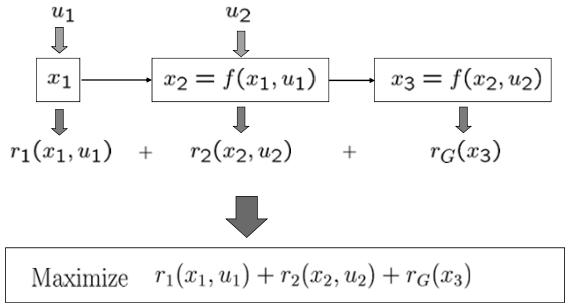
用語(1)

- state (space)
  - 状態（空間）
- decision (space)
  - 決定（空間）
  - action も同じ意味で使用
- reward
  - 利得
- terminal reward
  - 終端利得
- additive
  - 加法型
- transition (law)
  - 推移（法則）
  - transition system と transition law は同じ意味で使用
- deterministic
  - 確定的
- stochastic
  - 確率的
- stage
  - 期
  - time:時刻も同じ意味で使用

用語(2)

- decision function
  - 決定関数
- policy
  - 政策
  - Markov policy:マルコフ政策
  - general policy:一般政策
- subproblem
  - 部分問題
- recursive
  - 再帰的
  - recursive equation:再帰式 = recursive formula
- optimal value function
  - 最適値関数
  - value function:値関数も同じ意味で使用
- Markov
  - 現在にのみ依存する（過去に依存しない）性質を持つものに用いられる
  - Markov transition は stochastic transition の意
- decision process
  - 決定過程

確定システム



再帰式の導出

$$\text{Maxmize} \quad r_1(x_1, u_1) + r_2(x_2, u_2) + r_G(x_3)$$
$$\text{subject to} \quad \text{(i)} \ x_{n+1} = f(x_n, u_n)\,, \quad n = 1, 2$$
$$\text{(ii)} \ u_1, u_2 \in U$$

部分問題群

$$v^3(x_3) = r_G(x_3)\,, \quad x_3 \in X$$
$$v^2(x_2) = \underset{u_2 \in U}{\text{Max}}[r_2(x_2, u_2) + r_G(x_3)]\,, \quad x_2 \in X$$
$$v^1(x_1) = \underset{u_1, u_2 \in U}{\text{Max}}[r_1(x_1, u_1) + r_2(x_2, u_2) + r_G(x_3)]\,, \quad x_1 \in X$$

（ただし、$x_{n+1} = f(x_n, u_n),\ n = 1, 2$）

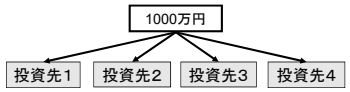
$$v^3(x_3) = r_G(x_3)\,, \quad x_3 \in X$$
$$v^2(x_2) = \underset{u_2 \in U}{\text{Max}}[r_2(x_2, u_2) + r_G(x_3)]\,, \quad x_2 \in X$$
$$v^1(x_1) = \underset{u_1 \in U}{\text{Max}}\left[r_1(x_1, u_1) + \underset{u_2 \in U}{\text{Max}}[r_2(x_2, u_2) + r_G(x_3)]\right]\,, \quad x_1 \in X$$

再帰式

$$v^3(x) = r_G(x)\,, \quad x \in X$$
$$v^2(x) = \underset{u \in U}{\text{Max}}[r_2(x, u) + v^3(f(x, u))]\,, \quad x \in X$$
$$v^1(x) = \underset{u \in U}{\text{Max}}[r_1(x, u) + v^2(f(x, u))]\,, \quad x \in X$$

例題1.1（投資問題）

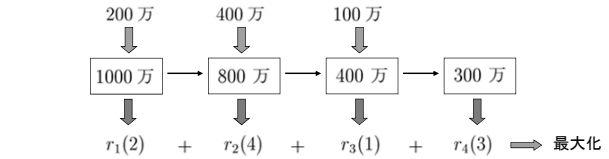
1000万円の投資資金を4つの投資先に分配し、投資した結果得られる利益を最大にしたい。1単位を100万円とし、10単位をどのように配分すれば良いだろうか。

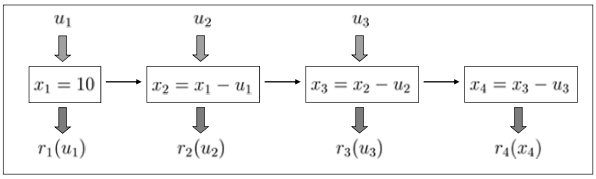


投資量 u ＼各投資先の回収利益	$r_1(u)$	$r_2(u)$	$r_3(u)$	$r_4(u)$
0	0	0	0	0
1	0.28	0.25	0.15	0.20
2	0.45	0.41	0.25	0.33
3	0.65	0.55	0.40	0.42
4	0.78	0.65	0.50	0.48
5	0.90	0.75	0.62	0.53
6	1.02	0.80	0.73	0.56
7	1.13	0.85	0.82	0.58
8	1.23	0.88	0.90	0.60
9	1.32	0.90	0.96	0.60
10	1.38	0.90	1.00	0.60

