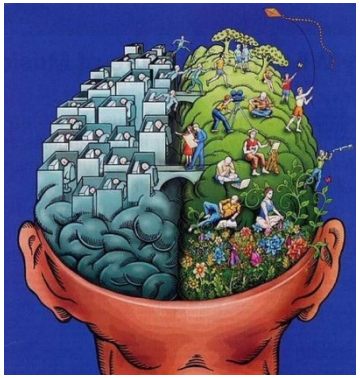
Budapesti Műszaki és Gazdaságtudományi Egyetem
Méréstechnika és Információs rendszerek Tanszék



Racionalitás, hasznosság, döntés
Markov döntési folyamat

Előadó: Hullám Gábor

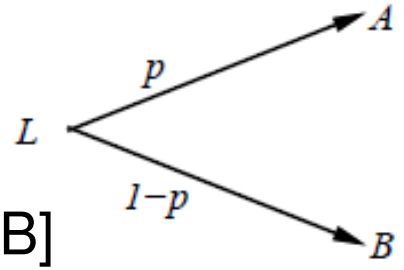
Előadás anyaga: Dobrowiecki Tadeusz

Preferenciák

Egy ágens választásai
A, B, … determinisztikus tételek,
ill. bizonytalan kimenetelű sorsjátékok

$A > B$: A preferált B-hez képest
$A \sim B$: nincs preferencia A és B között
$A \geq B$: B nem preferált A-val szemben

Sorsjáték:
$L = [p, A; (1-p), B]$



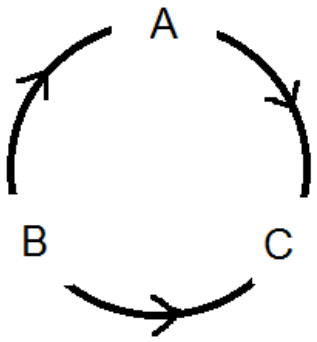
Sorrendezhetőség
$(A > B) \vee (B > A) \vee (A \sim B)$

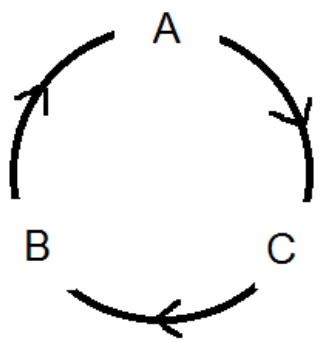
Tranzitivitás
$(B > A) \wedge (A > C) \rightarrow (B > C)$

Folytonosság
$A > B > C \rightarrow \exists p.\ [p, A; 1\text{-}p, C] \sim B$

Helyettesíthetőség
$A \sim B \rightarrow [p, A; 1\text{-}p, C] \sim [p, B; 1\text{-}p, C]$

Monotonitás
$A > B \rightarrow (p \geq q \leftrightarrow [p, A; 1\text{-}p, B] \geq [q, A; 1\text{-}q, B])$





Várható hasznosság maximalizálása

A korlátokat teljesítő preferenciákhoz létezik olyan valós értékű U(x) függvény, hogy (Ramsey, 1931, Neumann és Morgenstern, 1944):

$$U(A) \geq U(B) \leftrightarrow A \geq B$$

$$U([p_1, S_1; \ldots ; p_n, S_n] = \Sigma_i \, p_i \, U(S_i)$$

$$EU(A \mid E) = \sum_k P(Eredmény_k(A) \mid Tesz(A), E) \, U(Eredmény_k(A))$$

azt maximáló cselekvés megválasztása

Hasznosságok modellezése

Egy A állapot $\leftrightarrow$ standárt sorsolás:

- a lehető legjobb díj - u_{max} p valószínűséggel
- a lehető legnagyobb katasztrófa - u_{min} 1-p valószínűséggel
- p módosítása, amíg: $A \sim L_p$

Hasznossági skálák

- Normált: $u_{max} = 1$, $u_{min} = 0$
- Mikromort: halálesély/1000000, kb. 50 USD (2009)
  - pl. Mt.Everest: 39427 mm/ megmászás

....

A pénz hasznossága és az emberi (ir)racionalitás

Nem szabályos hasznosság! Ha L egy sorsjáték, aminek várható pénzbeli nyeresége EMV(L), akkor általában U(L) < U(EMV(L))

Hasznossági görbe: milyen p valószínűség esetén indifferens az x díj és a [p, M; 1-p, 0] sorsjáték értéke között, nagyon nagy M-re?

