Stochastic Decision Processes:

Theory, Computation, and Empirical Applications

John Rust
Yale University

November/December 1999

Lecture 1: Theory
Lecture 2: Computation
Lecture 3: Empirical Applications: Discrete Decision Processes
Lecture 4: Empirical Applications: Continuous Decision Processes

Optimal Strategies (Decision Rules) for Static Decision Problems

- An optimal strategy is the equilibrium solution to a *game against nature.*
- Consider a decision maker or *agent* who has a utility or reward function $u(y,a)$ that depends on an action $a \in A \subset R^m$ and a vector $y \in R^k$ of "payoff relevant" state variables. Nature is modeled as adopting an exogenously specified mixed strategy that generates a random state $\tilde{y} \sim P(y)$.
- Agents are assumed to be expected utility maximizers. The optimal strategy, or *decision rule* $\alpha(x)$ is the agent's *best response* to nature's draw of $\tilde{y} \sim P(y)$. It will generally be a function of all *information* x that is available to the agent at the time the decision is taken. We assume the decision maker knows $u(y,a)$ and $P(y)$ and faces no informational or computational costs in determining an optimal action.
- **Case 1: no information** *Suppose* $x = \emptyset$, *then* $\alpha(x) = a^*$ *where the optimal action* $a^* \in A$ *solves:*

$$a^* = \underset{a \in A}{argmax}\, E\{u(\tilde{y},a)\} = \int u(y,a)P(dy).$$

- **Case 2: full information** *Suppose* $x = y$, *then* $\alpha(y)$ *is the solution to:*

$$\alpha(y) = \underset{a \in A}{argmax}\, u(y,a).$$

Lemma 1: *Full information is better than no information, i.e.* $\max E \leq E \max$ *where*

$$E\{u(\tilde{y}, a^*)\} = \max_{a \in A} E\{u(\tilde{y}, a)\} \leq E\{\max_{a \in A} u(\tilde{y}, a)\} = E\{u(\tilde{y}, \alpha^*(\tilde{y}))\}.$$

Proof:

$$\max_{a \in A} E\{u(\tilde{y}, a)\} = \max_{a \in A} \int u(y, a) P(dy) \leq \int \max_{a \in A} u(y, a) P(dy) = E\{\max_{a \in A} u(\tilde{y}, a)\}.$$

Let V_{ni} and V_{fi} be the indirect utility functions or *value functions* given by

$$\begin{aligned} V_{ni} &= \max_{a \in A} E\{u(\tilde{y}, a)\} \\ V_{fi} &= E\left\{\max_{a \in A} u(y, a)\right\}. \end{aligned}$$

Then an equivalent way of stating Lemma 1 is:

$$V_{ni} \leq V_{fi}.$$

- **Case 3: Partial information** Suppose the agent observes a noisy signal $x \in R^n$ where $x \sim F(x)$ and the conditional probability of y given x is given by $P(y|x)$. Then the optimal decision rule $\alpha^*(x)$ and value function $V(x)$ are given by

$$\begin{aligned}
\alpha^*(x) &= \underset{a\in A}{argmax}\, E\{u(\tilde{y},a)|x\} = \underset{a\in A}{argmax} \int u(y,a)P(dy|x) \\
V(x) &= \max_{a\in A} E\{u(\tilde{y},a)|x\} = \max_{a\in A} \int u(y,a)P(dy|x).
\end{aligned}$$

Define the partial information value V_{pi} by

$$V_{pi} = E\{V(\tilde{x})\} = \int V(x)F(dx) = \int \max_{a\in A} u(y,a)P(dy|x)F(dx).$$

Lemma 2: *More information is better than less information, i.e. if $x = (x_1, x_2)$ then with probability 1 we have:*

$$E\{V(x_1,x_2)|x_j\} \geq V(x_j), \quad j \in \{1,2\}.$$

Proof: Using the Law of Iterated Expectations, we have:

$$\begin{aligned}
V(x_j) &= \max_{a\in A} E\{u(\tilde{y},a)|x_j\} \\
&= \max_{a\in A} E\left\{E\{u(\tilde{y},a)|x_1,x_2\} \mid x_j\right\} \\
&\leq E\left\{\max_{a\in A} E\{u(\tilde{y},a)|x_1,x_2\} \mid x_j\right\} \\
&= E\{V(x_1,x_2)|x_j\}.
\end{aligned}$$

Corollary: *We have:*

$$V_{ni} \leq V_{pi} \leq V_{fi}$$

Note: After taking expectations, problem becomes deterministic

$$V(x) = \max_{a \in A} Eu(x,a) \text{ where } Eu(x,a) \equiv \int u(y,a)P(dy|x).$$

