20 Bargaining

20.1 Bargaining Problems

Bargaining theory investigates how agents should cooperate when non-cooperation may result in outcomes that are Pareto dominated. Formally, a (two-player) *bargaining problem* is a pair (F, d) where $F \subseteq \mathbb{R}^2$ is a convex set of *feasible outcomes* and $d \in F$ is a *disagreement point* that results if players fail to agree on an outcome. Here, convexity corresponds to the assumption that any lottery over feasible outcomes is again feasible. A *bargaining solution* then is a function that assigns to every bargaining problem (F, d) a unique element of F.

The most basic example of a bargaining problem is the so-called ultimatum game given by $F = \{(v_1, v_2) \in \mathbb{R}^2 : v_1 + v_2 \leqslant 1\}$ and $d = (0, 0)$, in which two players receive a fixed amount of payoff if they can agree on a way to divide this amount among themselves. This game has many equilibria when viewed as a normal-form game, since disagreement results in a payoff of zero to both players. Players' preferences regarding these equilibria differ, and bargaining theory tries to answer the question which equilibrium should be chosen. More generally, a two-player normal-form game with payoff matrices $P, Q \in \mathbb{R}^{m \times n}$ can be interpreted as a bargaining problem where $F = \text{conv}(\{(p_{ij}, q_{ij}) : i = 1, \dots, m,\ j = 1, \dots, n\})$, $d_1 = \max_{x \in X} \min_{y \in Y} p(x, y)$, and $d_2 = \max_{y \in Y} \min_{x \in X} q(x, y)$, given that $(d_1, d_2) \in F$. Here, $\text{conv}(S)$ denotes the convex hull of set S.

Two kinds of approaches to bargaining exist in the literature: a strategic one that considers iterative procedures resulting in an outcome in F, and an axiomatic one that tries to identify bargaining solutions that possess certain desirable properties. We will focus on the axiomatic approach in this lecture.

20.2 Nash's Bargaining Solution

For a given bargaining problem (F, d), Nash proposed to

$$\begin{aligned} \text{maximize} \quad & (v_1 - d_1)(v_2 - d_2) \\ \text{subject to} \quad & v \in F \\ & v \geqslant d. \end{aligned} \tag{20.1}$$

The objective function of this optimization problem is strictly quasi-concave and therefore has a unique maximum. Formally, a function $f : S \to \mathbb{R}$ defined on a convex set S is *strictly quasi-concave* if for all $x, y \in S$ with $x \neq y$ and every $\delta \in (0, 1)$, $f(x) \geqslant f(y)$ implies $f((1 - \delta)x + \delta y) > f(y)$. In other words, strict quasi-concavity means that the interior of any line segment joining points on two level sets of f lies strictly above

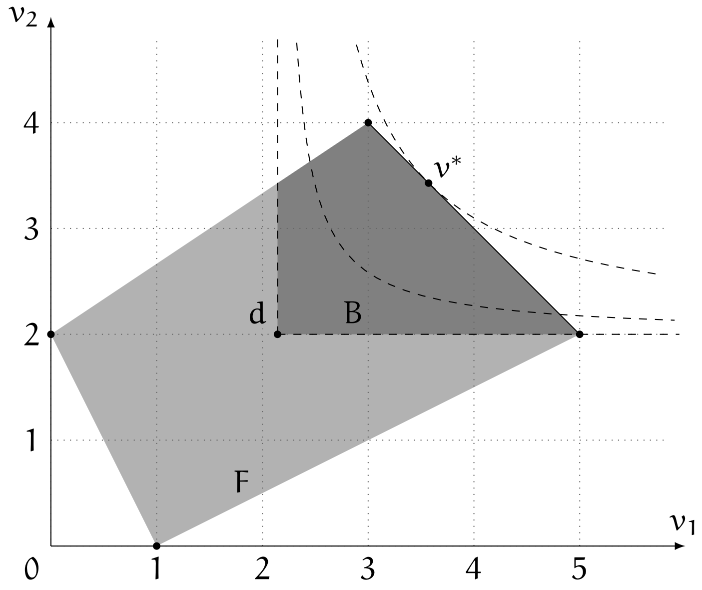

Figure 20.1: Illustration of the Nash bargaining solution

the level set corresponding to the lower value of the function. The objective function of (20.1) satisfies this criterion because its level sets are rectangular hyperbolae with horizontal and vertical asymptotes. Optimization problem (20.1) thus defines a bargaining solution, the so-called *Nash bargaining solution*.

Consider for example the two-player game with payoff matrices

$$P = \begin{pmatrix} 0 & 5 \\ 3 & 1 \end{pmatrix} \qquad \text{and} \qquad Q = \begin{pmatrix} 2 & 2 \\ 4 & 0 \end{pmatrix}.$$

In this game, the row player can guarantee a payoff of 15/7 by playing the two rows with probabilities 2/7 and 5/7, respectively. The column player can guarantee a payoff of 2 by playing the left column. The bargaining problem corresponding to this game is shown in Figure 20.1. The set F is the convex hull of the four payoff vectors $(0, 2)$, $(5, 2)$, $(3, 4)$, and $(1, 0)$, and it contains the feasible set $B = \{v \in F : v \geqslant d\}$ of (20.1). The disagreement point is $d = (15/7, 2)$. Level sets of the objective function corresponding to values 0 and 1 and to the optimal value are drawn as dashed curves. The Nash bargaining solution v^* is the unique point in the intersection of F with the optimal level set.