Tegyük fel, hogy nyert egy TV játékban. A műsorvezető most választásra kéri fel: elviheti az 1 milliós díjat, vagy felteheti azt egy pénzfeldobásos hazárdjátékon. Ha fej, nem kap semmit, ha írás, akkor kap 3 milliót. Ha hasonló a többi emberhez, akkor vonakodna játszani, és zsebre vágná a milliót. Ez irracionális volna?

1 millió < EMV(L) = 1.5 millió

De mi van, ha már van valami pénze (S_k)?

$$EU(Elfogad) = \frac{1}{2}U(S_k) + \frac{1}{2}U(S_{k+3M})$$

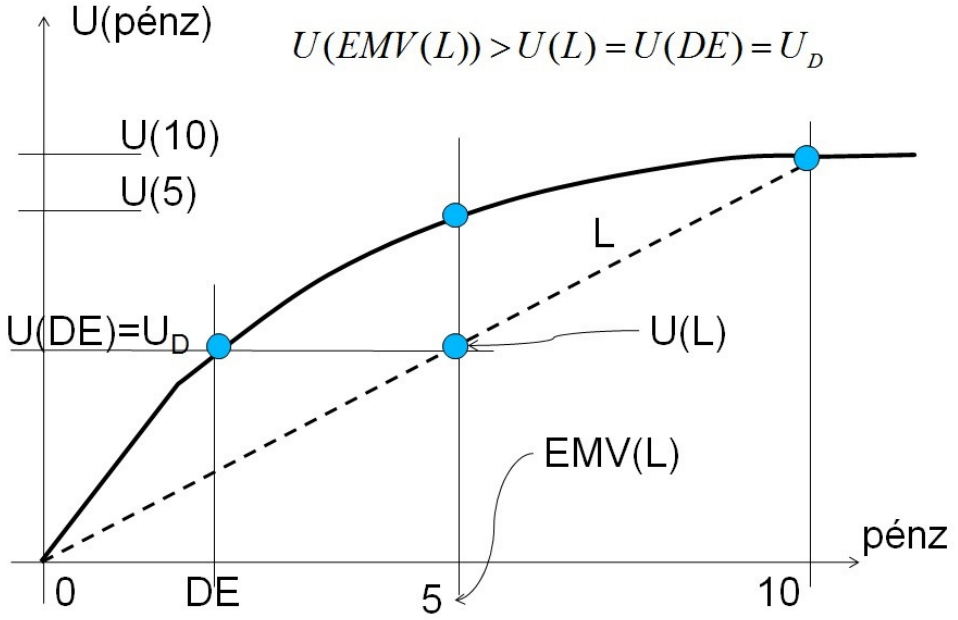
$$EU(Elutasít) = U(S_{k+1M})$$

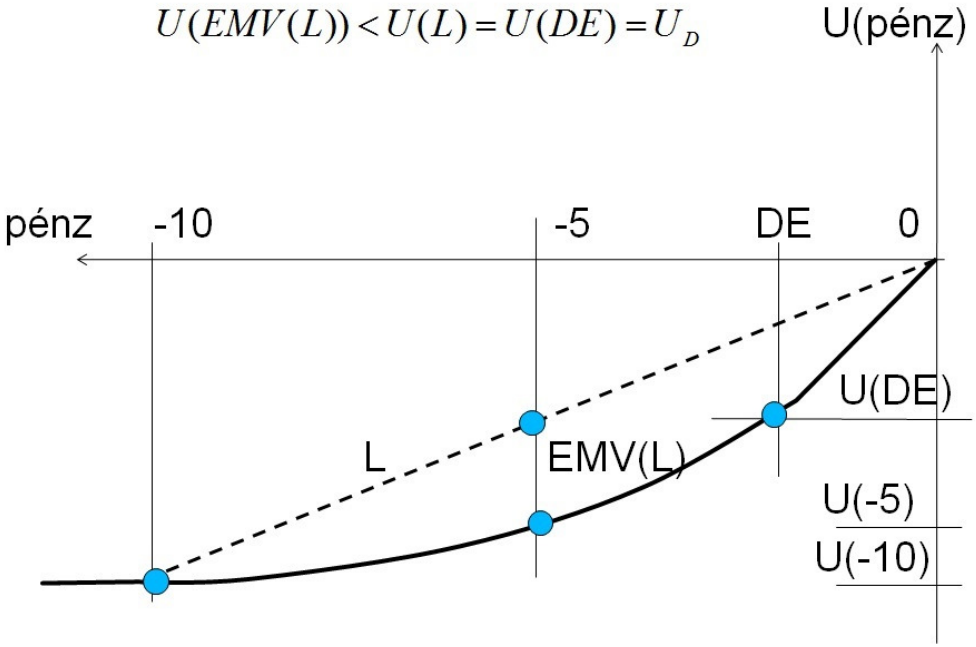
$U(S_k)$	5	5.0
$U(S_{k+1M})$	9	5.1
$U(S_{k+3M})$	11	5.3