Example 1: The discrete choice decision rule

- Let $x = (z, \epsilon)$ where $z \in R^k$ and $\epsilon \equiv (\epsilon_1, \ldots, \epsilon_{|A|}) \in R^{|A|}$ where $|A| = \{\text{number of elements in set } A\} < \infty$.
- Optimal decision rule is $\alpha^*(x, \epsilon)$ given by

$$\alpha^*(z,\epsilon) = \max_{a \in A} \left[Eu(z,a) + \sigma \epsilon_a \right],$$

where

$$Eu(z,a) = \int u(y,a)P(dy|z).$$

- Thus, the discrete choice model handles choice under uncertainty, and conditional on z, the random variable $\alpha^*(z,\epsilon)$ has a multinomial distribution:

$$Pr\left\{\alpha^*(z,\epsilon) = a|z\right\} = \int I\left\{\alpha^*(z,\epsilon) = a\right\} q(d\epsilon|z) = \frac{\exp\left(Eu(z,a)/\sigma\right)}{\sum_{j=1}^{|A|} \exp\left(Eu(z,j)/\sigma\right)},$$

where σ is the scale parameter for ϵ which has a standardized *IID* extreme value distribution.

Example 2: Let $y, a \in R^1$ and $u(y,a) = -(y-a)^2$. Assume $y \sim N(\mu, \sigma^2)$ and that our information is the noisy signal $x = y + \epsilon$ where $\epsilon \sim N(0, \gamma^2)$ and is independent of y. The no-information case corresponds to $\gamma^2 = \infty$, the full-information case corresponds to $\gamma^2 = 0$, and the partial information case corresponds to $0 < \gamma^2 < \infty$. In the latter case:

$$\alpha^*(x) = (1-b)\mu + bx$$

where

$$b = \frac{\sigma^2}{\sigma^2 + \gamma^2}.$$

The weight on x is decreasing in γ^2. In the limit we have

$$\begin{aligned} \lim_{\gamma^2 \to 0} \alpha^*(x) &= x \\ \lim_{\gamma^2 \to \infty} \alpha^*(x) &= \mu. \end{aligned}$$

It is not difficult to show that

$$\begin{aligned} V_{ni} &= -\sigma^2 \\ V_{pi} &= -\left[(1-b)^2\sigma^2 + b^2\gamma^2\right] \\ V_{fi} &= 0 \end{aligned}$$

and it is not hard to verify that $V_{ni} \leq V_{pi} \leq V_{fi}$ in this case. Note that

$$V(x) = -\frac{\sigma^2}{\sigma^2/\gamma^2 + 1} > -\sigma^2 = V_{ni}.$$

- Note that these results are crucially dependent on the assumption that the information x is exogenously given (i.e. draws from a distribution $F(x)$ that does not depend on a), and that there are no costs of computation. If there are costs to gathering and processing information, then the result that "more information is always better" no longer holds.

- Examples of problems where gathering information is costly and where it is generally not optimal to attempt to gather all possible information about an uncertain outcome include the *optimal search* problem and the *multi-armed bandit problem.*

- We will also consider the theory of *Information-based complexity* where it is costly to both gather and process information. This theory is useful in defining optimal algorithms for solving stochastic decision processes. However this theory has been applied to assist *economists* in computing solutions to stochastic decision processes, but not to the *agents* who we posit are also solving these problems. Later we will briefly touch on theories of *bounded rationality,* but for most of these lectures *we will ignore the costs of computation and assume that agents can implement any feasible decision rule.*

Optimal Decision Rules for non-expected utility maximizers

- I use the Russian notation $Mu(x,a)$ to denote the mean of the random variable $u(\tilde{y},a)$ given (x,a). However $Mu(x,a)$ could be some other *certainty equivalent* of the distribution of $u(\tilde{y},a)$ — not just its mean. For example we could hypothesize that people are *median utility maximizers*

$$Mu(x,a) = \underset{m(\cdot)}{\mathit{argmin}} \int |u(y,a) - m| \, P(dy|x) = \mathrm{med}(u(\tilde{y},a)|x).$$

- Note that we can also write $Mu(x,a) = \mu(u,P)$ to emphasize the dependence of the certainty equivalent on both the utility fuction u and the probability P. Expected utility is a special case where $\mu(u,P)$ takes the form of a conditional expectation operator $Mu(x,a) = Eu(x,a)$ and thus is *linear* in both u and P. Specifically for any $a,b \in R$ and functions u and v we have

$$\mu(au + bv, P) = a\mu(u,P) + b\mu(v,P).$$

For $\lambda \in (0,1)$ and any conditional distributions $P(y|x)$ and $Q(y|x)$ we have

$$\mu(u, \lambda P + (1-\lambda)Q) = \lambda\mu(u,P) + (1-\lambda)\mu(u,Q).$$

- Unfortunately, anomalous findings in numerous human experiments, including direct violations such as the *Allais Paradox* suggest that human decision-makers do not obey one or more of the underlying *axioms,* the necessary and sufficient conditions for the expected utility representation. The ones that have received the most scrutiny are the *sure thing principal* and the *independence axiom* (see Machina, 1982).

$$p \succcurlyeq q \quad \textit{iff} \quad \forall r \in \mathcal{P}, \forall \lambda \in (0,1) \quad \lambda p + (1-\lambda) r \succcurlyeq \lambda q + (1-\lambda) r.$$

- However non-expected utility can be tricky. Need minimal regularity such as, $P(y|x) = \delta_x \Longrightarrow \mu(u, \delta_x) = u(x, a)$ and other weak continuity conditions.

