \{1, \ldots, n\}$ una permutación. Entonces se cumple que $\sum_{i=1}^{n} \beta_i = \sum_{i=1}^{n} \beta_{\pi(i)}$ y, por lo tanto, tenemos que:

$$M(\boldsymbol{\beta}) = \operatorname{sign}\left(\sum_{i=1}^{n} \beta_i\right) = \operatorname{sign}\left(\sum_{i=1}^{n} \beta_{\pi(i)}\right)$$

2. **Neutralidad**. Tenemos que se cumple lo siguiente:

$$\begin{aligned} M(\boldsymbol{\beta}) &= M(\beta_1, \ldots, \beta_n) \\ &= \operatorname{sign}\left(\sum_{i=1}^{n} \beta_i\right) \\ &= \operatorname{sign}\left(-\sum_{i=1}^{n} (-\beta_i)\right) \\ &= -\operatorname{sign}\left(\sum_{i=1}^{n} (-\beta_i)\right) \\ &= -M(-\beta_1, \ldots, -\beta_n) = -M(-\boldsymbol{\beta}) \end{aligned} \tag{2.1}$$

3. **Monotonía**. Asumamos que $M(\boldsymbol{\beta}) = M(\beta_1, \ldots, \beta_n) \geq 0$, entonces $\operatorname{sign}\left(\sum_{i=1}^{n} \beta_i\right) \geq 0$, lo que implica que $\sum_{i=1}^{n} \beta_i \geq 0$. Tomemos $(\beta_1', \ldots, \beta_n') \geq (\beta_1, \ldots, \beta_n)$ tal que $(\beta_1', \ldots, \beta_n') \neq (\beta_1, \ldots, \beta_n)$. Entonces, $\sum_{i=1}^{n} \beta_i' > 0$, lo que implica que $\operatorname{sign}\left(\sum_{i=1}^{n} \beta_i'\right) > 0$ y esto, finalmente, implica que $M\left(\beta_1', \ldots, \beta_n'\right) = 1$.

□

Por lo tanto, acabamos de ver que una condición necesaria de la regla de la mayoría es que se cumplan estas tres propiedades. Sin embargo, estas condiciones no son sólo necesarias, si no también suficientes [14].

Teorema 2.9. *(Teorema de May) Una función de bienestar social F es una es una función de bienestar social de mayoría simple si y sólo si es simétrica entre votantes, neutral entre alternativas y monótona.*

Demostración. Ya hemos argumentado que las tres propiedades son necesarias en la proposición 2.8. Para establecer la suficiencia, nótese en primer lugar que la propiedad de simetría entre los votantes significa que la preferencia social sólo depende del número total de votantes que prefieren la alternativa x a y, del número total que es indiferente y del número total que prefiere y a x. Dados $(\beta_1, \dots, \beta_n)$, denotamos

$$n^+(\beta_1, \dots, \beta_n) = \#\{i : \beta_i = 1\} \text{ y } n^-(\beta_1, \dots, \beta_n) = \#\{i : \beta_i = -1\}\,[2].$$

Entonces la simetría entre votantes nos permite expresar $F(\beta_1, \dots, \beta_n)$ de la forma

$$F(\beta_1, \dots, \beta_n) = G\left(n^+(\beta_1, \dots, \beta_n), n^-(\beta_1, \dots, \beta_n)\right).$$

Ahora, supongamos que $(\beta_1, \dots, \beta_n)$ es tal que $n^+(\beta_1, \dots, \beta_n) = n^-(\beta_1, \dots, \beta_n)$. Entonces $n^+(-\beta_1, \dots, -\beta_n) = n^-(\beta_1, \dots, \beta_n) = n^+(\beta_1, \dots, \beta_n) = n^-(-\beta_1, \dots, -\beta_n)$, y, por tanto,

$$\begin{aligned}
F(\beta_1, \dots, \beta_n) &= G\left(n^+(\beta_1, \dots, \beta_n), n^-(\beta_1, \dots, \beta_n)\right) \\
&= G\left(n^+(-\beta_1, \dots, -\beta_n), n^-(-\beta_1, \dots, -\beta_n)\right) \\
&= F(-\beta_1, \dots, -\beta_n) \\
&= -F(\beta_1, \dots, \beta_n)
\end{aligned}$$

La última igualdad se deriva de la neutralidad entre alternativas. Como el único número que es igual a su negativo es el cero, concluimos que si $n^+(\beta_1, \dots, \beta_n) = n^-(\beta_1, \dots, \beta_n)$ entonces $F(\beta_1, \dots, \beta_n) = 0$.

Supongamos a continuación que $n^+(\beta_1, \dots, \beta_n) > n^-(\beta_1, \dots, \beta_n)$. Denotamos $H = n^+(\beta_1, \dots, \beta_n)$, $J = n^-(\beta_1, \dots, \beta_n)$; entonces $J < H$. Digamos, sin pérdida de generalidad, que $\beta_i = 1$ para $i \leq H$ y $\beta_i \leq 0$ para $i > H$. Definamos un nuevo perfil $(\beta'_1, \dots, \beta'_n)$ como

$$\begin{cases}
\beta'_i = \beta_i = 1 & \text{para } i \leq J < H, \\
\beta'_i = 0 & \text{para } J < i \leq H, \\
\beta'_i = \beta_i \leq 0 & \text{para } i > H.
\end{cases}$$

Entonces $n^+(\beta'_1, \dots, \beta'_n) = J$ y $n^-(\beta'_1, \dots, \beta'_n) = n^-(\beta_1, \dots, \beta_n) = J$. Por lo tanto, $F(\beta'_1, \dots, \beta'_n) = 0$. Pero, por construcción, la alternativa x ha perdido fuerza en el nuevo perfil de ordenación. En efecto, $(\beta_1, \dots, \beta_n) > (\beta'_1, \dots, \beta'_n)$ ya que se cumple que $\beta_{J+1} = 1 > 0 = \beta'_{J+1}$. Por lo tanto, por la propiedad de monotonía, debemos tener $F(\beta_1, \dots, \beta_n) = 1$.

[2] Recordemos que la notacion $\#A \equiv$ “cardinal del conjunto A” $\equiv$ “número de elementos del conjunto A”.

Por el contrario, $n^-(\beta_1,\ldots,\beta_n) > n^+(\beta_1,\ldots,\beta_n)$ entonces $n^+(-\beta_1,\ldots,-\beta_n) > n^-(-\beta_1,\ldots,-\beta_n)$ y, por tanto, $F(-\beta_1,\ldots,-\beta_n) = 1$. Por lo tanto, por la neutralidad entre las alternativas:

$$F(\beta_1,\ldots,\beta_n) = -F(-\beta_1,\ldots,-\beta_n) = -1.$$

Concluimos así que $F(\beta_1,\ldots,\beta_n)$ es efectivamente una función de bienestar social de mayoría simple. □

2.1.2. Caso general

Una vez visto el caso concreto de dos alternativas, en esta sección intentaremos generalizar los conceptos e ideas vistas con el fin de poder aplicarlas a contextos un poco más complejos, pero, a la vez, más interesantes a nivel de estudio. Para ello, comenzaremos generalizando la definición de función de Pareto o, lo que es lo mismo, el criterio de Pareto.

Definición 2.10. La función de bienestar social $F : \mathcal{A} \longrightarrow \mathcal{R}$ es **de Pareto** (o, dicho de otro modo, satisface el **Principio de Pareto**) si, para cualquier par de alternativas $\{x, y\} \subset S$ y cualquier perfil de preferencia $(\succsim_1,\ldots,\succsim_n) \in \mathcal{A}$, tenemos que x es socialmente preferida (estrictamente) a y, es decir, $xF_p(\succsim_1,\ldots,\succsim_n)\,y$, siempre que $x \succ_i y$ para cada $i \in I$.

En el siguiente ejemplo observamos una regla que cumple con la definición del Principio de Pareto.

Ejemplo 2.11. La regla de Borda, definida en 1.2.1. Esta regla cumple que es de Pareto, puesto que si $x \succ_i y$ para todo $i \in I$, entonces $\boldsymbol{\alpha}_i(x) > \boldsymbol{\alpha}_i(y)$ y, por tanto, $\sum_i \boldsymbol{\alpha}_i(x) > \sum_i \boldsymbol{\alpha}_i(y)$, provocando esto que $xF_p(\succsim_1,\ldots,\succsim_n)\,y$.

Ahora que hemos visto la definición del criterio de Pareto para el caso general, veamos también la definición de función dictatorial para este caso [1].

Definición 2.12. Se dice que una función de bienestar social es **dictatorial** si existe un individuo d tal que, para todo $x, y \in S$ (siendo S el conjunto de alternativas), $x \succ_d y$ implica que $x \succ y$ independientemente de las ordenaciones del perfil $(\succsim_1,\ldots,\succsim_n) \in \mathcal{R}^n$ de todos los individuos distintos de d, donde $\succ$ es la relación de preferencia social estricta correspondiente al perfil $(\succsim_1,\ldots,\succsim_n)$.

A continuación, exponemos una importante restricción sobre las funciones de bienestar social sugerida por primera vez por Arrow [1]. La restricción dice que las preferencias sociales entre dos alternativas cualesquiera dependen sólo de las preferencias individuales entre las mismas dos alternativas. Hay dos posibles líneas de justificación para este supuesto. La primera es estrictamente normativa y tiene un atractivo considerable: argumenta que al establecer una clasificación social entre x e y, la presencia o ausencia de alternativas distintas de x e y no debería importar. Son irrelevantes para el hecho en cuestión. La segunda es de carácter práctico. El supuesto facilita enormemente la tarea de tomar decisiones sociales porque ayuda a separar los problemas. La determinación de la clasificación social sobre un subconjunto de alternativas no necesita ninguna información sobre las preferencias individuales sobre las alternativas fuera de este subconjunto.

Definición 2.13. La función de bienestar social $F : \mathcal{A} \to \mathcal{R}$ definido en el dominio $\mathcal{A} \subset \mathcal{R}^n$ satisface la condición de **independencia de alternativas irrelevantes** (o IIA, por sus siglas en inglés) si la preferencia social entre dos alternativas cualesquiera $\{x, y\} \subset S$ depende únicamente del perfil de las preferencias individuales sobre las mismas alternativas. Formalmente, para cualquier par de alternativas $\{x, y\} \subset S$ y para cualquier par de perfiles de preferencias $(\succsim_1, \ldots, \succsim_n) \in \mathcal{A}$ y $(\succsim_1', \ldots, \succsim_n') \in \mathcal{A}$ con la propiedad de que, para cada $i \in I$,

$$x \succsim_i y \Leftrightarrow x \succsim_i' y \text{ e } y \succsim_i x \Leftrightarrow y \succsim_i' x,$$

tenemos que

$$xF\left(\succsim_1, \ldots, \succsim_n\right) y \quad \Leftrightarrow \quad xF\left(\succsim_1', \ldots, \succsim_n'\right) y$$

y que

$$yF\left(\succsim_1, \ldots, \succsim_n\right) x \quad \Leftrightarrow \quad yF\left(\succsim_1', \ldots, \succsim_n'\right) x.$$

Ejemplo 2.14. Veamos ahora como la regla de Borda no satisface la condición de IIA. La razón es sencilla: el rango de una alternativa depende de la colocación de cada una de las otras alternativas. Supongamos, por ejemplo, que hay dos votantes con las relaciones de orden $\succsim_1, \succsim_2 \in \mathcal{R}$ y un conjunto de tres alternativas $S = \{x, y, z\}$. Para las preferencias

$$x \succ_1 z \succ_1 y$$
$$y \succ_2 x \succ_2 z$$

tenemos que x es socialmente preferida a y. De hecho, $\boldsymbol{\alpha}(x) = 3$ and $\boldsymbol{\alpha}(y) = 2$. Pero para las preferencias

$$x \succ_1' y \succ_1' z,$$
$$y \succ_2' z \succ_2' x$$

tenemos que y es socialmente preferida a x. De hecho, en este caso $\boldsymbol{\alpha}(x) = 2$ y $\boldsymbol{\alpha}(y) = 3$. Sin embargo, la ordenación relativa de x e y no ha cambiado para ninguno de los dos votantes.