Grayson (1960): a pénz hasznossága majdnem teljesen arányos a mennyiségének logaritmusával (először Bernoulli, 1783).

A pénz hasznossága és az emberi (ir)racionalitás

Nyereségekre: (kockázatkerülő)
U(L) < U(EMV(L) biztos kifizetése)
Veszteségekre: (kockázatkereső)
U(L) > U(EMV(L) biztos kifizetése)
Kis értékek szakasza lineáris
- kockázat-semleges
Sorsjáték determinisztikus ekvivalense
DE (játék helyett fogad el)





Többváltozós hasznosságfüggvények

U(Halálesetek, Zaj, Költség)?

$U(x_1, x_2, \ldots, x_n) = ?$ (1) teljes körű beazonosítás
(2) függetlenségek, kanonikus alakok

Additív értékfüggvény $U = k_1 U_1 + k_2 U_2 + k_3 U_3$

Pl. U(Zaj, Költség, Halálesetek) =
– Zaj[dB] × 10^4 – Költség[mFt] – Halálesetek[mikromort] × 10^{12}

Multiplikatív értékfüggvény

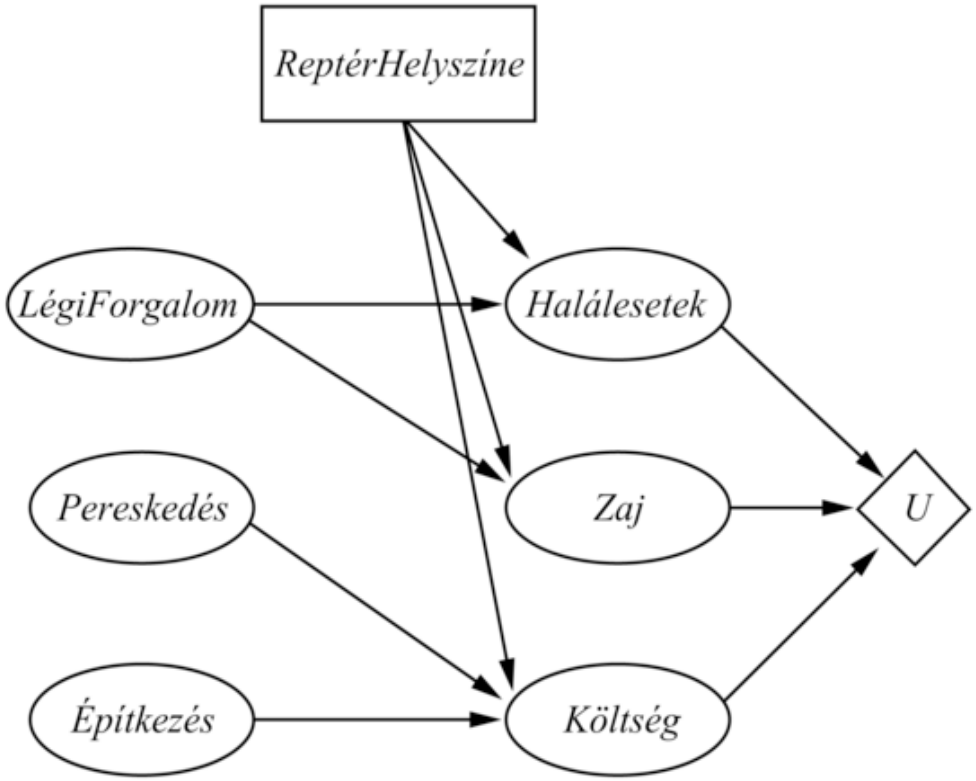
$$U = k_1 U_1 + k_2 U_2 + k_3 U_3 + k_1 k_2 U_1 U_2 + k_2 k_3 U_2 U_3 + k_3 k_1 U_3 U_1 + k_1 k_2 k_3 U_1 U_2 U_3$$

csak 3 paraméter

stb.