定式化

決定：各投資先への投資額　　**状態**：投資可能額　　**利得**：利益





期数	N	:	$N = 3$
状態空間	X	:	$X = \{0, 1, 2, \ldots, 10\}$
決定空間	U	:	$U = \{0, 1, 2, \ldots, 10\}$
決定制約	$U_n = U$	:	$U(x) = \{0, 1, 2, \ldots, x\}$
推移法則	$f_n = f$	:	$f(x, u) = x - u$
利得	r_n	:	$r_n(x, u) = r_n(u)$ ， $n = 1, 2, 3$
終端利得	r_G	:	$r_G(x) = r_4(x)$
初期状態	x_1	:	$x_1 = 10$

多段決定過程問題としての定式化

$$\text{Max} \quad r_1(u_1) + r_2(u_2) + r_3(u_3) + r_G(x_4)$$

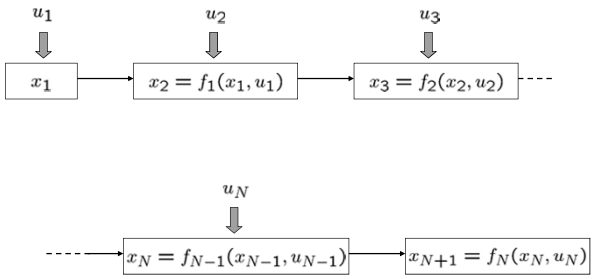
$$\text{s. t.} \quad \begin{cases} x_{n+1} = f(x_n, u_n) & n = 1, 2, 3 \\ u_n \in U(x_n) = \{0, 1, 2, \ldots, x_n\} & n = 1, 2, 3 \end{cases}$$

DPによる再帰式

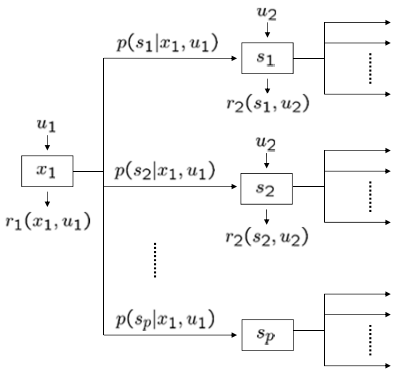
$$v^4(x) = r_4(x), \quad x \in X$$

$$v^n(x) = \max_{u \in U(x)} \{r_n(u) + v^{n+1}(x - u)\} \quad n = 1, 2, 3$$

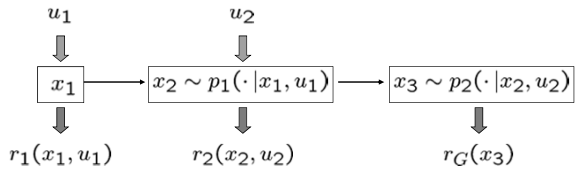
推移図（確定システム、一般）



推移図（確率システム、一般）



確率的推移システム上での加法型評価

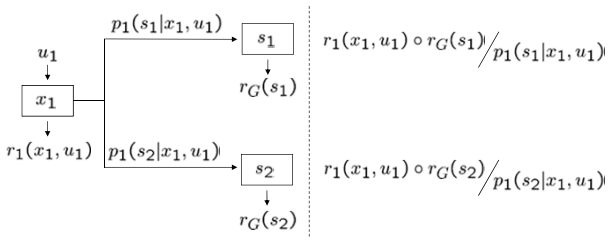


確率変数！

加法型評価

$$r_1(x_1, u_1) + r_2(x_2, u_2) + r_G(x_3)$$

確率システムの評価



$$E[r_1(x_1, u_1) \circ r_G(x_2)]$$
$$= \{r_1(x_1, u_1) \circ r_G(s_1)\} \times p_1(s_1|x_1, u_1) + \{r_1(x_1, u_1) \circ r_G(s_2)\} \times p_1(s_2|x_1, u_1)$$

例題1.2 (Deterministic Maximization)