Para ilustrarlo de otra forma, esta vez con tres votantes y cuatro alternativas $\{x, y, z, w\}$, consideremos

$$z \succ_1 x \succ_1 y \succ_1 w,$$
$$z \succ_2 x \succ_2 y \succ_2 w,$$
$$y \succ_3 z \succ_3 w \succ_3 x.$$

Aquí, y es socialmente preferida a x, ya que $\boldsymbol{\alpha}(x) = 4$ y $\boldsymbol{\alpha}(y) = 5$. Pero supongamos ahora que las alternativas z y w se desplazan hacia el fondo para todos los votantes (lo que, debido a la propiedad de Pareto que hemos definido en 2.2, es una forma de decir que las dos alternativas se eliminan del conjunto de alternativas):

$$\begin{aligned} &x \succ_1' y \succ_1' z \succ_1' w, \\ &x \succ_2' y \succ_2' z \succ_2' w, \\ &y \succ_3' x \succ_3' z \succ_3' w. \end{aligned} \tag{2.2}$$

Entonces x se prefiere socialmente a y, pues en este caso $\boldsymbol{\alpha}(x) = 8$ y $\boldsymbol{\alpha}(y) = 7$. Por tanto, la presencia o ausencia de las alternativas z y w importa para la preferencia social entre x y y. Otra modificación llevaría la alternativa x al fondo para el votante 3:

$$\begin{aligned} &x \succ_1'' y \succ_1'' z \succ_1'' w, \\ &x \succ_2'' y \succ_2'' z \succ_2'' w \\ &y \succ_3'' z \succ_3'' w \succ_3'' x \end{aligned}$$

Ahora y es socialmente preferido a x, pues $\boldsymbol{\alpha}(x) = 6$ y $\boldsymbol{\alpha}(y) = 7$ (lo que, en relación con el resultado visto en 2.2, es un buen resultado desde el punto de vista del votante 3).

2.2. Teorema de imposibilidad

Hemos visto en la sección anterior una serie de propiedades que pueden poseer las funciones de bienestar. De hecho hemos observado la dificultad de la búsqueda de funciones de bienestar que cumplan varias de estas propiedades. En esta sección ahondaremos más en esta cuestión y la remataremos con el enunciado y demostración del Teorema de Imposibilidad de Arrow.

Para hacer esto, en primer lugar, agruparemos las propiedades explicadas en la sección anterior como condiciones requeridas para una función de bienestar. Una vez hecho esto, demostraremos que estas condiciones se cumplen en el caso particular de emplear dos alternativas. Para finalizar, veremos que estas condiciones no se cumplen en el caso general con al menos tres alternativas enunciado en forma del Teorema de Imposibilidad de Arrow.

2.2.1. Condiciones

En su trabajo, Arrow propuso que las funciones de bienestar deben cumplir una serie de condiciones para ser una "buena" función de bienestar o, dicho de otro modo, Arrow propuso una serie de condiciones deseables que una función de bienestar debe poseer para poder generar como resultado una ordenación social que agregue de forma satisfactoria las relaciones de orden de cada individuo. Las condiciones son las siguientes.

1. **Condición 1**. Entre todas las alternativas existe un conjunto S de tres alternativas tal que, para cualquier conjunto de ordenaciones individuales $(\succsim_1^*, \ldots, \succsim_n^*) \in \mathcal{R}^n$ de las alternativas en S, existe un conjunto admisible de ordenaciones individuales $(\succsim_1, \ldots, \succsim_n) \in \mathcal{R}^n$ de todas las alternativas tal que, para cada individuo $i \in I$, $x \succsim_i y$ si y sólo si $x \succsim_i^* y$ para $x, y \in S$.

2. **Condición 2**. Sean $(\succsim_1, \ldots, \succsim_n)$ y $(\succsim_1', \ldots, \succsim_n')$ dos perfiles de relaciones de ordenación individual, $\succsim$ y $\succsim'$ las correspondientes ordenaciones sociales, y $\succ$ y $\succ'$ las correspondientes relaciones de preferencia social estricta. Supongamos que para cada $i \in I$, dada una alternativa $x \in S$, las dos relaciones de ordenación individual están conectadas de la siguiente manera:
$$\begin{cases} x' \succsim_i' y' \Leftrightarrow x' \succsim_i y' & x \neq x', x \neq y', \\ x \succsim_i y' \Rightarrow x \succsim_i' y' & \forall y', \\ x \succ_i y \Rightarrow x \succ_i' y' & \forall y'. \end{cases}$$
Entonces, si $x \succ_i y$, $x \succ_i' y$.

3. **Condición 3**. La función de bienestar social debe cumplir la propiedad IIA de la cual ya hemos visto la definición en 2.13.

4. **Condición 4**. La función de bienestar social no debe ser impuesta. Veamos la definición de ser impuesta.

   Definición 2.15. Se dirá que una función de bienestar social es **impuesta** si, para algún par de alternativas distintas $x, y \in S$, $x \succsim y$ para cualquier perfil de ordenaciones individuales $(\succsim_1, \ldots, \succsim_n) \in \mathcal{R}^n$ donde $\succsim$ es el ordenamiento social correspondiente a $(\succsim_1, \ldots, \succsim_n)$.

5. **Condición 5**. La función de bienestar social no puede cumplir la propiedad dictatorial de la cual ya hemos visto la definición en 2.12.

De aquí en adelante supondremos que las funciones de bienestar social de las cuales hablaremos cumplen las Condiciones 1-5, a no ser que se especifique lo contrario. El resultado central de esta sección y de la teoría de elección social de Arrow es que la suposición de que una función de bienestar social cumple las cinco condiciones anteriores nos lleva a una contradicción, es decir, el resultado central es el Teorema de Imposibilidad de Arrow. Para poder probar esto debemos definir y demostrar una serie de resultados que utilizaremos posteriormente en la demostración de susodicho teorema. Comencemos viendo qué es un conjunto decisivo de votantes [13].

Definición 2.16. Se dice que el conjunto $V \subset I$ de votantes se dice que V es:

1. **Decisivo para x frente a y** si para todo $i \in V$ se cumple que $x \succ_i y$ y para todo $i \in I \backslash V$ se cumple que $y \succ_i x$, entonces se tiene que $x \succ y$.

2. **Decisivo** si, para cualquier par $\{x, y\} \subset S$, V es decisivo para x frente a y.

3. **Completamente decisivo para x frente a y** si para todo $i \in V$ se cumple que $x \succ_i y$ e, indiferentemente de las preferencias de los votantes $i \in I \backslash V$, se tiene que $x \succ y$.

En alguna literatura, como es el caso de Arrow [1], esta definición que damos de un conjunto completamente decisivo es en realidad como se define un conjunto decisivo. Estos dos conceptos muy similares y veremos posteriormente cómo se relacionan.

Antes de eso, vamos a tratar de explicar la definición de completamente decisivo. Sea el perfil de ordenaciones individuales $(\succsim_1, \ldots, \succsim_n)$. La condición $x \succ_i y$ para todo $i \in V$ restringe la familia perfiles de ordenaciones $(\succsim_1, \ldots, \succsim_n)$ restringiendo el rango de variación de aquellos componentes cuyos subíndices están en V a las relaciones de ordenación que tienen la propiedad dada con respecto a x e y. A cada $(\succsim_1, \ldots, \succsim_n)$, una determinada función de bienestar social le asigna un ordenamiento social $\succsim$; según esta escala podemos tener, en general, $x \succ y$ o $x \sim y$ o $y \succ x$. Supongamos que resulta que, para todo $(\succsim_1, \ldots, \succsim_n)$ consistente con la condición de que $x \succ_i y$ para toda $i \in V$, la resultante $\succsim$ es tal que $x \succ y$; entonces podemos decir que V es completamente decisivo para x frente a y. Intuitivamente, el concepto de conjunto completamente decisivo puede explicarse como sigue: Un conjunto de individuos es completamente decisivo si, siempre que todos ellos prefieran x a y, la sociedad prefiere x a y independientemente de las preferencias que cualquier individuo pueda tener respecto a otras alternativas distintas de x o y. Obsérvese que un conjunto puede ser completamente decisivo para x frente a y sin serlo para y frente a x. Hay que destacar que la cuestión de si un conjunto dado de individuos es completamente decisivo o no con respecto a un par dado de alternativas, x e y, está determinada por la función de bienestar social.

El concepto de conjunto decisivo es muy similar a este, pero más restrictivo, pues, además de tener la condición de que $x \succ_i y$ para toda $i \in V$, se tiene la condición de que $y \succ_i x$ para toda $i \in I \backslash V$. Por este motivo, es lógico pensar que si un conjunto es decisivo, entonces también es completamente decisivo. Esto, de hecho, sucede, pero antes de verlo es necesario demostrar un resultado previo.

Proposición 2.17. *Sean $(\succsim_1, \ldots, \succsim_n), (\succsim'_1, \ldots, \succsim'_n) \in \mathcal{R}^n$ dos perfiles de ordenaciones individuales tales que, para unas alternativas $x, y \in S$ dadas, $x \succ'_i y$ para todo $i \in I$ cumpliendo que $x \succsim_i y$. Por lo tanto, si $x \succ y$, entonces $x \succ' y$ (donde $\succ$ y $\succ'$ son las preferencias sociales estrictas correspondientes a $(\succsim_1, \ldots, \succsim_n)$ y $(\succsim'_1, \ldots, \succsim'_n)$, respectivamente).*

Demostración. Para probar esto, suponemos que sólo hay tres alternativas $x, y, z \in S$. Para cada

$i \in I$, definimos la relación de ordenación $\succsim_i''$, como sigue:

$$x' \succsim_i'' y' \Leftrightarrow \begin{cases} x' \succsim_i y' \text{ y } x' \neq z \\ \text{o bien} \\ y' = z. \end{cases} \tag{2.3}$$

Esto equivale a mover z de su posición en $\succsim_i$ a la parte inferior, pero, por lo demás, dejar $\succsim_i$ sin cambios. Es fácil comprobar que $\succsim_i''$ es una relación de orden, es decir, satisface los axiomas 1.1 de conexidad y 1.2 de transitividad. Además, para cada $i \in S$, $\succsim_i''$ ordena los elementos x e y de la misma manera que $\succsim_i$; es decir,

$$x' \succsim_i'' y' \Leftrightarrow x' \succsim_i y' \text{ para } x', y' \in \{x, y\}. \tag{2.4}$$