Döntési hálók

véletlen csomópontok
    FVT
döntési csomópontok
    döntési lehetőségek
hasznosság csomópontok
    hasznosságok leírása
    cselekvéshasznosság táblák



Következtetés:
- evidencia változók beállítása
- a döntési csomópont minden egyes értékére:
    - állítsuk be a döntési csomópontot erre az értékre
    - **számítsuk ki az a posteriori valószínűségeket a hasznosság-csomópont szüleire** (szabványos valószínűségi hálós következtetés)
    - számítsuk ki a cselekvések hasznosságát
- ? a legnagyobb hasznosságértékű cselekvés

Információ hasznossága

Legyen a meglévő evidencia E, az aktuális legjobb cselekvés α, melynek lehetséges kimenetelei Eredmény_i, az új lehetséges evidencia E_j.
A pillanatnyi legjobb cselekvés értéke:

$$EU(\alpha \mid E) = \max_A \sum_k P(Eredmény_k(A) \mid Tesz(A), E)\, U(Eredmény_k(A))$$

A pillanatnyi legjobb cselekvés értéke új evidencia után:

$$EU(\alpha_{E_j} \mid E, E_j) = \max_A \sum_i U(Eredm_i(A)) P(Eredm_i(A) \mid Tesz(A), E, E_j)$$

A teljes információ értéke (TIÉ) (az előre még nem ismert új evidencia értékeire vett átlag):

$$TIÉ_E(E_j) = \left(\sum_k P(E_j = e_{jk} \mid E) EU(\alpha_{e_{jk}} \mid E, E_j = e_{jk})\right) - EU(\alpha \mid E)$$

Racionális ágensek tranzitív preferenciáiról

Három ágens preferenciái:

(Ág1) Körte > Szőlő > Alma
(Ág2) Szőlő > Alma > Körte
(Ág3) Alma > Körte > Szőlő

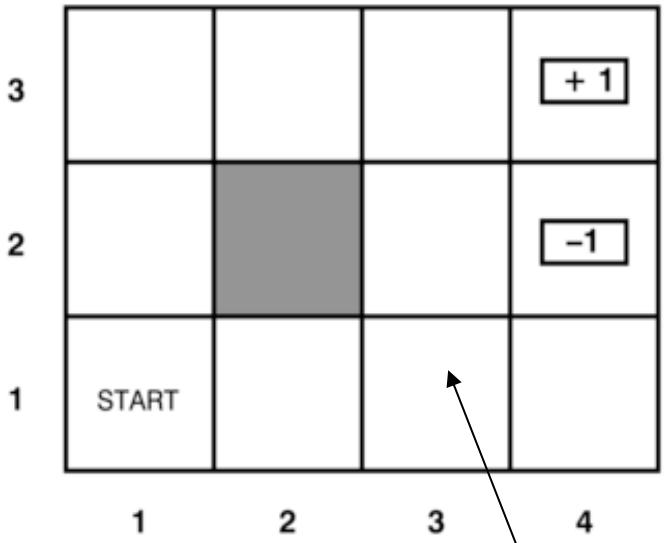
Mi a csoport véleménye, a csoport preferenciasora?

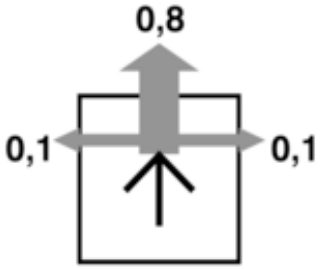
Legyen annak kifejezője a többségi választás (itt 2 az 1 ellen):