To compute v^*, we first observe that $v^* \in \{(v_1, v_2) : v_2 = 7 - v_1, 3 \leqslant v_1 \leqslant 5\}$. The objective function becomes

$$(v_1 - d_1)(v_2 - d_2) = (v_1 - \frac{15}{7})(5 - v_1) = \frac{50}{7}v_1 - v_1^2 - \frac{75}{7},$$

and has a stationary point if $50/7 - 2v_1 = 0$. We obtain $v^* = (25/7, 24/7)$, which is indeed a maximum.

While it is not obvious that maximizing the product of the excess of the two players is a good idea, it turns out that the Nash bargaining solution can be characterized using a set of simple axioms. Bargaining solution f is

- *Pareto efficient* if $f(F, d)$ is not Pareto dominated in F for any bargaining problem (F, d);
- *symmetric* if $(f(F, d))_1 = (f(F, d))_2$ for every bargaining problem (F, d) such that $(y, x) \in F$ whenever $(x, y) \in F$ and $d_1 = d_2$;
- *invariant under positive affine transformations* if $f(F', d') = \alpha \circ f(F, d) + \beta$ for any $\alpha, \beta \in \mathbb{R}^2$ with $\alpha > 0$ and any two bargaining problems (F, d) and (F', d') such that $F' = \{\alpha \circ x + \beta : x \in F\}$ and $d' = \alpha \circ d + \beta$; and
- *independent of irrelevant alternatives* if $f(F, d) = f(F', d)$ for any two bargaining problems (F, d) and (F', d) such that $F' \subseteq F$ with $d \in F'$ and $f(F, d) \in F'$.

Here, $\circ$ denotes component-wise multiplication of vectors, i.e., $(s \circ t)^T = (s_1 t_1, s_2 t_2)$ for all $s, t \in \mathbb{R}^2$.

In the context of bargaining, Pareto efficiency means that no payoff is wasted, and symmetry is an obvious fairness property. Invariance under positive affine transformations should hold because payoffs are just a representation of the underlying ordinal preferences. The intuition behind independence of irrelevant alternatives is that an outcome only becomes easier to justify as a solution when other outcomes are removed from the set of feasible outcomes.

Theorem 20.1. *Nash's bargaining solution is the unique bargaining solution that is Pareto efficient, symmetric, invariant under positive affine transformations, and independent of irrelevant alternatives.*

Proof. We denote the Nash bargaining solution by f^N and begin by showing that it satisfies the axioms. For Pareto efficiency, this follows directly from the fact that the objective function is increasing in v_1 and v_2. For symmetry, assume that $d_1 = d_2$ and let $v^* = (v_1^*, v_2^*) = f^N(F, d)$. Clearly (v_2^*, v_1^*) maximizes the objective function, and by uniqueness of the optimal solution $(v_2^*, v_1^*) = (v_1^*, v_2^*)$ and thus $f_1^N(F, d) = f_2^N(F, d)$. For invariance under positive affine transformations, define F' and d' as above, and observe that $f^N(F', d')$ is an optimal solution of the problem to maximize $(v_1 - \alpha_1 d_1 - \beta_1)(v_2 - \alpha_2 d_2 - \beta_2)$ subject to $v \in F'$, $v_1 \geqslant d_1$, and $v_2 \geqslant d_2$. By setting $v' = \alpha \circ v + \beta$, it follows that $f^N(F', d') = \alpha \circ f^N(F, d) + \beta$. For independence of irrelevant alternatives, let $v^* = f^N(F, d)$ and $F' \subseteq F$. If $v^* \in F'$, it remains optimal and thus $v^* = f^N(F', d)$.

Now consider a bargaining solution f that satisfies the axioms, and fix F and d. Let $z = f^N(F, d)$, and let F' be the image of F under an affine transformation that maps z to $(1/2, 1/2)$ and d to the origin, i.e.,

$$F' = \{\alpha \circ v + \beta : v \in F, \alpha \circ z + \beta = (1/2, 1/2)^T, \alpha \circ d + \beta = 0\}.$$

Since both f and f^N are invariant under positive affine transformations, $f(F, d) = f^N(F, d)$ if and only if $f(F', 0) = f^N(F', 0)$. It thus suffices to show that $f(F', 0) = (1/2, 1/2)$.

We begin by showing that for all $v \in F'$, $v_1 + v_2 \leqslant 1$. Assume for contradiction that there exists $v \in F$ with $v_1 + v_2 > 1$, and let $t^\delta = (1-\delta)(1/2, 1/2)^T + \delta v$. By convexity of F', $t^\delta \in F'$ for $\delta \in (0, 1)$. Moreover, since the objective function has a unique maximum, we can choose δ sufficiently small such that $t_1^\delta t_2^\delta > 1/4 = f^N(F', 0)$, contradicting optimality of $f^N(F', 0)$.

Now let F'' be the closure of F' under symmetry, and observe that for all $v \in F''$, $v_1 + v_2 \leqslant 1$. Therefore, by Pareto optimality and symmetry of f, $f(F'', 0) = (1/2, 1/2)^T$. Since f is independent of irrelevant alternatives, $f(F', 0) = (1/2, 1/2)^T$ as required. $\square$