A partir de 2.4 y de la Condición 3, $E(\{x,y\}) = E''(\{x,y\})$, donde $E(S)$ y $E''(S)$ son las elecciones sociales realizadas para un conjunto de alternativas S cuando $(\succsim_1, \ldots, \succsim_n)$ y $(\succsim_1'', \ldots, \succsim_n'')$ son los perfiles de relaciones de orden individuales, respectivamente. Por hipótesis, $x \succ y$; a partir del lema 1.7, $E(\{x,y\})$ contiene el único elemento x. Por tanto, $E''(\{x,y\})$ contiene el único elemento x, o, por el Lema 1.7,

$$x \succ'' y. \tag{2.5}$$

Definir los ordenaciones individuales $(\succsim_1^*, \ldots, \succsim_n^*)$, como sigue:

$$x' \succsim_i^* y' \Leftrightarrow \begin{cases} x' \succsim_i^* y' \text{ y } x' \neq z \\ \text{o bien} \\ y' = z. \end{cases} \tag{2.6}$$

lo cual es exactamente paralelo a 2.3. A partir de 2.3, 2.6 y la definición 1.3, se cumple que $y \succ_i'' z$, $y \succ_i^* z$, para todo $i \in I$. Por lo tanto, si

$$[x' \neq x, \;\; y' \neq x, \;\; x' \succsim_i'' y'] \Leftrightarrow x' \succsim_i^* y'. \tag{2.7}$$

Además, $x \succ_i'' z$, $x \succ_i^* z$ para todo $i \in I$. Por 2.3, para todo $i \in I$ tal que $x \succsim_i'' y$, $x \succsim_i y$; por hipótesis, $x \succ_i' y$ para tal i, y por tanto, por la ecuación 2.6, $x \succ_i^* y$. Por lo tanto,

$$\forall y', x \succsim_i' y' \Rightarrow x \succsim_i *y'; \tag{2.8}$$

$$\forall y', x \succ_i'' y' \Rightarrow x \succ_i *y'. \tag{2.9}$$

Por las ecuaciones 2.7, 2.8, 2.9 y 2.5, se cumplen las hipótesis de la Condición 2; por tanto, $x \succ^* y$. De 2.7 se deduce, de la misma manera que antes, que $E^*(\{x,y\}) = E'(\{x,y\})$, por lo que $x \succ' y$. □

Así pues, estamos en condiciones de demostrar la implicación que antes señalábamos.

Proposición 2.18. *Sea $V \subset I$ y $x, y \in S$. Si el conjunto V es decisivo para x frente a y, entonces V es completamente decisivo para x frente a y.*

Demostración. Sea $(\succsim'_1, \ldots, \succsim'_n) \in \mathcal{R}^n$ cualquier conjunto de ordenaciones individuales sujetas únicamente a la condición de que

$$x \succ'_i y \ \ \forall i \in V. \tag{2.10}$$

Para demostrar que V es completamente decisivo, es necesario, según la definición 2.16, demostrar que, para cada conjunto $(\succsim'_1, \ldots, \succsim'_n) \in \mathcal{R}^n$, el correspondiente ordenamiento social $\succsim'$ es tal que $x \succ' y$. Pero de la condición anterior 2.10 y de la hipótesis de que $x \succ_i y$ para todo $i \in V$ e $y \succ_i x$ para todo $i \in I \backslash V$, se deduce que $x \succ'_i y$ siempre que $x \succsim_i y$. Vemos ahora que, por la proposición 2.17, ésto implica que $x \succ' y$. □

El significado de esta consecuencia puede formularse un poco como sigue: imaginemos que un observador ve que los individuos escriben sus ordenaciones individuales y las entregan a las autoridades centrales, que entonces forman una relación de orden social basada en las ordenaciones individuales de acuerdo con la función de bienestar social. Supongamos además que este observador se da cuenta de que, para un par específico de alternativas $x, y \in S$, todos los individuos de un determinado conjunto V de individuos prefieren x a y, mientras que todos los que no están en V prefieren y a x, y que el ordenamiento social resultante sitúa a x por encima de y. Entonces, el observador tiene derecho a decir, sin mirar ningún otro aspecto de los ordenaciones individuales y sociales, que V es un conjunto completamente decisivo para x frente a y, es decir, que, si los gustos cambian, pero de tal manera que todos los individuos de V siguen prefiriendo x a y (aunque hayan cambiado su clasificación para todas las demás alternativas y aunque los individuos que no están en V hayan cambiado completamente su escala), entonces el ordenamiento social seguirá clasificando a x por encima de y.

Así pues, estamos en condiciones de enunciar el Teorema de Imposibilidad de Arrow, también denominado, Teorema de Posibilidad General [2].

Teorema 2.19. *Si existen, al menos, tres alternativas que los miembros de la sociedad pueden ordenar libremente de cualquier manera, entonces cualquier función de bienestar social satisfaciendo las Condiciones 2 y 3, y dando lugar a una relación de orden que cumpla los axiomas 1.1 y 1.2, entonces ha de ser impuesta o dictatorial.*

Sin embargo, Blau en [5] se dio cuenta de que el Teorema de Imposibilidad de Arrow estaba declarado incorrectamente. Blau comprobó que intercambiando las Condiciones 1 y 2 por las dos siguientes, el problema se evitaba. No entraremos a aclarar cuál era el problema que encontró Blau por salirse del alcance de este proyecto.

1. **Condición** $1'$. Todas las posibles ordenaciones lógicas de las alternativas contenidas en S son admisibles.

2. **Condición** $2'$. Para un par de alternativas dadas $x, y \in S$, sea el perfil de relaciones de orden individuales $(\succsim_1, \dots, \succsim_n) \in \mathcal{R}^n$. Supongamos que x es elevada dentro de alguna de las relaciones de orden individuales. Entonces, si x era socialmente preferida a y, se mantendrá como socialmente preferida a y después de los cambios.

Cabe decir que la Condición $1'$ es innecesariamente fuerte, pero facilitará el desarrollo una demostración del teorema más sencilla. Por otro lado, la Condición $2'$, si nos fijamos, es equivalente a la Condición 2 en presencia de las Condiciones $1'$ y 3, de hecho, esta condición es exactamente la misma que la proposición 2.17 la cual ya hemos demostrado.

Por último, nótese que en ningún momento nombramos el criterio de Pareto en el enunciado del teorema y hemos hablando bastante a lo largo de esta memoria sobre él. Esto es porque la enunciación original del teorema, es decir, la que hemos realizado en 2.19 no lo incluye; sin embargo, enunciaciones posteriores sí. Para relacionar estos dos enunciados del teorema, demostraremos la siguiente proposición que relaciona el criterio de Pareto con las Condiciones $2'$, 3 y 4. De aquí en adelante lo denominaremos la **Condición** P, siguiendo la notación de Arrow [1].

Proposición 2.20. *La Condición P es deducible de las Condiciones $2'$, 3 y 4.*

Demostración. En primer lugar, el criterio de Pareto equivale a decir que, para $x, y \in S$ tal que $x \neq y$, I es un conjunto decisivo para x frente a y [1]. Por lo tanto, vamos a probar esto.

Si intercambiamos x e y en la definición 2.15, entonces la Condición 4 dice que existe un perfil de ordenaciones individuales $(\succsim_1, \dots, \succsim_n) \in \mathcal{R}^n$ tal que no se cumple $y \succsim x$, donde $\succsim$ es la relación de orden social correspondiente al conjunto de relaciones de orden individuales $(\succsim_1, \dots, \succsim_n)$. Es decir, por el apartado 5 del lema 5

$$x \succ y. \tag{2.11}$$

Sea $(\succsim_1', \dots, \succsim_n') \in \mathcal{R}^n$ cualquier perfil de ordenaciones individuales tal que

$$x \succ_i' y \tag{2.12}$$

para todo $i \in I$. De 2.12 se cumple que $x \succ_i' y$ para todo i tal que $x \succsim_i y$. Entonces, a partir de 2.11 y de la Condición $2'$, $x \succ' y$. Como esto es válido para cualquier conjunto de ordenaciones que satisfagan 2.12, se deduce de la de la definición de I que $x \succ' y$ para cualquier perfil de ordenaciones tal que $x \succ_i' y$ para $i \in I$, $y \succ_i' x$ para $i \notin I$. Por lo tanto, por la proposición 2.18, tenemos que I es decisivo para x frente a y. $\square$

Así pues, en las posteriores secciones, las condiciones que le exigiremos a una función de bienestar serán las Condiciones $1'$, 3, P y 5; en lugar de las Condiciones iniciales 1-5. Primero, veremos de nuevo el caso particular para dos alternativas, donde sí existe una función de bienestar que satisfaga las condiciones y, de hecho, es la regla de la mayoría por pares. Segundo, veremos el caso general donde, si hay al menos tres alternativas, ninguna función de bienestar satisfará todas las condiciones a la vez.

2.2.2. Caso con dos alternativas

En el caso de que haya únicamente dos alternativas, es posible construir una función de bienestar social que cumpla las Condiciones $1'$, 3, P y 5. Por supuesto, la Condición $1'$ debe modificarse para este caso. Exigimos ahora que todo conjunto de ordenaciones individuales de las dos alternativas en cuestión dé lugar a una ordenación social que satisfaga los axiomas 1.1 y 1.2.

La función de bienestar que satisfará estas condiciones no será otra que la regla de la mayoría y demostraremos que esta regla cumple las condiciones solicitadas una a una.

Teorema 2.21. *(Teorema de la posibilidad para dos alternativas): Si el número total de alternativas es dos, el método de decisión por mayoría es una función de bienestar social que satisface las Condiciones $1'$, 3, P y 5 y produce una relación de orden social de las dos alternativas para cada perfil de ordenaciones individuales.*

Demostración. Para demostrar que satisface la Condición $1'$ debemos demostrar que $M(\boldsymbol{\beta})$, tal y como definíamos en 2.3, es una ordenación débil, es decir, satisface los axiomas de conexidad y transitividad, 1.1 y 1.2, respectivamente. En primer lugar, tomaremos $H = n^+(\beta_1, \ldots, \beta_n)$ y $J = n^-(\beta_1, \ldots, \beta_n)$ tal y como denotábamos en la demostración del teorema de May 2.9. Así pues, por construcción, tenemos que

$$M(\beta) \geq 0 \Leftrightarrow H \geq J. \tag{2.13}$$

Claramente, siempre va a suceder que, o bien $H \geq J$, o bien $J \geq H$; lo cual, por 2.13 tenemos que:

$$M(\beta) \geq 0 \text{ o } M(\beta) \leq 0, \tag{2.14}$$

respectivamente. Por lo tanto, la regla es conexa y se cumple el axioma 1.1. En cuanto a la transitividad, supongamos que, siendo $\succsim = M(\beta)$, $x \succsim y$ y que $y \succsim z$; debemos ver que $x \succsim z$. Como vemos, para poder comprobar la transitividad de manera formal debemos cambiar la notación específica que habíamos creado para dos alternativas y emplear la general. Esto es lógico, pues al tener dos alternativas la transitividad se cumple de manera trivial por el hecho de cumplirse la conexidad. Si lo vemos de manera formal, observamos que se tiene que dos de las alternativas x, y y z han de ser iguales. Las posibilidades son que $x = y$, $x = z$ o que $y = z$. Comprobémoslo para cada caso:

1. Caso $x = y$. En este caso hay que ver que $y \succsim z$, pero esto se cumple por hipótesis.

2. Caso $x = z$. En este caso hay que ver que $x \succsim x$. Como hemos visto que $\succsim$ es conexa, es decir, cumple el axioma 1.1, esto se cumple.

3. Caso $y = z$. En este caso hay que ver que $x \succsim y$, pero esto se cumple por hipótesis.

Por lo tanto, la regla de la mayoría M cumple la Condición $1'$.