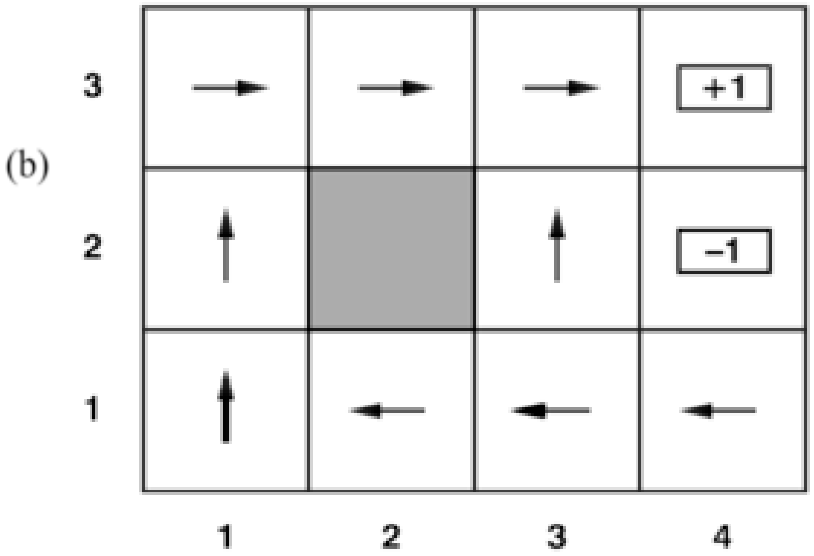
Körte > Szőlő
Szőlő > Alma
Alma > Körte

Egyenként racionális (tranzitív), együtt már nem?

Szekvenciális döntési probléma







- 0.04

Fix út: fel, fel, jobbra, jobbra, jobbra?

(optimális) Eljárásmód: $\pi(s) = a$

Szekvenciális döntési probléma

Markov döntési folyamat

Kezdőállapot:	S0
Állapotátmenet-modell:	T(s, a, s′)
Jutalomfüggvény:	R(s), vagy R(s, a, s′)

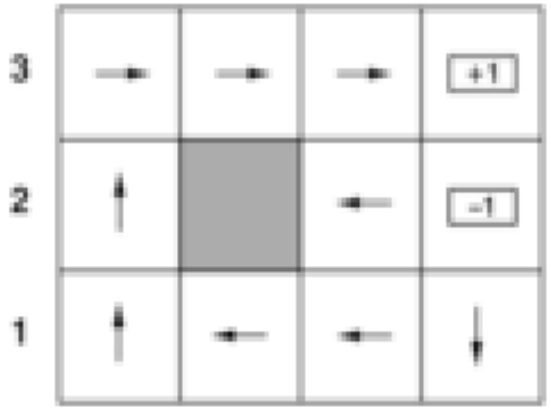
Optimális eljárásmód = optimális mozgás, döntés cselekvés megválasztására, de nem elég egyszer, folyamatosan kell, amíg nincs a probléma vége.

$\pi(s) = a$

$\pi^*(s) = a$

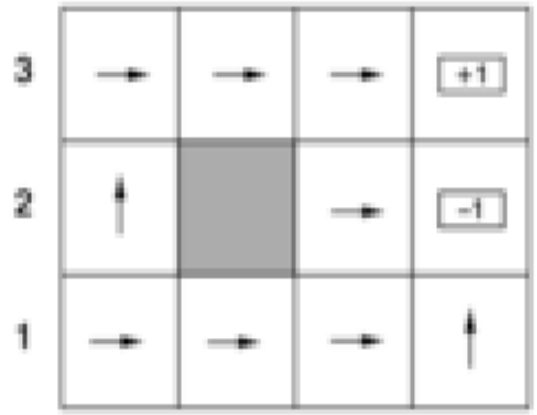
Szekvenciális döntési probléma

$-0{,}0221 < R(s) < 0$



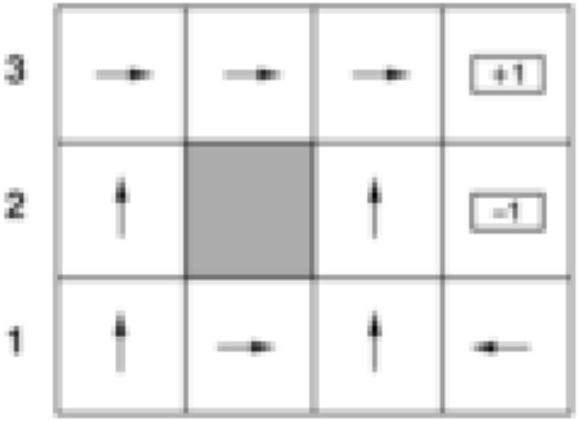
az élet csak kevéssé bánatos,
ne legyen kockázat!

$R(s) \leq -1{,}6284$



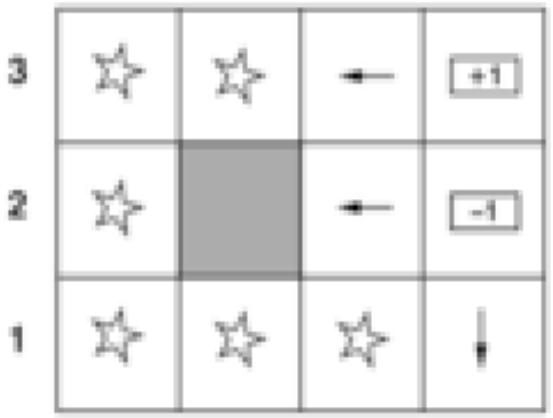
élet elviselhetetlen, ki!