数値例

2 期間 - 3 状態 - 2 決定 問題:

$$\begin{aligned}&\text{Maximize} \quad r_1(u_1)+r_2(u_2)+r_G(x_3)\\ &\text{subject to} \quad \text{(i)}\ x_{n+1}=f(x_n,u_n) \quad n=1,2\\ &\qquad\qquad \text{(ii)}\ u_1,u_2\in U\end{aligned}$$

ただし，データは以下のとおり：

$$r_G(s_1)=0.4,\quad r_G(s_2)=1.0,\quad r_G(s_3)=0.8$$

$$r_2(a_1)=0.8\quad r_2(a_2)=0.6$$

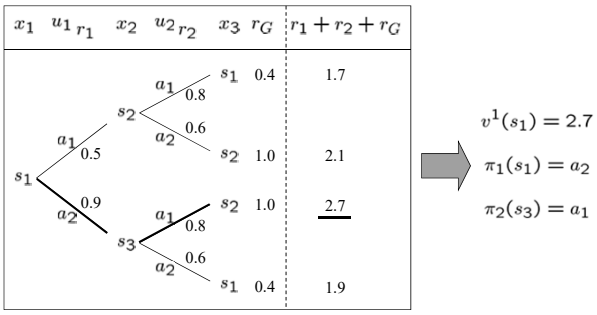
$$r_1(a_1)=0.5,\quad r_1(a_2)=0.9$$

$f(x,u)$

$x \setminus u$	a_1	a_2
s_1	s_2	s_3
s_2	s_1	s_2
s_3	s_2	s_1

$$\left[\begin{array}{c} X=\{s_1,\ s_2,\ s_3\},\ \ U_1(x)=U_2(x)=U=\{a_1,\ a_2\}\\ f_1=f_2=f,\ \ r_n(x,u)=r_n(u)\end{array}\right]$$

列挙法による解法



再帰式による解法 (1/3)

定理 4.2(4.1) より

$$v^3(x)=r_G(x),\quad x=s_1,s_2,s_3$$

$$v^2(x)=\underset{u=a_1,a_2}{\mathrm{Max}}\,[r_2(u)+v^3(f(x,u))],\quad x=s_1,s_2,s_3$$

$$v^1(x)=\underset{u=a_1,a_2}{\mathrm{Max}}\,[r_1(u)+v^2(f(x,u))],\quad x=s_1,s_2,s_3$$

を順に計算していけばよい.
まず，v^3 を求める

$$v^3(s_1)=r_G(s_1)=0.3$$
$$v^3(s_2)=r_G(s_2)=1.0$$
$$v^3(s_3)=r_G(s_3)=0.8$$

再帰式による解法 (2/3)

次に，v^2 を求める

$$\begin{aligned}v^2(s_1) &= \underset{u=a_1,a_2}{\mathrm{Max}}\,[r_2(u)+v^3(f(s_1,u))]\\ &= \mathrm{Max}[\{r_2(a_1)+v^3(f(s_1,a_1))\},\{r_2(a_2)+v^3(f(s_1,a_2))\}]\\ &= \mathrm{Max}[\{0.8+v^3(s_2)\},\ \{0.6+v^3(s_3)\}]\\ &= \mathrm{Max}[\{0.8+1.0\},\ \{0.6+0.8\}]\\ &= \mathrm{Max}[1.8,\ 1.4]=1.8,\qquad \pi_2^*(s_1)=a_1\end{aligned}$$

同様にして

$$v^2(s_2)=1.6,\ \ \pi_2^*(s_2)=a_2$$

再帰式による解法 (3/3)

最後に，v^1 を求めると

$$v^1(s_1)=\_\_\_\_\_,\ \ \pi_1^*(s_1)=\_\_\_\_\_$$
$$v^1(s_2)=\_\_\_\_\_,\ \ \pi_1^*(s_2)=\_\_\_\_\_$$
$$v^1(s_3)=\_\_\_\_\_,\ \ \pi_1^*(s_3)=\_\_\_\_\_$$

従って，最大値は初期状態 $x_1=s_1,s_2,s_3$ に対しそれぞれ

_____

となり，最適政策は次で与えられる.

$$\pi^*=\{\pi_1^*,\ \pi_2^*\}$$

終了集合

終了時刻 N が未定の問題に対しては終了集合という概念を導入する：

終了集合 $T\subset X$ が与えられたとき、システムは

$$x_n\in T$$

を満たした時点で終了するものとする。

必要に応じて以下も用いられる

$X_n(x,u)\subset X$ ： 第 $n-1$ 期の状態 x と決定 u に対し、第 n 期に生じ得る状態の集合

$U_n(x)\subset U$ ： 第 n 期において、状態 x に対し取り得る決定の集合
（より一般には $U_n(x_1,u_1,x_2,u_2,\ldots,x_n)$）