Consideremos ahora que la Condición 3 de IIA es trivial en este caso porque el único conjunto T que contiene más de un miembro contiene al conjunto entero, que consta de dos miembros, es decir, $T = S$. Si T contiene un elemento, $E(T)$ es ese elemento independiente de las alternativas que no están en T; si T contiene dos elementos y, por lo tanto, $T = S$; entonces $E(T) = E(S)$ está ciertamente determinado por las ordenaciones individuales de los elementos en S ya que no hay otros.

En cuanto a la Condición P, ya dijimos en el ejemplo 2.3 que la regla de la mayoría satisfacía el criterio de Pareto. De hecho, construimos la función de la mayoría como un caso particular de un conjunto de funciones de la forma

$$F(\boldsymbol{\beta}) = \operatorname{sign}\left(\sum_i \gamma_i \beta_i\right) \text{ con } \gamma_i \geq 0 \quad \forall i \in I,$$

donde todas cumplen que son de Pareto, independientemente del valor de γ_i, para todo $i \in I$.

Por último, en cuanto a la Condición 5 (no dictadura), supongamos que hubiera un individuo d que cumpliera las condiciones de la definición 2.4. Supongamos que $\beta_d = 1$, mientras que $\beta_i = -1$ para todo $i \neq d$. Entonces, por $\beta_d = 1$ y no $\beta_i \neq 1$ para $i \neq 1$ tenemos que $H = 1$. Además, $\beta_i \neq 1$ para $i \neq d$, de modo que $J \geq 1 = H$. Por 2.13, tenemos que $M(\boldsymbol{\beta}) \leq 0$. Sin embargo, por la definición 2.4, $\beta_d = 1$ implica $M(\boldsymbol{\beta}) = 1$. Por lo tanto, no puede haber ningún dictador y se cumple así la Condición 5. □

Este teorema es, en cierto modo, el fundamento lógico del sistema bipartidista angloamericano. Para una referencia posterior, obsérvese que la demostración dada anteriormente de que el método de decisión por mayoría satisface las Condiciones P y 5 era independiente de la suposición de que sólo había dos alternativas. También es cierto que el método de decisión por mayoría satisface la Condición 3 de IIA independientemente del número total de alternativas. A partir de la definición de la regla de la mayoría por pares dada en la Sección 1.2.2, es obvio que la verdad o falsedad de la afirmación $M(\boldsymbol{\beta}) > 0$ es invariante bajo cualquier cambio de ordenaciones individuales que deje invariantes, para cada individuo, las posiciones relativas de x e y, por la propia construcción basada en la comparación por pares de la función. Por la definición 1.6, $E(S)$ está completamente determinado por el conocimiento de la verdad o falsedad de la afirmación $M(\boldsymbol{\beta}) > 0$ para cada par $x, y \in S$; por tanto, $E(S)$ es ciertamente invariante bajo cualquier cambio de las ordenaciones individuales que deje invariantes las ordenaciones dentro de S.

Lema 2.22. *Para cualquier espacio de alternativas, el método de decisión por mayoría por pares es una función de bienestar social que satisface las Condiciones* 3*,* P *y* 5*.*

El ejemplo de la paradoja de Condorcet dado en la Sección 1.3 muestra que la regla de la mayoría por pares no satisface la Condición $1'$ cuando hay más de dos alternativas. Ahora estamos preparados para examinar la construcción de funciones de bienestar social en este último caso.

2.2.3. Caso general

El siguiente teorema es, precisamente, el Teorema de Imposibilidad de Arrow, el resultado central de este capítulo. Esencialmente nos dice que la paradoja de Condorcet no se debe a ninguna de las propiedades fuertes de la regla de mayoría por pares que, recordemos del teorema de May 2.9, son: la simetría entre los votantes, la neutralidad entre las alternativas y la monotonía. La paradoja va al fondo de la cuestión: con independencia de alternativas irrelevantes, es decir, con independencia entre pares, no hay ninguna función de bienestar social definida en $\mathcal{R}$ que satisfaga una forma mínima de simetría entre votantes (no dictadura) y una forma mínima de monotonía (el criterio de Pareto).

Teorema 2.23. *(Teorema de Imposibilidad de Arrow) Supongamos que el conjunto de alternativas* S *tiene más de tres alternativas diferentes. Entonces, dada una función de bienestar* $F : \mathcal{A} \subset \mathcal{R}^n$ *cumpliendo las Condiciones* $1'$*,* 3 *y* P*, la función* F *nunca podrá satisfacer la Condición 5.*

Demostración. La prueba procederá mediante una investigación detallada de la estructura de la familia de conjuntos decisivos. Lo haremos en una serie de pequeños pasos. Los pasos 1 a 3 muestran que si un subconjunto de votantes es decisivo para algún par de alternativas, entonces es decisivo para todos los pares. Los pasos 4 a 6 establecen algunas propiedades algebraicas de la familia de conjuntos decisivos. Los pasos 7 y 8 las utilizan para demostrar que existe un conjunto decisivo más pequeño formado por un solo votante. Los pasos 9 y 10 demuestran que este votante es un dictador.

Paso 1: Si para algún $\{x, y\} \subset S$, $V \subset I$ es decisivo para x sobre y, entonces, para cualquier alternativa $z \neq x$, V es decisivo para x sobre z. Del mismo modo, para cualquier $z \neq y$, V es decisivo para z sobre y.

Demostramos que si V es decisivo para x sobre y entonces es decisivo para x sobre cualquier $z \neq x$. El razonamiento para z sobre y es idéntico.

Si $z = y$ no hay nada que demostrar. Por tanto, suponemos que $z \neq y$. Consideremos un perfil de preferencias $(\succsim_1, \ldots, \succsim_n) \in \mathcal{A}$ donde

$$\begin{array}{ll} x \succ_i y \succ_i z & \text{para todo } i \in V \text{ y} \\ y \succ_i z \succ_i x & \text{para todo } i \in I \backslash V. \end{array}$$

Entonces, dado que V es decisivo para x sobre y, tenemos que x es socialmente preferido a y, es decir, $xF_p\,(\succsim_1, \ldots, \succsim_n)\,y$. Además, dado que $y \succsim_i z$ para cada $i \in I$, y $F(\cdot)$ satisface la Condición P, se deduce que $yF_p\,(\succsim_1, \ldots, \succsim_1)\,z$. Por lo tanto, por la transitividad de la relación de preferencia social, concluimos que $xF_p\,(\succsim_1, \ldots, \succsim_1)\,z$. Por la Condición 3 de IIA, se deduce que x es socialmente preferida a z siempre que todo votante en V prefiera x a z y todo votante que no esté en V prefiera z a x, con total independencia del perfil concreto escogido. Es decir, V es decisivo para x sobre z.

Paso 2: Si para algún $\{x, y\} \subset S$, $V \subset I$ es decisivo para x sobre y y z es una tercera alternativa, entonces V es decisivo para z sobre w y para w sobre z, donde $w \in S$ es cualquier alternativa distinta de z.

Por el *Paso 1*, V es decisivo para x sobre z y para z sobre y. Pero entonces, aplicando de nuevo el *Paso 1*, esta vez al par $\{x, z\}$ y a la alternativa w, concluimos que V es decisivo para w sobre z. Del mismo modo, aplicando el *Paso 1* a $\{z, y\}$ y w, concluimos que V es decisivo para z sobre w.

Paso 3: Si para algún $\{x, y\} \subset S$, $V \subset I$ es decisivo para x sobre y, entonces V es decisivo.

Esto es una consecuencia inmediata del *Paso 2* y del hecho de que hay alguna alternativa $z \in S$ distinta de x o y. En efecto, tomemos cualquier par $\{v, w\}$. Si $v = z$ o $w = z$, entonces el *Paso 2* implica directamente el resultado. Si $v \neq z$ y $w \neq z$, aplicamos el *Paso 2* para concluir que V es decisivo para z sobre w, y luego el *Paso 1* aplicado al par $\{z, w\}$ para concluir que V es decisivo para v sobre w.

Paso 4: Si $V \subset I$ y $T \subset I$ son decisivos, entonces $V \cap T$ es decisivo.

Tome cualquier triple de alternativas distintas $\{x, y, z\} \in S$ y consideremos un perfil de preferencias $(\succsim_1, \ldots, \succsim_n) \in \mathcal{A}$ donde

$$\begin{aligned}
z \succ_i y \succ_i x \quad & \text{para todo } i \in V \backslash (V \cap T), \\
x \succ_i z \succ_i y \quad & \text{para todo } i \in V \cap T, \\
y \succ_i x \succ_i z \quad & \text{para todo } i \in T \backslash (V \cap T), \\
y \succ_i z \succ_i x \quad & \text{para todo } i \in I \backslash (V \cup T).
\end{aligned} \tag{2.15}$$

Entonces $zF_p(\succsim_1, \ldots, \succsim_n)\, y$ porque V es un conjunto decisivo y porque de 2.15 se tiene que, por la Condición 3 de IIA,

$$\begin{aligned}
z \succ_i y \quad & \text{para todo } i \in V, \\
y \succ_i z \quad & \text{para todo } i \in I \backslash V,
\end{aligned}$$

pues $V = [V \backslash (V \cap T)] \cup [V \cap T]$. Del mismo modo, $xF_p(\succsim_1, \ldots, \succsim_n)\, z$ porque T es un conjunto decisivo y de 2.15 se tiene que, por la Condición 3 de IIA,

$$\begin{aligned}
x \succ_i z \quad & \text{para todo } i \in T, \\
z \succ_i x \quad & \text{para todo } i \in I \backslash T.
\end{aligned}$$

pues $T = [V \cap T] \cup [T \backslash (V \cap T)]$. Por lo tanto, por la transitividad de la preferencia social, tenemos que $xF_p(\succsim_1, \ldots, \succsim_n)\, y$. Utilizando de nuevo el mismo argumento, se deduce que $V \cap T$ es decisivo para x sobre y, ya que, por la Condición 3 de IIA, se observa en 2.15 que

$$\begin{aligned}
x \succ_i z \quad & \text{para todo } i \in V \cap T, \\
z \succ_i x \quad & \text{para todo } i \in I \backslash (V \cap T),
\end{aligned}$$

pues $V \cap T = [V \backslash (V \cap T)] \cup [T \backslash (V \cap T)] \cup [I \backslash (V \cup T)]$. Por lo tanto, por el *Paso 3*, $V \cap T$ es un conjunto decisivo.

Paso 5: Para cualquier subconjunto $V \subset I$, tenemos que o bien V o bien su complementario, $I \backslash V \subset I$, es decisivo.

Tomemos cualquier triple de alternativas distintas $\{x, y, z\} \subset S$ consideremos un perfil de ordenaciones $(\succsim_1, \ldots, \succsim_n) \in \mathcal{A}$ donde

$$\begin{aligned}
x \succ_i z \succ_i y \quad & \text{para todo } i \in V \\
y \succ_i x \succ_i z \quad & \text{para todo } i \in I \backslash V.
\end{aligned}$$

Entonces hay dos posibilidades: o bien $xF_p(\succsim_1, \ldots, \succsim_n)\, y$, en cuyo caso, por la Condición 3 de IIA, V es decisivo para x sobre y (por tanto, por el *Paso 3*, decisivo), o bien $yF(\succsim_1, \ldots, \succsim_n)\, x$. Dado que, por la Condición P, tenemos $xF_p(\succsim_1, \ldots, \succsim_n)\, z$, la transitividad de la relación de preferencia social produce que $yF_p(\succsim_1, \ldots, \succsim_n)\, z$ en este caso. Pero entonces, utilizando de nuevo la Condición 3 de IIA, concluimos que $I \backslash V$ es decisivo para y sobre z (por tanto, por el *Paso 3*, decisivo).