$-0{,}4278 \leq R(s) \leq -0{,}0850$



élet kellemetlen; +1 állapot,
–1 kockázattal

$R(s) > 0$



élet kifejezetten élvezhető,
az ágens benn akar maradni

Szekvenciális döntési probléma

Optimalis szekvenciális döntési probléma

végtelen horizont
véges horizont
optimális eljárásmód nem-stacionárius
stacionárius

Többattribútumú hasznosságelmélet:
ágens preferenciái az állapotsorozatok között stacionáriusok

additív jutalmak $U_h([s_0, s_1, s_2, \ldots]) = R(s_0) + R(s_1) + R(s_2) + \ldots$

leszámított jutalmak

$$U_h([s_0, s_1, s_2, \ldots]) = R(s_0) + \gamma R(s_1) + \gamma^2 R(s_2) + \ldots$$

Szekvenciális döntési probléma

Leszámított jutalmak, egy végtelen sorozat hasznossága

$$U_h([s_0, s_1, s_2, \ldots]) = \Sigma_{t=0\ldots\infty} \gamma^t R(s_t) =< \Sigma_{t=0\ldots\infty} \gamma^t R_{\max} = {R_{\max}}/{(1-\gamma)}$$

Ha van végállapot, ha garantált, hogy az ágens végül bele kerül, akkor nincs szükség végtelen sorozatok összehasonlítására.
Egy eljárásmód, ami garantáltan végállapotba juttat, véges eljárásmód, $\gamma = 1$

Végtelen sorozatok összehasonlítása:
az időegységenkénti átlagjutalom

Optimális eljárásmód

$$\pi^* = \arg\max_{\pi} E[\Sigma_{t=0\ldots\infty} \gamma^t R(s_t) \mid \pi]$$

Optimális eljárásmód meghatározása - Értékiteráció

Egy **állapot hasznossága –** a belőle kiinduló állapotsorozatok várható hasznossága

Az állapotsorozatok függnek a végrehajtott eljárásmódtól, így elsőként egy adott π eljárásmódra definiáljuk a hasznosságot:

$$U^{\pi}(s) = E\left[\sum_{t=0}^{\infty} \gamma^t R(s_t) \,|\, \pi, s_0 = s\right]$$

$$\pi^*(s) = \arg\max_a \sum_{s'} T(s,a,s')\,U(s')$$

Optimális eljárásmód

Az **állapot hasznossága** - az állapotban tartózkodás közvetlen jutalmának és a következő állapot várható leszámított hasznosságának az összege, feltéve, hogy az ágens az optimális cselekvést választja (**Bellman** egyensúlyi **egyenlet**)

$$U(s) = R(s) + \gamma \max_a \sum_{s'} T(s,a,s')\,U(s')$$

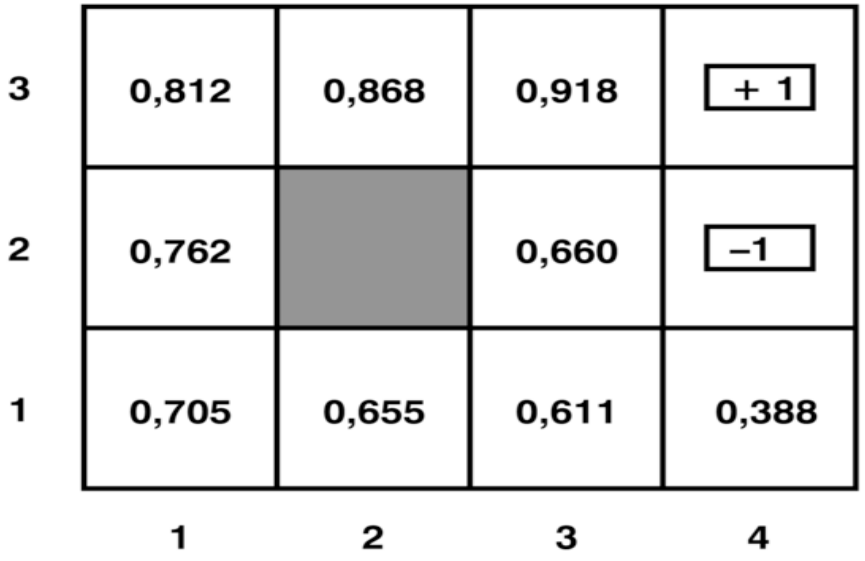
Legyen $\gamma = 1$ és a nem végállapotoknál $R(s) = -0{,}04$

Nézzük meg a 4 × 3-as világ Bellman-egyenleteinek egyikét.