- Some specifications/axiomatizations that generate certainty equivalents $\mu(u, P)$ which are not linear in u ane P are thought to be capable of resolving some of the experimental anomalies.

- **Exersize:** Is $\mu(u, P) = \mathrm{med}(u(\tilde{y}, a)|x)$ linear in u and P?

- **Example** For $\gamma > 0$ let $\mu(u, P)$ be given by:

$$\mu(u, P) = \left[\int u(y, a)^{\gamma} P(dy|x) \right]^{\frac{1}{\gamma}}.$$

  Is this $\mu(u, P)$ linear in u and P?

Modeling the distinction between *risk* **and** *uncertainty.*

- **Definition:** *A decision problem is* **uncertain** *if the agent does not know the probability distribution $P(y|x)$ generating the payoff relevant outcomes $\tilde{y}$.*
- We model decision maker as believing $P(y|x)$ belongs to some class of probability models $\mathcal{P}$. Agents make decisions using their best guess of the true $P(y|x)$.
- **Parametric Bayesian Decision Theory** $\mathcal{P}$ can be described by a parametric mapping $\Theta \to \mathcal{P}$ with typical element $P(y|x,\theta)$. Also need beliefs about how the signal x is generated, represented by $f(x|\theta)$. The **likelihood** for (y,x) is

$$L(y,x,\theta) = P(y|x,\theta)f(x|\theta).$$

$L(y,x,\theta^*)$ for some unknown $\theta^* \in \Theta$ is **true data generating process.** Agent has *prior belief* $\pi(\theta)$ which is a probability measure on Θ. Optimal (*Bayes Rule* is then given by:

$$V(x) = \max_{a\in A} Eu(x,a) = \max_{a\in A} \int_\Theta \int_Y u(y,a)P(dy|x,\theta)\rho(d\theta|x),$$

where $\rho(\theta|x)$ is the **posterior distribution** given by:

$$\rho(\theta|x) = \frac{\int_Y P(dy|x,\theta)f(x|\theta)\pi(\theta)}{\int_\Theta P(y|x,\theta)f(x|\theta)\pi(d\theta)}.$$

- **Question:** *Where do priors and beliefs come from? What if the agent does not know these objects either?*

- **Example: Robust Decision Making** Anderson, Hansen and Sargent (1998) consider a decison maker who is concerned about whether the "reference model" $P(y|x)$ is correct. Consider the expected utility of the worst case deviation model given by:

$$\sup_{w>0} J(w) \Longrightarrow \quad J(w) = -\frac{1}{\theta} I(w,P) + \int u(y,a)P(dy|x)$$

where $\theta > 0$ and $w > 0$ represents a direction of departure from $P(y|x)$, and $I(w,P)$ is the Kullback-Liebler distance between $P(y|x)$ and $w(y|x)$. Then the robust decision maker prefers action a to a' if

$$\mathcal{R}(x,a) = \frac{1}{\theta}\log\left\{\int \exp(\theta u(y,a))P(dy|x)\right\} > \mathcal{R}(x,a').$$

Maximin Expected Utility (Gilboa and Schmeidler, 1989, Chamberlain 1999).

- These papers explicitly distinguish between choices under uncertainty (unknown outcome probabilities) which they call *horse lotteries*, and choices under risk (known outcome probabilities) which they call *roulette lotteries*.
- Let $\theta \in \Theta$ index the parameters of the distribution $P(y|x,\theta)$ the decision maker does not know. But now, instead of a unique prior $\pi(\theta)$ over θ the agent has a set Π of possible priors $\pi(\theta)$ over Θ.
- **Generalized Independence Axiom** Given a preference relation $\succcurlyeq$ over conditional probability distributions of payoffs, $P(y|\theta,x)$ on a set Y, where x is a signal observed by the agent at the time they make their decision and θ is an unknown parameter vector. $P(\cdot|\theta,x) \succcurlyeq Q(\cdot|\theta,x)$ *iff* $\forall\lambda \in (0,1)$ and all roulette lotteries $R(\cdot|\theta,x)$ we have
$$\lambda P + (1-\lambda)R \quad \succcurlyeq \quad \lambda Q + (1-\lambda)R.$$
- **Theorem** *The generalized independence axiom implies that there if a utility function $u(y)$ and a set of probability measures on $\Theta \times X$ such that $P(\cdot|\theta,x) \succcurlyeq Q(\cdot|\theta,x)$ iff*
$$\min_{\pi\in\Pi} \int_\Theta \int_X \int_Y u(y)P(dy|x,\theta)f(dx|\theta)\pi(d\theta) \geq \min_{\pi\in\Pi} \int_\Theta \int_X \int_Y u(y)Q(dy|\theta,x)f(dx|\theta)\pi(d\theta).$$