Paso 6: Si $V \subset I$ es decisivo y $V \subset T$, entonces T también es decisivo.

Debido a la Condición P, el conjunto vacío de votantes no puede ser decisivo ya que, si ningún votante prefiere x a y y todos los votantes prefieren y a x, entonces x no es socialmente preferible a y, pues incumpliría esta condición. Por lo tanto, $I \backslash T$ no puede ser decisivo porque, de lo contrario, en el *Paso 4*, $V \cap (I \backslash T) = \varnothing$ sería decisivo. Por lo tanto, por el *Paso 5*, T es decisivo.

Paso 7: Si $V \subset I$ es decisivo e incluye más de un votante, entonces hay un subconjunto estricto $V' \subset V$, $V' \neq V$, tal que V' es decisivo.

Tomemos cualquier $h \in V$. Si $V \backslash \{h\}$ es decisivo, entonces hemos terminado. Supongamos, por lo tanto, que $V \backslash \{h\}$ no es decisivo. Entonces, por el *Paso 5*, $I \backslash (V \backslash \{h\}) = (I \backslash V) \cup \{h\}$ es decisivo. Se deduce, por el *Paso 4*, que $\{h\} = V \cap [(I \backslash V) \cup \{h\}]$ también es decisivo. Por lo tanto, hemos terminado de nuevo ya que, por supuesto, $\{h\}$ es un subconjunto estricto de V.

Paso 8: Hay un votante $h \in I$ tal que $V = \{h\}$ es decisivo.

Esto se consigue iterando el *Paso 7* y teniendo en cuenta, primero, que el conjunto de votantes I es finito y, segundo, que, por la condición P, el conjunto I de todos los votantes es decisivo.

Paso 9: Si $V \subset I$ es decisivo entonces, para cualquier $\{x, y\} \subset S$, V es completamente decisivo para x sobre y.

Ya hemos demostrado que esto ocurre cuando se cumplen las Condiciones 1-5 en la proposición 2.18, sin embargo, ahora lo demostraremos usando las Condiciones $1'$, 3 y P. Queremos demostrar que, para cualquier $T \subset I \backslash V$, x es socialmente preferido a y siempre que cada votante en V prefiera x a y, cada votante en T considere que x es al menos tan bueno como y, y cualquier otro votante prefiere y a x. Para verificar esta propiedad, tomemos una tercera alternativa $z \in S$, distinta de x e y. Por la Condición 3 de IIA, basta con considerar un perfil de preferencias $(\succsim_1, \ldots, \succsim_n) \in \mathcal{A}$ donde

$$\begin{aligned} x \succ_i z \succ_i y \quad & \text{para todo } i \in V, \\ x \succ_i y \succ_i z \quad & \text{para todo } i \in T, \\ y \succ_i z \succ_i x \quad & \text{para todo } i \in I \backslash (V \cup T). \end{aligned}$$

Entonces $x F_p (\succsim_1, \ldots, \succsim_n) z$ porque, por el *Paso 6*, $V \cup T$ es decisivo, y $z F_p (\succsim_1, \ldots, \succsim_n) y$ porque V es decisivo. Por lo tanto, por la transitividad de la preferencia social, tenemos que $x F_p (\succsim_1, \ldots, \succsim_n) y$, como queríamos demostrar.

Paso 10: Si, para algún $h \in I$, $V = \{h\}$ es decisivo, entonces h es un dictador.

Si $\{h\}$ es decisivo, entonces, por el *Paso 9*, $\{h\}$ es completamente decisivo para cualquier x sobre cualquier y. Es decir, si el perfil $(\succsim_1, \ldots, \succsim_n)$ es tal que $x \succ_h y$, entonces $x F_p (\succsim_1, \ldots, \succsim_n) y$. Pero esto es precisamente lo que significa que $h \in I$ sea un dictador, incumpliéndose entonces la Condición 5.

Por lo tanto, siguiendo todos estos pasos hemos comprobado que, efectivamente, una función de bienestar social que satisfaga las Condiciones $1'$, 3 y P nunca podrá cumplir también la Condición 5. □

2.3. Resultados de posibilidad

El resultado del Teorema de Imposibilidad de Arrow es algo desalentador. Demuestra que no debemos esperar que un conjunto de individuos se comporte con el tipo de coherencia que podemos esperar de un único individuo.

No obstante, existe la pregunta de hasta qué punto podemos escapar de la conclusión de la dictadura si relajamos algunas de las exigencias impuestas por el teorema de Arrow. Existen dos debilitamientos principales [13]:

1. Relajamiento de los requisitos de racionalidad impuestos a las preferencias agregadas.
2. Planteamiento la cuestión de la agregación en un ámbito restringido. En particular, consideraremos una restricción (la de los picos) que ha resultado ser significativa y útil en las aplicaciones.

En primer lugar, supongamos que mantenemos las condiciones de Pareto y de IIA, pero permitimos que las preferencias sociales no sean "totalmente racionales". Esta racionalidad total hace referencia al entendimiento que tenemos nosotros de las relaciones de orden. Esta relajación se plasma en la siguiente definición.

Definición 2.24. Supongamos que la relación de preferencia $\succsim$ sobre S es reflexiva y completa. Decimos entonces que:

1. $\succsim$ es cuasitransitiva si la preferencia estricta $\succ$ inducida por $\succsim$ es transitiva.
2. $\succsim$ es acíclica si $\succsim$ tiene un elemento máximo en cada subconjunto finito $S' \subset S$, s decir, $\{x \in S' : x \succsim y \text{ para todo } y \in S'\} \neq \varnothing$

Una relación de preferencia cuasitransitiva es acíclica, pero lo contrario puede no ser cierto. Asimismo, una relación de orden es cuasitransitiva, pero, de nuevo, la inversa puede no ser válida. Por tanto, la condición más débil es la acíclica. Sin embargo, la aciclicidad no es un debilitamiento drástico de la racionalidad: obsérvese, por ejemplo, que los ordenamientos sociales de la paradoja de Condorcet también violan la aciclicidad. No vamos a discutir en detalle las posibilidades que nos abren estos debilitamientos de la racionalidad social, ya que no son muy sustanciales. Remitimos a [19] para una exposición detallada.

En segundo lugar, tenemos el debilitamiento asociado a la restricción del dominio. Este fue desarrollado principalmente por Duncan Black **??**, el cual definió una restricción a los dominios de las funciones de elección social denominada “preferencias de un solo pico”. Según este principio, todas las elecciones tienen una posición predeterminada a lo largo de una línea, lo que les da un ordenamiento lineal. Cada votante tiene un lugar especial que le gusta más a lo largo de esa línea. Su ordenación de las opciones viene determinada por sus distancias a ese punto. Por ejemplo, si se vota sobre dónde poner el volumen de la música, sería razonable suponer que cada votante tiene su propia preferencia de volumen ideal y que, a medida que el volumen sea progresivamente demasiado alto o demasiado bajo, estará cada vez más insatisfecho. Black demostró que sustituyendo el dominio no restringido por las “preferencias de un solo pico” en el teorema de Arrow se elimina la imposibilidad: hay funciones de bienestar social cumpliendo el criterio de Pareto y no dictatoriales que satisfacen el IIA.

Capítulo 3

Teorema de Imposibilidad de Gibbard-Satterthwaite

En las secciones anteriores hemos teorizado sobre las funciones de bienestar social y hemos tratado de averiguar si existe una función cumpliendo una serie de condiciones deseadas. Es así como llegamos al Teorema de Imposibilidad de Arrow.

Las funciones de bienestar, como ya hemos visto, tienen como resultado una relación de orden, es decir, orden de preferencia de las alternativas. Este orden de preferencia es usado después para tomar decisiones, es decir, para escoger la mejor opción. En este apartado nos vamos a centrar precisamente en esto: en tomar decisiones.

Para ello, en esta sección vamos a retomar el concepto de función de elección social o regla de votación tal y como definíamos en 1.9. En este caso, también impondremos unas condiciones similares sobre estas funciones y, de nuevo, obtendremos una conclusión similar a la de la sección anterior, topándonos de nuevo con la imposibilidad de satisfacer una serie de condiciones deseables.

Como resulta evidente, vamos a solicitar condiciones muy similares a las que solicitábamos para las funciones de bienestar. Esto resulta lógico teniendo en cuenta que estos dos conceptos están íntimamente ligados, como ya hemos visto al comienzo de esta memoria en la Sección 1.1. Así pues, definamos la primera condición.

Definición 3.1. La función de elección social $f : \mathcal{A} \to S$ definida como $\mathcal{A} \subset \mathcal{R}^n$ cumple el **criterio de Pareto** si para cualquier perfil de relaciones de orden $(\succsim_1, \ldots, \succsim_n) \in \mathcal{A}$ la elección $f(\succsim_1, \ldots, \succsim_n) \in S$ satisface que, para un par de alternativas $\{x, y\} \subset S$, si $x \succ_i y$ para todo $i \in I$, entonces $y \neq f(\succsim_1, \ldots, \succsim_n)$.

De manera informal, esta condición nos dice que si para un $x \in S$ existe $x \in S$ tal que todos las relaciones de orden individuales prefieren a x antes que a y, entonces y no puede ser escogido como ganador por delante de x. Como podemos observar, este criterio de Pareto es similar al criterio de Pareto definido en 2.2. Antes de seguir con otra condición importante, debemos de definir un concepto previo.

Definición 3.2. La alternativa $x \in S$ **mantiene su posición** del perfil de ordenación $(\succsim_1, \dots, \succsim_n) \in \mathcal{R}^n$ al perfil $(\succsim'_1, \dots, \succsim'_n) \in \mathcal{R}^n$ si

$$x \succsim_i y \Rightarrow x \succsim'_i y$$

para todo $i \in I$ y todo $y \in S$.

En otras palabras, x mantiene su posición de $(\succsim_1, \dots, \succsim_n)$ a $(\succsim'_1, \dots, \succsim'_n)$ si para cada $i \in I$ el conjunto de alternativas inferiores o indiferentes a x expande o se mantiene cuando pasamos de $\succsim_i$ a $\succsim'_i$. Obsérvese que la condición establecida en la definición 3.2 no impone ninguna restricción sobre cómo otras alternativas diferentes de x pueden cambiar su orden mutuo al pasar de $\succsim_i$ a $\succsim'_i$.

Una vez enunciada y explicada este concepto de "mantener la posición", podemos definir la condición citada anteriormente.

Definición 3.3. La función de elección social $f : \mathcal{A} \to S$ definida en $\mathcal{A} \subset \mathcal{R}^n$ es **monótona** si para cualquier par de perfiles de ordenación $(\succsim_1, \dots, \succsim_n) \in \mathcal{A}$, $(\succsim'_1, \dots, \succsim'_n) \in \mathcal{A}$ con la propiedad de que la alternativa escogida $x = f(\succsim_1, \dots, \succsim_n)$ mantenga su posición al pasar de $(\succsim_1, \dots, \succsim_n)$ a $(\succsim'_1, \dots, \succsim'_n)$, tenemos que $f(\succsim'_1, \dots, \succsim'_n) = x$.

Dicho informalmente, la función de elección social es monótona si una alternativa no puede dejar de ser escogida a menos que para algún votante se deteriore su preferencia.