Az (1, 1) állapothoz tartozó egyenlet:

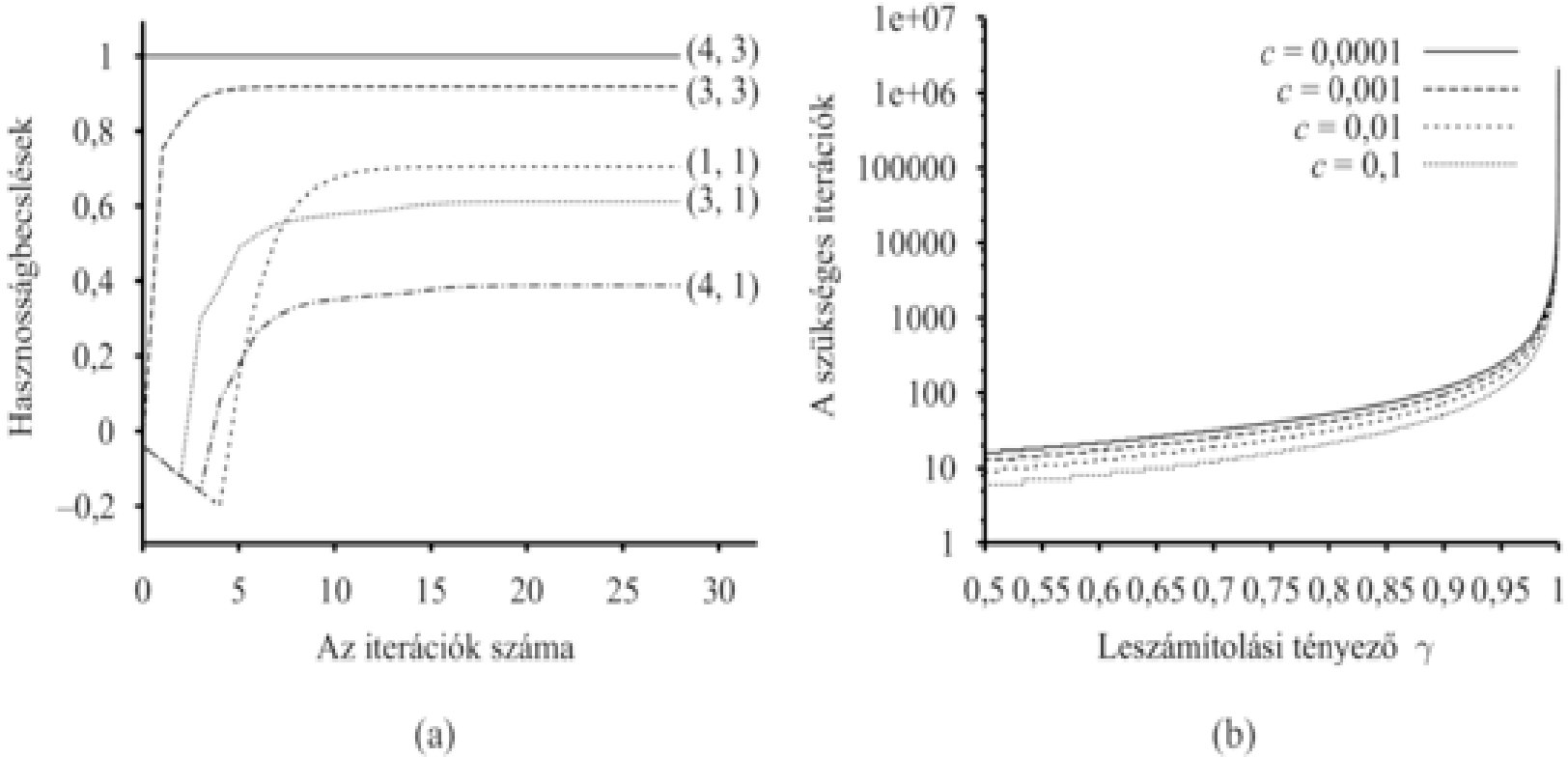
$$U(1,1) = -0{,}04 + \gamma \max\{$$
$$0{,}8\,U(1,2) + 0{,}1\,U(2,1) + 0{,}1\,U(1,1) \quad \text{(Fel)},$$
$$0{,}8\,U(2,1) + 0{,}1\,U(1,2) + 0{,}1\,U(1,1) \quad \text{(Jobbra)},$$
$$0{,}9\,U(1,1) + 0{,}1\,U(1,2) \quad \text{(Balra)},$$
$$0{,}9\,U(1,1) + 0{,}1\,U(2,1)\} \quad \text{(Le)}$$

Sajnos nemlineáris – iteráció!