- Suppose the set Π is the convex hull of a finite number of "prior distributions" on Θ, $\{\pi_1, \ldots, \pi_J\}$. The *Minimax Theorem* yields:

$$\overline{V} = \sup_{a \in \mathcal{A}} \inf_{\pi \in \Pi} Eu(a, \pi) = \inf_{\pi \in \Pi} \sup_{a \in \mathcal{A}} Eu(a, \pi) = \underline{V}$$

where

$$Eu(a, \pi) = \int_{\Theta} \int_{X} \int_{Y} u(y, a) P(dy|\theta, x) f(dx|\theta) \pi(d\theta).$$

A minimax expected utility maximizer behaves *cautiously* evaluating any decision rule at the *least favorable prior* in Π. The least favorable prior π^* can be found from the solution to a concave programming problem

$$\pi^* = \underset{\pi \in \Pi}{argmin} \max_{a \in \mathcal{A}} Eu(a, \pi).$$

- **Application:** Assume decision maker wants to estimate the mean of an AR-(1) model of dynamic earnings data using a signal x, where $x \equiv \{x_{it}\}$ is given by:

$$x_{it} = \gamma x_{it-1} + \zeta + \eta_{it}$$

where $\{\eta_{it}\} \sim \mathit{IID}\, N(0, \sigma^2)$, $\zeta \sim N(0, \kappa^2)$, independent of $\{\eta_{it}\}$. Let $\lambda = \kappa/\sigma$, and $\theta = (\gamma, \lambda)$. $Y = \Theta$ and set $u(y, a) = u(\theta, a) = -(\gamma - a)^2$.

- For any prior π optimal decision rule is Bayes Rule $\alpha^*(x, \pi)$ given above. If we can find least favorable prior π^* then optimal decision rule for a minimax expected utility maximizer is $\alpha^*(x, \pi^*)$.

- Note that since Π is convex hull of $\{\pi_1, \dots, \pi_J\}$ we have $\pi \in \Pi$ *iff* $\pi = \sum_{j=1}^{J} \delta_j \pi_j$ where $\delta_j \in [0, 1]$ and $1 = \sum_{j=1}^{J}$. Call this π^δ. We have:

$$Eu(a, \pi^\delta) = \sum_{j=1}^{J} \delta_j Eu(a, \pi_j).$$

Let $g(\delta) = \max_{a \in \mathcal{A}} Eu(a, \pi^\delta)$. Since $Eu(a, \pi^\delta)$ is a linear function of $\delta = (\delta_1, \dots, \delta_J)$ for each a, $g(\delta)$, as a maximum of linear functions must be convex in δ.

- The least favorable prior can be computed from the following convex minimization problem, $\pi^* = \pi^{\delta^*}$ where

$$\delta^* = \underset{\delta \in \Delta}{\mathit{argmin}}\, g(\delta)$$

where Δ is the $J - 1$-dimensional simplex.

- Numerical example where $J = 120$, $N = 100$, $T = 2$, put prior over $\theta = (\gamma, \lambda)$ for 15 values of *gamma* from $(0, \dots, 1.4)$ and 8 values of λ from $10^{-4}, \dots, 1.4$. Optimal solution gives $\sqrt{\overline{V}} = .115$ and $\delta_j^* = 0$ for 69 points and $\delta_j^* > 0$ for 51 points.

Now add dynamics to the game against nature:

Definition: A *stochastic decision process* consists of the following objects:

- A *time index* $t \in \{0, 1, 2, \ldots, T\}$, $T \leq \infty$
- A *state space* S
- An *action space* A
- A set of *histories* $\{H_{t-1}\}_{t=1}^{T}$ of realized states and actions up to time $t-1$, where $H_{t-1} = (s_0, d_0, s_1, a_1, \ldots, s_{t-1}, a_{t-1})$.
- A family of *constraint sets* $\{A_t(s_t, H_{t-1}) \subseteq A\}$
- A family of *transition probabilities* $\{p_t(ds_t|H_{t-1})\}$
- A *utility functional* $U(\mathbf{s}, \mathbf{a})$ *where* $\mathbf{s} = (s_0, s_1, s_2, \ldots) \in S^T$ and $\mathbf{a} = (a_0, a_1, a_2, \ldots) \in A^T$.