Una clase no muy atractiva de funciones de elección social que tienen las dos propiedades en este dominio son las funciones de elección social dictatoriales.

Definición 3.4. Un votante $d \in I$ es un dictador para la función de elección social $f : \mathcal{A} \to S$, siendo $\mathcal{A} \subset \mathcal{R}^n$, si, para todo perfil $(\succsim_1, \dots, \succsim_n) \in \mathcal{A}$, se tiene que $x = f(\succsim_1, \dots, \succsim_n)$ es una de las alternativas preferidas para $\succsim_d$ en S; es decir,

$$f(\succsim_1, \dots, \succsim_n) \in \{x \in S : x \succsim_d y \ \ \forall y \in S\}.$$

Una función de elección social que admite dictador es una función **dictatorial**.

Como ya vimos en el caso de las funciones de bienestar social, la condición de dictatorialidad no es deseable para tener una buena función de elección social. También hemos dicho que las funciones dictatoriales son una clase de funciones monótonas y de Pareto, Veamos que, efectivamente, cumplen estas dos propiedades para el conjunto universal de relaciones de orden $\mathcal{R}^n$.

Proposición 3.5. *En el dominio $\mathcal{R}^n$, una función de elección social dictatorial es de Pareto y monótona.*

Demostración. Para probar el criterio de Pareto, tomemos cualquier par de alternativas $\{x, y\} \subset S$ siendo x una de las preferidas de $\succsim_d$. Entonces, nunca ocurrirá que $y \succ_i x$ para todo $i \in I$, pues, si $i = d$, entonces $x \succ_d y$.

Para probar la monotonía, simplemente hay que explicitar que si $x = f(\succsim_1, \ldots, \succsim_n)$ (una de las alternativas preferidas por el dictador) mantiene su posición, entonces también se mantendrá como una de las preferidas del dictador. Luego seguiremos teniendo $x = f(\succsim'_1, \ldots, \succsim'_n)$ para cualquier perfil $(\succsim'_1, \ldots, \succsim'_n) \in \mathcal{R}^n$. □

De hecho, en el dominio universal de funciones de elección social que cumplen estas condiciones no podemos obtener nada mejor que las funciones de elección social dictatoriales. Esto demostraremos a continuación, pero antes de esto es importante definir un concepto similar al de mantener la posición.

Definición 3.6. Dado el conjunto finito $S' \subset S$ y el perfil de ordenaciones $(\succsim_1, \ldots, \succsim_n) \in \mathcal{R}^n$, decimos que el perfil $(\succsim'_1, \ldots, \succsim'_n) \in \mathcal{R}^n$ **lleva S' a la cumbre** desde $(\succsim_1, \ldots, \succsim_n)$ si para cada $i \in I$

$$\begin{aligned} &x \succ'_i y && \text{para } x \in S' \text{ y } y \notin S', \\ &x \succsim_i y \Leftrightarrow x \succsim'_i y && \forall x, y \in S'. \end{aligned}$$

De manera informal, la relación de preferencia $\succsim'_i$ se obtiene de $\succsim_i$ simplemente llevando a cada alternativa de S' a la cumbre, preservando el orden interno (débil o estricto) entre estas alternativas. El ordenamiento entre las alternativas que no están en S' es arbitrario. Por ejemplo, si $x \succ_i y \succ_i z \succ_i w$, entonces la relación de preferencia estricta $\succ'_i$ definida por $y \succ'_i w \succ'_i z \succ'_i x$ lleva a $\{y, w\}$ a la cumbre desde $\succsim_i$. Nótese también que si $(\succsim'_1, \ldots, \succsim'_n)$ lleva S' a la cumbre desde $(\succsim_1, \ldots, \succsim_n)$, entonces $x \in S'$ mantiene su posición desde $(\succsim_1, \ldots, \succsim_n)$ a $(\succsim'_1, \ldots, \succsim'_n)$.

Una vez explicada el concepto de "llevar a la cumbre" para un conjunto de alternativas, podemos demostrar el resultado que antes enunciábamos informalmente.

Proposición 3.7. *Supongamos que el número de alternativas es al menos tres y que el dominio de los perfiles de ordenación admisibles es $\mathcal{A} = \mathcal{R}^n$ o $\mathcal{A} = \mathcal{P}^n$. Entonces, cada función de elección social $f : \mathcal{A} \to S$ que cumpla el criterio de Pareto y sea monótona es dictatorial.*

Demostración. La prueba del resultado se obtendrá como corolario del Teorema de Imposibilidad de Arrow 2.23. Para ello, procederemos a construir una función de bienestar social $F(\cdot)$ que racionalizará $f(\succsim_1, \ldots, \succsim_n)$ para cada perfil $(\succsim_1, \ldots, \succsim_n) \in \mathcal{A}$. Veremos más adelante qué significa esto. Después, mostraremos que $F(\cdot)$ satisface los supuestos del Teorema de Imposibilidad de Arrow, con lo que se obtiene la conclusión de la dictadura. Para llevar acabo esta prueba, al igual que hicimos con la de Arrow, procederemos empleando una serie de pasos.

Paso 1. Si ambos perfiles $(\succsim_1', \ldots, \succsim_n') \in \mathcal{A}$ y $(\succsim_1'', \ldots, \succsim_n'') \in \mathcal{A}$ llevan $S' \subset S$ a la cumbre desde $(\succsim_1, \ldots, \succsim_n)$, entonces $f(\succsim_1', \ldots, \succsim_n') = f(\succsim_1'', \ldots, \succsim_n'')$.

Para cada $i \in I$ y $x \in S'$ tenemos que

$$\left\{y \in S : x \succsim_i' y\right\} = \left\{y \in S : x \succsim_i'' y\right\} = \{y \in S : x \succsim_i y\} \cup S \backslash S'.$$

Por el criterio de Pareto, $f(\succsim_1', \ldots, \succsim_n') \in S'$. Por lo que, $f(\succsim_1', \ldots, \succsim_n') \in S'$ mantiene su posición pasando de $(\succsim_1', \ldots, \succsim_n')$ a $(\succsim_1'', \ldots, \succsim_n'')$. Por tanto, por la monotonía de $f(\cdot)$, concluimos que $f(\succsim_1', \ldots, \succsim_n') = f(\succsim_1'', \ldots, \succsim_n'')$.

Paso 2. Definición de $F(\succsim_1, \ldots, \succsim_n)$.

Para cada perfil $(\succsim_1, \ldots, \succsim_n) \in \mathcal{A}$ definimos una relación binaria $F(\succsim_1, \ldots, \succsim_n)$ en S. Concretamente, sea $xF(\succsim_1, \ldots, \succsim_n)\, y$, si $x = y$ o si $x = f(\succsim_1', \ldots, \succsim_n')$ cuando $(\succsim_1', \ldots, \succsim_n') \in \mathcal{A}$ es cualquiera de los perfiles que lleva $\{x, y\} \subset S$ a la cumbre desde el perfil $(\succsim_1, \ldots, \succsim_n)$. En el *Paso 1* esto está bien definido, es decir, es independiente del perfil particular $(\succsim_1', \ldots, \succsim_n')$ escogido.

Paso 3. Para cada perfil $(\succsim_1, \ldots, \succsim_n) \in \mathcal{A}$, $F(\succsim_1, \ldots, \succsim_n)$ es una relación de orden. No sólo eso, sino que también se cumple que $F(\succsim_1, \ldots, \succsim_n) \in \mathcal{P}$; es decir, no hay dos alternativas distintas que sean socialmente indiferentes. Para ver que es una relación de orden probaremos primero que es conexa y después que es transitiva, es decir, probaremos los axiomas 1.1 y 1.2.

1. **Axioma de conexidad**.

   Porque $f(\cdot)$ cumple el criterio de Pareto, se tiene que, cuando $(\succsim_1', \ldots, \succsim_n')$ lleva $\{x, y\}$ a la cumbre desde $(\succsim_1, \ldots, \succsim_n)$, debemos tener $f(\succsim_1', \ldots, \succsim_n') \in \{x, y\}$. Por lo tanto, concluimos que $xF(\succsim_1, \ldots, \succsim_n)\, y$ o $yF(\succsim_1, \ldots, \succsim_n)\, x$, pero, por el *Paso 1*, no ocurrirán las dos cosas la vez (a no ser que $x = y$). En particular, $F(\succsim_1, \ldots, \succsim_n)$ es conexa, es decir, satisface el axioma 1.1.

2. **Axioma de transitividad**.

   Para verificar la transitividad, supongamos que $xF(\succsim_1, \ldots, \succsim_n)\, y$ e $yF(\succsim_1, \ldots, \succsim_n)\, z$. Podemos asumir que las tres alternativas $\{x, y, z\}$ son distintas. Sea $(\succsim_1'', \ldots, \succsim_n'') \in \mathcal{A}$ un perfil que lleva $\{x, y, z\}$ a la cumbre desde $(\succsim_1, \ldots, \succsim_n)$. Por el hecho de que $f(\cdot)$ cumple el criterio de Pareto, tenemos que $f(\succsim_1'', \ldots, \succsim_n'') \in \{x, y, z\}$.

   Supongamos que tenemos $y = f(\succsim_1'', \ldots, \succsim_n'')$. Considérese un perfil $(\succsim_1', \ldots, \succsim_n') \in \mathcal{A}$ que lleva $\{x, y\}$ a la cumbre desde $(\succsim_1'', \ldots, \succsim_n'')$. Debido a que y mantiene su posición de $(\succsim_1'', \ldots, \succsim_n'')$ a $(\succsim_1', \ldots, \succsim_n')$, se sigue de la monotonía de la función que $f(\succsim_1', \ldots, \succsim_n') = y$.

Pero $(\succsim_1', \ldots, \succsim_n')$ también lleva $\{x, y\}$ a la cumbre desde $(\succsim_1, \ldots, \succsim_n)$: el orden relativo entre x e y, las dos alternativas en la cumbre, las preferidas, no se ha alterado en ninguna preferencia individual al pasar de $(\succsim_1, \ldots, \succsim_n)$ a $(\succsim_1', \ldots, \succsim_n')$. Luego concluimos que $yF(\succsim_1, \ldots, \succsim_n)x$, lo que contradice la suposición de que $xF(\succsim_1, \ldots, \succsim_n)y$, $x \neq y$. Así pues, $y \neq f(\succsim_1'', \ldots, \succsim_n'')$. De manera similar, obtenemos que $z \neq f(\succsim_1'', \ldots, \succsim_n'')$. Sólo tenemos que repetir el mismo argumento utilizando el par $\{y, z\}$.

La única posibilidad restante es $x = f(\succsim_1'', \ldots, \succsim_n'')$. Por tanto, sea $(\succsim_1', \ldots, \succsim_n') \in \mathcal{A}$ la relación que lleva $\{x, z\}$ a la cumbre desde $(\succsim_1'', \ldots, \succsim_n'')$. Como x mantiene su posición desde $(\succsim_1'', \ldots, \succsim_n'')$ a $(\succsim_1', \ldots, \succsim_n')$, se tiene que $x = f(\succsim_1', \ldots, \succsim_n')$. Pero $(\succsim_1', \ldots, \succsim_n')$ también lleva $\{x, z\}$ a la cumbre desde $(\succsim_1, \ldots, \succsim_n)$. Así, $xF(\succsim_1, \ldots, \succsim_n)z$, y se satisface la transitividad.