Bellman-frissítés:

$$U_{t+1}(s) = R(s) + \gamma \max_a \sum_{s'} T(s,a,s') U_t(s')$$



A hasznosságok fejlődése

A szükséges értékiterációk száma, hogy a hiba garantáltan legfeljebb $\varepsilon = c\, R_{max}$ legyen

Eljárásmód-iteráció

(optimális eljárásmódot is kaphatunk, ha a hasznosság becslése pontatlan - ha egy cselekvés egyértelműen jobb, akkor a releváns állapotok pontos hasznosságát nem szükséges precízen tudnunk)

$$U(s) = R(s) + \gamma \max_a \sum_{s'} T(s,a,s')\,U(s')$$

Eljárásmód-értékelés

Egy adott π_i eljárásmódnál számítsuk ki $U_i = U^{\pi i}$-t, az egyes állapotok hasznosságát mintha π_i volna végrehajtva.

lineáris!

$$U_t(s) = R(s) + \gamma \sum_{s'} T(s,\pi_t(s),s')\,U_t(s')$$

$$U_{t+1}(s) \leftarrow R(s) + \gamma \sum_{s'} T(s,\pi_t(s),s')\,U_t(s')$$

Eljárásmód-javítás

Módosított eljárásmód-iteráció

for each s állapotra **in** S **do**

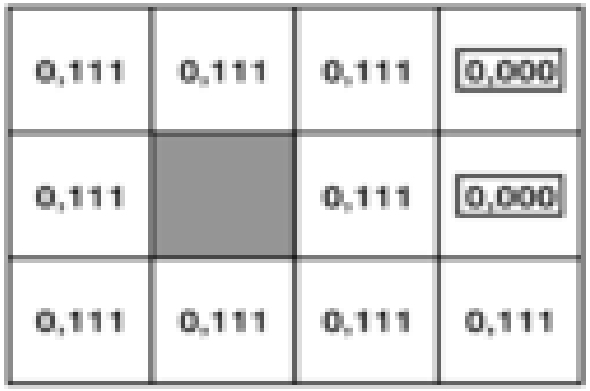
if $\max_a \sum_{s'} T(s,a,s')U[s'] > \sum_{s'} T(s,\pi[s],s')U[s']$ **then**

$$\pi[s] \leftarrow \text{argmax}_a \sum_{s'} T(s,a,s')U[s']$$

Részlegesen megfigyelhető Markov döntési folyamat

Kezdőállapot: S0
Állapotátmenet-modell: T(s, a, s′)
Jutalomfüggvény: R(s), v. R(s, a, s′)
Megfigyelési modell,
az s állapotban az o megfigyelés érzékelésének a valószínűsége
O(s, o)

Hiedelmi állapot = $b(s)$ = eloszlás állapotok felett



$$\left\langle \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, 0, 0 \right\rangle$$

$$b'(s') = \alpha O(s', o) \sum_{s} T(s, a, s') b(s)$$

(szűrés)

$$P(o \mid a,b) = \sum_{s'} P(o \mid a,s',b)\, P(s' \mid a,b)$$

$$= \sum_{s'} O(s',o) P(s' \mid a,b) = \sum_{s'} O(s',o) \sum_{s} T(s,a,s') b(s)$$

$$\tau(b,a,b') = P(b' \mid a,b) = \sum_{o} P(b' \mid o,a,b)\, P(o \mid a,b)$$

$$= \sum_{o} P(b' \mid o,a,b) \sum_{s'} O(s',o) \sum_{s} T(s,a,s') b(s)$$

$$\rho(b) = \sum_{s} b(s) R(s)$$

RMMDF megoldása a fizikai (véges) állapottérben redukálható egy MDF megoldására a hozzá tartozó hiedelmi állapot térben
(val. eloszlások folytonos terében)