Game Theory Fall 2003

Debraj Ray

The One-Shot Deviation Principle

The one-shot deviation principle is fundamental to the theory of extensive games. It was originally formulated by David Blackwell (1965) in the context of dynamic programming. As the strategy of other players induces a normal maximization problem for any one player, we can formulate the principle in the context of a single-person decision tree.

Consider a possibly infinite tree. A *path* y is an ordered collection of nodes in the tree, with adjacent entries connected by immediate succession, and having the property that if y has a last entry, it must be a terminal node of the tree. *Note*: paths don't necessarily start at the initial node of the tree.

Every path y and node x in y induces a "subpath" y_x with initial node x in the obvious way. Two paths y and y' *diverge at* x if they share the same nodes up to x but have distinct subpaths thereafter.

To each path y attach a return $\pi(y)$. We make the following assumptions on π:

[A.1] (consistency) If $\pi(y) \geq \pi(y')$, and if y and y' diverge at x, then $\pi(y_x) \geq \pi(y'_x)$.

[A.2] (continuity) Fix any path y. For every $\epsilon > 0$, there exists an integer N such that if $n \geq N$ and if another path y' shares the first n nodes as y, then $|\pi(y) - \pi(y')| < \epsilon$.

Remarks:

(a) Finite decision trees with payoffs at terminal nodes automatically satisfy [A.1] and [A.2].

(b) Infinite optimization problems with discounting and additively separable payoffs also satisfy [A.1] and [A.2]. This is the case studied in Blackwell (1965).

(c) If nodes are interpreted as information sets and payoffs as expected payoffs then stochastic decision problems (or responses to opponent behavior strategies) can easily be included in this framework.

A *strategy* σ assigns to each non-terminal node x a probability distribution over $A(x)$, the set of immediate successors of x. Starting from any node x, a strategy induces probability distributions over paths with initial node x in the obvious way. Define $\pi(\sigma, x)$ to be the expectation of $\pi(y)$ over all such paths.

A strategy σ is *optimal* if there is no strategy σ' and node x such that $\pi(\sigma', x) > \pi(\sigma, x)$.

For any strategy σ, node x, and any action (node) $a \in A(x)$, define σ_a to be the strategy obtained by simply substituting the deterministic choice a at x, instead of what was prescribed by σ, and leaving all else unchanged.[1]

A strategy σ is *unimprovable* if there is no node x, $a \in A(x)$ and corresponding σ_a such that $\pi(\sigma_a, x) > \pi(\sigma, x)$.

Observe that σ_a is a special strategy, differing as it does from σ by only "a one-shot deviation" at the node x. It is therefore obvious that an optimal strategy is unimprovable. The converse is what we're after:

[1] I don't use x in the notation for the alternative strategy because each action (node) has a distinct name and a unique immediate predecessor, so x is identifiable from this information.

Theorem **1.** *Under [A.1] and [A.2], an unimprovable strategy must be optimal.*

Proof. Suppose, on the contrary, that σ is unimprovable, and yet it is not optimal. Then there exists σ' and node x_0 such that $\pi(\sigma', x_0) > \pi(\sigma, x_0)$. Because stochastic strategies add nothing to best payoff, this is equivalent to the following assertion: there is a path y starting from x_0 such that

$$\pi(y) \geq \pi(\sigma, x_0) + 2\epsilon$$

for some $\epsilon > 0$. Now using [A.2], choose an integer N such that if any path y' starting from x_0 shares the first $N+1$ nodes as y,

$$\pi(y') \geq \pi(y) - \epsilon.$$

For all such paths y', it follows from the two inequalities above that

$$\pi(y') \geq \pi(\sigma, x_0) + \epsilon.$$

Call the first $N+1$ nodes of y $x_0, \ldots, x_N$. In particular, this means that a finite number of one-shot deviations at the nodes x_i, with σ applied everywhere else, is enough to generate a payoff improvement at x_0.

Define a family of N different strategies α_i, for $i = 0, \ldots, N-1$, by the property that α_i chooses x_{j+1} at the node x_j, for every j between 0 and i, and concides with σ elsewhere. Then the conclusion of the previous paragraph informs us that

$$\pi(\alpha_{N-1}, x_0) > \pi(\sigma, x_0). \tag{1}$$

Notice that α_{N-2} fully coincides with σ from the node x_{N-1} "downwards", while α_{N-1} is a one-shot deviation from σ at that node. Because σ is unimprovable by assumption, we have

$$\pi(\alpha_{N-2}, x_{N-1}) = \pi(\sigma, x_{N-1}) \geq \pi(\alpha_{N-1}, x_{N-1}),$$

and applying [A.1], we conclude that — since α_{N-2} and α_{N-1} will share the same nodes $x_0, \ldots, x_{N-1}$ along every path generated by the two —

$$\pi(\alpha_{N-2}, x_0) \geq \pi(\alpha_{N-1}, x_0). \tag{2}$$

Combining (1) and (2), we may conclude that

$$\pi(\alpha_{N-2}, x_0) > \pi(\sigma, x_0). \tag{3}$$

Proceeding step by step in this way (and using unimprovability and [A.1] each time), we can finally see that

$$\pi(\alpha_0, x_0) > \pi(\sigma, x_0). \tag{4}$$

But α_0 is just a one-shot deviation from σ. Formally, $\alpha_0 = \sigma_{x_1}$. Therefore (4) contradicts the unimprovability of σ. □

References

Blackwell, D. (1965), "Discounted Dynamic Programming," *Annals of Mathematical Statistics* **36**, 226–235.