Paso 4: La función de bienestar social $F : \mathcal{A} \to \mathcal{P}$ racionaliza $f : \mathcal{A} \to S$; esto significa que, para cada perfil $(\succsim_1, \ldots, \succsim_n) \in \mathcal{A}$, $f(\succsim_1, \ldots, \succsim_n)$ es la alternativa preferida para $F(\succsim_1, \ldots, \succsim_n)$ en S.

Esto es natural teniendo en cuenta que $F(\cdot)$ se ha construido partiendo de $f(\cdot)$. Denótese $x = f(\succsim_1, \ldots, \succsim_n)$ y sea $y \neq x$ cualquier otra alternativa. Considérese el perfil de ordenaciones $(\succsim_1', \ldots, \succsim_n') \in \mathcal{P}^n$ que lleva $\{x, y\}$ a la cumbre desde $(\succsim_1, \ldots, \succsim_n)$. Como x mantiene la posición de $(\succsim_1, \ldots, \succsim_n)$ a $(\succsim_1', \ldots, \succsim_n')$, tenemos $x = f(\succsim_1', \ldots, \succsim_n')$. Así pues, $xF(\succsim_1, \ldots, \succsim_n)y$.

Paso 5: La función de bienestar social $F : \mathcal{A} \to \mathcal{P}$ es de Pareto.

Es claro que si $x \succ_i y$ para todo $i \in I$, entonces, por el criterio de Pareto para funciones de elección social y que cumple $f(\cdot)$, debemos tener $x = f(\succsim_1', \ldots, \succsim_n')$ siempre que $(\succsim_1', \ldots, \succsim_n')$ lleve $\{x, y\}$ a la cumbre de $(\succsim_1, \ldots, \succsim_n)$. Así, $xF(\succsim_1, \ldots, \succsim_n)y$ y, por el *Paso 3*, podemos concluir que $xF_p(\succsim_1, \ldots, \succsim_n)y$.

Paso 6. La función de bienestar social $F : \mathcal{A} \to \mathcal{P}$ satisface la condición de IIA.

Esto se sigue del *Paso 1*. Supongamos que $(\succsim_1, \ldots, \succsim_n) \in \mathcal{A}$ y $(\succsim_1', \ldots, \succsim_n') \in \mathcal{A}$ ordenan de la misma forma $\{x, y\}$ para cada $i \in I$ (esto es, para cada $i \in I$, $x \succsim_i y$ si y sólo si $x \succsim_i'$ y). Supongamos que $(\succsim_1'', \ldots, \succsim_n'') \in \mathcal{A}$ lleva $\{x, y\}$ a la cumbre de $(\succsim_1, \ldots, \succsim_n)$ y que $x = f(\succsim_1'', \ldots, \succsim_n'')$. Entonces $xF(\succsim_1, \ldots, \succsim_n)y$. Pero $(\succsim_1'', \ldots, \succsim_n'')$ también lleva $\{x, y\}$ a la cumbre de $(\succsim_1', \ldots, \succsim_n')$. Así, $xF(\succsim_1', \ldots, \succsim_n')y$, como queríamos probar.

Paso 7. La función de elección social $f : \mathcal{A} \to S$ es dictatorial.

Por el Teorema de Imposibilidad de Arrow 2.23, existe un votante $d \in I$ tal que, para cada perfil $(\succsim_1, \ldots, \succsim_n) \in \mathcal{A}$, tenemos que $xF_p(\succsim_1, \ldots, \succsim_n)y$ siempre que $x \succ_d y$. Por tanto, $f(\succsim_1, \ldots, \succsim_n)$, que por el *Paso 4* es la alternativa preferida para $F(\succsim_1, \ldots, \succsim_n)$ en S; también debe ser la alternativa más preferida para d. Es decir, $f(\succsim_1, \ldots, \succsim_n) \succsim_d x$ para cada $x \in S$. Por lo tanto, concluimos que d es un dictador. □

Esta proposición que acabamos de ver es el preámbulo del teorema que da nombre a esta sección: el Teorema de Imposibilidad de Gibbard-Satterthwaite. Pero, como también pasaba con la anterior proposición, antes de enunciar y demostrar este teorema, es necesario definir un último concepto.

Definición 3.8. Sea $f : \mathcal{P}^n \longrightarrow S$ una función de elección social. Diremos que $f(\cdot)$ es **no manipulable** si satisface que

$$f\left(\succsim_1, \ldots, \succsim_{i-1}, \succsim_i, \succsim_{i+1}, \ldots, \succsim_n\right) \succsim_i f\left(\succsim_1, \ldots, \succsim_{i-1}, \succsim_i', \succsim_{i+1}, \ldots, \succsim_n\right)$$

para cada $i \in I$, cada $\succsim_i' \in \mathcal{P}$ y cada perfil $(\succsim_1, \ldots, \succsim_n) \in \mathcal{P}^n$.

Por lo tanto, ya estamos en condiciones de enunciar y de demostrar el Teorema de Imposibilidad de Gibbard-Satterthwaite [11] [18].

Teorema 3.9. *(Teorema de Imposibilidad de Gibbard-Satterthwaite) Supongamos que el número de alternativas es al menos tres y que $f : \mathcal{P}^n \to S$ es una función de elección social que es de Pareto y satisface la condición de no manipulabilidad. Entonces $f(\cdot)$ es dictatorial.*

Demostración. Por lo visto en la proposición 3.7 es suficiente con ver que $f : \mathcal{P}' \to S$ debe ser monótona.

Supongamos que no es así. Entonces, sin pérdida de generalidad, podemos suponer que, para algún votante d, existen preferencias $\succsim_i \in \mathcal{P}$ para los votantes $i \neq d$, y preferencias $\succsim_d''$, $\succsim_d''' \in \mathcal{P}$ para el votante d, tales que, denotando

$$x = f\left(\succsim_1, \ldots, \succsim_{d-1}, \succsim_d'', \succsim_{d+1}, \ldots, \succsim_n\right)$$

y

$$y = f\left(\succsim_1, \ldots, \succsim_{d-1}, \succsim_d''', \succsim_{d+1}, \ldots, \succsim_n\right),$$

tenemos que $x \succsim_d'' z$ implica que $x \succsim_d''' z$ para cada $z \in S$, y, aún así, $y \neq x$.

Hay dos posibilidades: o bien $y \succ_d'' x$ o bien $x \succsim_d'' y$.

1. Si $y \succ_d'' x$ entonces la condición de no manipulabilidad se viola para la relación de preferencia "verdadera" $\succsim_d = \succsim_d''$ y la falsa representación $\succsim_d' = \succsim_d'''$.

2. Si $x \succsim_d'' y$, entonces $x \succsim_d''' y$. Por lo tanto, como no hay dos alternativas distintas que puedan ser indiferentes, $x \succ_d''' y$. Pero si $x \succ_d''' y$ entonces la condición de no manipulabilidad se viola para la relación de preferencia "verdadera" $\succsim_d = \succsim_d'''$ y la falsa representación $\succsim_d' = \succsim_d''$.

Por lo tanto, la suposición que hacíamos al principio nos lleva a un absurdo y podemos concluir que la función de elección social $f(\cdot)$ es monótona y, por lo tanto, estamos en las condiciones de aplicar la proposición 3.7 y concluir que $f(\cdot)$ es dictatorial. $\square$

Capítulo 4

Experimento práctico

En esta sección final trataremos de complementar el estudio teórico sobre el Teorema de Imposibilidad de Arrow realizado en el capítulo 2. Para ello, vamos a analizar el comportamiento de las funciones de bienestar explicadas en la Sección 1.2 (exceptuando la regla de mayoría por pares, que fue definida para su uso teórico a lo largo de la memoria) con respecto a las Condiciones 3 y P, es decir, con respecto a la independencia de alternativas irrelevantes y al criterio de Pareto. Cabe decir que hemos intentado replicar las pruebas que se realizan en el artículo [15].

4.1. Exposición del escenario

En primer lugar, es necesario indicar que en el artículo [15] realizan un estudio práctico con el que generan perfiles de preferencias reales. Para hacer esto, efectúan un cuestionario a un número determinado de individuos, los cuales establecen sus niveles de preferencia con respecto a unos tipos de tazas. Ésto lo emplean para, a través de un método de regresión, generar unas funciones de utilidad con las que producir las relaciones de orden. Además de esto, también utilizan relaciones de orden totalmente aleatorias y comparan los resultados.

En nuestro caso, ningún estudio será llevado a cabo: nosotros emplearemos únicamente perfiles de preferencia totalmente aleatorios. Esto lo hacemos porque un estudio tan amplio se sale del alcance del proyecto y con los perfiles totalmente aleatorios podemos estudiar el peor de los casos, pues, como se observa en el citado artículo, las funciones de bienestar social responden mucho mejor a perfiles de relaciones de preferencia con correlación entre las preferencias individuales.

Así pues, se compararon las funciones de agregación utilizando perfiles de preferencia con un número variable de votantes (de 3 a 15) y de alternativas (de 3 a 6). Además, el experimento se repitió un total de 1000 veces para poder asegurarnos de que los resultados son robustos y, por lo tanto, fiables.

4.2. Comprobaciones

Como ya hemos dicho, hemos comprobado que se cumplen las condiciones de independencia de alternativas irrelevantes y de Pareto. Es importante decir que la condición de no-dictadura depende sólo de la función de agregación y las funciones elegidas para este trabajo satisfacen todas esta condición, por eso no realizamos condición alguna sobre esta condición.

Sin embargo, las condiciones de Pareto e IIA dependen del perfil de preferencias específico y por eso las ponemos a prueba. La condición de Pareto se comprobó encontrando primero las preferencias por parejas que compartían todos los individuos. Si estas preferencias unánimes se encontraban también en la preferencia del grupo, entonces el perfil de preferencia satisfacía la condición.

La condición de IIA se evaluó mediante un procedimiento de eliminación y otro de inclusión. El procedimiento de eliminación consistía en calcular primero la ordenación social para el perfil de preferencias original. A continuación, se eliminó un subconjunto de alternativas del conjunto original. El perfil de preferencias se actualizaba según las funciones de utilidad de los individuos y se calculaba una nueva clasificación de grupo. Si la posición relativa de las alternativas originales (o restantes) en la nueva clasificación de grupo no cambiaba con respecto a la clasificación original, el escenario de preferencias satisfacía la condición de IIA. Esto se repitió para cada posible subconjunto de alternativas del conjunto original. El procedimiento de inclusión fue similar, pero se añadieron alternativas adicionales. En concreto, se añadieron hasta 27 alternativas al espacio con el fin de imitar totalmente lo aplicado en el artículo [15].

Veamos ahora un ejemplo de cada método de comprobación que realizamos. Comencemos probando la comprobación de IIA empleando el método de eliminación. Para ello, pondremos a prueba la regla de Borda.

Ejemplo 4.1. Veamos cómo falla la regla de Borda cuando, para comprobar, empleamos el método de eliminación (ya hemos visto en el ejemplo 2.14 como Borda no satisface la condición de IIA). Tomemos un conjunto de alternativas $S = \{x, y, z\}$. Supongamos que las relaciones de orden son:

1. Votante 1: $x \succsim_1 y \succsim_1 z$. Es decir, el orden del votante será xyz.
2. Votante 2: $x \succsim_2 y \succsim_2 z$. Es decir, el orden del votante será xyz.
3. Votante 3: $z \succsim_3 x \succsim_3 y$. Es decir, el orden del votante será zxy.

Como estamos hablando de la regla de Borda, el vector de puntuación para tres alternativas será $\boldsymbol{\alpha} = (2, 1, 0)$. Por lo tanto, tenemos que $\boldsymbol{\alpha}(x) = 5$, $\boldsymbol{\alpha}(y) = 2$ y $\boldsymbol{\alpha}(z) = 2$. Así pues, $y \sim z$. Si eliminamos la alternativa x, tenemos que las relaciones de orden son:

1. Votante 1: $y \succsim_1 z$. Es decir, el orden del votante será yz.

2. Votante 2: $y \succsim_2 z$. Es decir, el orden del votante será yz.

3. Votante 3: $z \succsim_3 y$. Es decir, el orden del votante será zy.

En este caso, el vector de puntuación será $\boldsymbol{\alpha}' = (1, 0)$. Por tanto, tenemos que $\boldsymbol{\alpha}'(y) = 2$ y $\boldsymbol{\alpha}'(z) = 1$, lo cual implica que $y \succ z$, a diferencia del caso original donde veíamos que $y \sim z$.

Veamos ahora un ejemplo comprobando IIA con el método de inclusión y obteniendo el fallo.

Ejemplo 4.2. Veamos, en este caso, cómo falla la regla de Copeland cuando, para comprobar, empleamos el método de inclusión. Tomemos, otra vez, un conjunto de tres alternativas $S = \{x, y, z\}$. Supongamos que las relaciones de orden son las mismas que en el ejemplo anterior. Como es Copeland, no habrá un vector de puntuación, por el contrario, tendremos la siguiente tabla de enfrentamientos:

Mayoría por pares			
	$x?y$	$x?z$	$y?z$
Votante 1	x	x	y
Votante 2	x	x	y
Votante 3	x	z	z
Resultados	x	x	y

Por lo tanto, tenemos que las puntuaciones son $\mathrm{C}^{\frac{1}{2}}(x) = 2$, $\mathrm{C}^{\frac{1}{2}}(y) = 1$ y $\mathrm{C}^{\frac{1}{2}}(z) = 0$. Ahora bien, añadamos una nueva alternativa a de la siguiente forma:

1. Votante 1: $x \succsim_1 y \succsim_1 z \succsim_1 a$. Es decir, el orden del votante será $xyza$.

2. Votante 2: $a \succsim_2 x \succsim_2 y \succsim_2 z$. Es decir, el orden del votante será $axyz$.

3. Votante 3: $z \succsim_3 a \succsim_3 x \succsim_3 y$. Es decir, el orden del votante será $zaxy$.

En este caso, tenemos la tabla de enfrentamientos:

Mayoría por pares						
	$x?y$	$x?z$	$x?a$	$y?z$	$y?a$	$z?a$
Votante 1	x	x	x	y	y	z
Votante 2	x	x	a	y	a	a
Votante 3	x	z	a	z	a	z
Resultados	x	x	a	y	a	z

Por lo tanto, tenemos que las puntuaciones son $\mathrm{C}^{\frac{1}{2}}(x) = 2$, $\mathrm{C}^{\frac{1}{2}}(y) = 1$, $\mathrm{C}^{\frac{1}{2}}(z) = 1$ y $\mathrm{C}^{\frac{1}{2}}(a) = 2$. Por lo tanto, la condición de IIA falla puesto que antes teníamos que $y \succ z$ y, sin embargo, ahora tenemos que $y \sim z$.

Por último, vamos a poner un ejemplo de la comprobación del criterio de Pareto. Esto lo haremos sobre la regla de veto, aunque sería análoga la comprobación en el caso de mayoría simple, por la antisimetría que comparten y que ya comentamos cuando explicamos esta regla de agregación.

Ejemplo 4.3. Veamos cómo incumple el criterio de Pareto la regla de veto. Tomemos un conjunto de alternativas $S = \{x, y, z\}$. Supongamos que las relaciones de orden son:

1. Votante 1: $z \succsim_1 y \succsim_1 x$. Es decir, el orden del votante será zyx.

2. Votante 2: $z \succsim_2 y \succsim_2 x$. Es decir, el orden del votante será zyx.

3. Votante 3: $z \succsim_3 y \succsim_3 x$. Es decir, el orden del votante será zyx.

Como es Borda, el vector de puntuación para tres alternativas será $\boldsymbol{\alpha} = (1, 1, 0)$. Por lo tanto tenemos que $\boldsymbol{\alpha}(x) = 0$, $\boldsymbol{\alpha}(y) = 3$ y $\boldsymbol{\alpha}(z) = 3$. Así pues, $y \sim z$. Sin embargo, vemos que en todas las ordenaciones individuales se cumple que $z \succ y$, por lo que se incumple el criterio de Pareto.

4.3. Resultados y conclusiones

Así pues, las comprobaciones implementadas en el apartado anterior se las hemos aplicado a cuatro funciones de bienestar: veto, mayoría simple, Borda y Copeland. En la siguiente tabla encontramos los resultados obtenidos en cada uno de los métodos.

Cumplimiento reglas		
	Pareto	IIA
Veto	0.949	0.418
Mayoría simple	0.950	0.418
Borda	1	0.599
Copeland	1	0.720

Como ya hemos dicho en este capítulo, hemos basado nuestro análisis de las condiciones de Pareto e IIA en los análisis realizados en el artículo [15]. Sin embargo, se puede observar que, si bien los resultados del cumplimiento del criterio de Pareto son bastante similares, los resultados de IIA son, sin duda, diferentes. Mientras que en este estudio nos encontramos con un 1,2 % de cumplimiento de la condición de IIA para veto y pluralidad, un 11,8 % para la de Borda y un 15,1 % para la de Copeland; vemos como en nuestro estudio los resultados son bien distintos, siendo el cumplimiento de veto y pluralidad del 41,8 %, el de Borda un 59,9 % y el de Copeland un 72 %.

Esto no significa que ninguna de las dos comprobaciones estén mal per se, sino que, a la hora de codificar la comprobación de IIA en el artículo citado, se hizo una aproximación más restrictiva y estricta de lo que entendemos como IIA que en nuestro caso. Aunque nosotros, como se puede ver comparando las explicaciones de las comprobaciones seguidas, intentamos seguir la misma aproximación, al no ser accesible el código del estudio no podemos determinar exactamente las causas de tal diferencia.

Sin embargo, pese a que los órdenes de magnitud son claramente diferentes, se puede observar aún así la misma relación entre las reglas que en el artículo. Es decir, en ambos casos Copeland es la más efectiva en el cumplimiento de IIA, le sigue de cerca Borda y, por último, veto y mayoría simple tienen la misma efectividad y cierran la lista con el resultado más bajo.

Esto tiene mucho sentido, pues, como vimos al construir Copeland, ésta es una función de bienestar de comparación por pares, luego es natural que su rendimiento en la condición de IIA sea correcto, pues, recordemos, esta condición mide la independencia entre pares. Por otro lado, es muy lógico que veto y mayoría simple tengan resultados muy parecidos por la antisimetría que guardan entre sí estos dos métodos, pues, como ya vimos, veto se considera la contraparte de mayoría simple: en lugar de premiar la alternativa que queda más veces primera, se castiga a la que queda más veces última. Por lo tanto, es lógico que el cumplimiento de las condiciones sea similar.

Con todo esto concluimos que, si bien el Teorema de Imposibilidad de Arrow arroja un resultado poco alentador, podemos tratar de buscar reglas que cumplan las condiciones deseadas el mayor número de veces posibles. Es importante recalcar que los casos que hemos estudiado en este capítulo son con perfiles de ordenación totalmente aleatorios, es decir, hemos estudiado el peor de los casos. En el artículo de referencia [15] se puede observar como, para unos perfiles de preferencia reales, es decir, perfiles con una cierta correlación; los resultados son claramente mejores, sobre todo para Copeland y Borda.

Bibliografía

[1] Kenneth J Arrow. *Social Choice and Individual Values*. Vol. 12. Yale University Press, 1963.

[2] Kenneth J Arrow. *Social Choice and Individual Values*. Vol. 12. Yale university press, 2012.

[3] J. Bartholdi, C. A. Tovey y M. A. Trick. "Voting schemes for which it can be difficult to tell who won the election". En: *Social Choice and Welfare* 6 (1989), págs. 157-165. DOI: 10.1007/BF00303169.

[4] Abram Bergson. "A Reformulation of Certain Aspects of Welfare Economics". En: *The Quarterly Journal of Economics* 52.2 (1938), págs. 310-334. DOI: doi.org/10.2307/1881737. (Visitado 17-06-2022).

[5] Julian H. Blau. "The existence of social welfare functions". En: *Econometrica* 25 (1957), pág. 302.

[6] Jean-Charles de Borda. "Mémoire sur les élections au scrutin". En: *Histoire de l'Académie Royale des Sciences* (1781).

[7] Walter Bossert y John A. Weymark. "Utility in Social Choice". En: *Handbook of Utility Theory: Volume 2 Extensions*. Ed. por Salvador Barberà, Peter J. Hammond y Christian Seidl. Springer US, 2004, págs. 1099-1177. DOI: 10.1007/978-1-4020-7964-1_7.

[8] Josep M. Colomer. "Ramon Llull: from 'Ars electionis' to social choice theory". En: *Social Choice and Welfare* 40 (2013), págs. 317-328. DOI: 10.1007/s00355-011-0598-2.

[9] Jean Antoine Marie Nicolas Caritat Condorcet. "Essai sur l'application de l'analyse à la probabilité des décisions, rendues à la pluralité des voix". En: (1785).

[10] Marc Fleurbaey y Maurice Salles. "A Brief History of Social Choice and Welfare Theory". En: *Conversations on Social Choice and Welfare Theory - Vol. 1*. Ed. por Marc Fleurbaey y Maurice Salles. Springer International Publishing, 2021, págs. 1-16. DOI: 10.1007/978-3-030-62769-0_1.

[11] Allan Gibbard. "Manipulation of Voting Schemes: A General Result". En: *Econometrica* 41 (1973), págs. 587-601.

[12] A.R. Karlin e Y. Peres. *Game Theory, Alive*. American Mathematical Society, 2017.

[13] Andreu Mas-Colell, Michael Dennis Whinston, Jerry R Green y col. *Microeconomic theory*. Vol. 1. Oxford university press New York, 1995.

[14] Kenneth O. May. "A Set of Independent Necessary and Sufficient Conditions for Simple Majority Decision". En: *Econometrica* 20 (1952), pág. 680.

[15] Christopher McComb, Kosa Goucher-Lambert y Jonathan Cagan. "Impossible by design? Fairness, strategy, and Arrow's impossibility theorem". En: *Design Science* 3 (2017), e2. DOI: 10.1017/dsj.2017.1.

[16] Efe A Ok. *Real analysis with economic applications*. Princeton University Press, 2007.

[17] J. Rothe e I. Rothe. *Economics and Computation: An Introduction to Algorithmic Game Theory, Computational Social Choice, and Fair Division*. Springer Berlin Heidelberg, 2015.

[18] Mark Allen Satterthwaite. "Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions". En: *Journal of Economic Theory* 10 (1975), págs. 187-217. ISSN: 0022-0531. DOI: https://doi.org/10.1016/0022-0531(75)90050-2.

[19] Amartya Sen. *Collective Choice and Social Welfare*. Holden Day, 1970.

[20] Amartya Sen. "Social Choice". En: *The New Palgrave Dictionary of Economics*. Palgrave Macmillan UK, 2017, págs. 1-20. DOI: 10.1057/978-1-349-95121-5_1856-2.