TERESINHA DE MARIA BEZERRA SAMPAIO XAVIER
Professor Assistente do Instituto de Matemática
da Universidade Federal do Ceará

PROGRAMAÇÃO DINÂMICA ESTOCÁSTICA

Trabalho apresentado no IME da USP
para obtenção do grau de Mestre em
Probabilidade e Estatística Teórica

JULHO DE 1972

- SÃO PAULO -

§ 0 - INTRODUÇÃO

Nêste trabalho, consideraremos a formulação geral de problemas de Progração de Dinâmica estocástica, na linha dos artigos de D. Blackwell (1) (2), sendo particularmente estudada a questão da existência de estratégias ótimas, considerado um fator de desconto β, $0 \leq \beta < 1$.

Inicialmente, abordaremos o "caso discreto", significando que tanto o conjunto de "estados" de um sistema, bem como o conjunto de "ações" capazes de serem empreendidas, são conjuntos discretos, ou mais precisamente, finitos; o tempo será, sempre, uma variável assumindo os valores 0, 1, 2, ... (ao tratarmos o "caso discreto", teremos também oportunidade de examinar a situação limite $\beta = 1$). Nêste caso, o problema de programação dinâmica estocástica fica completamente especificado quando são fornecidos 4 objetos, $\mathcal{S}$, A, q, r, a saber:

$\mathcal{S}$: o conjunto de "estados" s do sistema;

A : o conjunto de "ações";

q : aplicação servindo para nos indicar como é selecionado (probabilisticamente) cada estado $\underline{s}$; num determinado instante, sabendo-se que no instante anterior o sistema se encontrava no estado $\underline{s}$ e que foi escolhida a ação $\underline{a}$, sendo estas probabilidades condicionais, de transição, notadas $q(s'|s,a)$;

r : aplicação servindo para medir os ganhos $r(s,a,s')$, nas transições sucessivas de cada estado $\underline{s}$ a cada estado $\underline{s'}$ (com escolha da ação $\underline{a}$); observe-se que um ganho $r(s,a,s') < 0$ se interpreta como uma perda ou custo.

Definição 0.1

Uma "regra de ação" é cada aplicação $f: \mathcal{S} \to A$ (que, portanto, faz corresponder a cada estado $s \in \mathcal{S}$, uma ação $a = f(s)$). Representaremos por F o conjunto de tôdas estas funções.

Definição 0.2

Uma estratégia (ou política) π é uma sequência $(f_1, f_2, \ldots)$, tal que $f_n \in F$, para $n = 1, 2, \ldots$.

Cada f_n se interpreta como a regra de ação a ser utilizada na transição do $(n-1)$-ésimo instante ao n-ésimo, de sorte que se $\underline{s}$ é o estado do sistema no $(n-1)$-ésimo instante, a ação escolhida será $f(s)$ ficando a probabilidade do mesmo evoluir para o estado $\underline{s'}$, no n-ésimo instante, cada por $q(s'|s,f_n(s))$, sendo o ganho correspondente $r(s,f_n(s),s)$.

Precisamente, estaremos interessados na existência de estratégias ótimas, ou seja, estratégias que garantem um ganho total esperado máximo, na perspectiva de toda a evolução do sistema (ou seja, em relação ao futuro infinito, pois consideraremos esta evolução ao longo de todos os instantes $n = 0, 1, 2, \ldots$); no momento devido, definiremos o que se deve entender por "ganho total esperado" mas que de resto, adiantamos se tratar da definição a mais natural possível. Para garantir a convergência matemática do ganho total esperado, somos obrigados a considerar um "desconto" β, $0 \leq \beta < 1$ ($\beta = 1$ corresponde à não existência de desconto).

O problema de Programação Dinâmica estocástica será também estudado numa situação mais geral, em que $\mathcal{S}$ e A são "conjuntos de Borel"; adiamos para o § 6 e subsequentes a apresentação de mais detalhes sôbre o assunto.

Para fins de uniformidade, tivemos de fazer algumas modificações nas notações originais, dadas nos artigos de Brackwell.

A G R A D E C I M E N T O S

Aproveitamos a oportunidade para expressar nossa gratidão ao Prof. Dr. Carlos Alberto Barbosa Dantas pela inestimável orientação, não só no preparo dêste trabalho, como também na fase de consolidação de nossos conhecimentos na área da Probabilidade e Estatística Teórica, feita especialmente através de curso de leitura em "Probabilidade Avançada", de curso regular sôbre "Convergência de Medidas", bem como seminário, todos por êle dirigidos, no Instituto de Matemática e Estatística, no 2º semestre de 1970 e durante o ano de 1971.

Desejamos, ainda, deixar consignados nossos agradecimentos ao Chefe do Departamento de Estatística - Prof. Dr. Lindo Fava - e a todos os colegas e amigos do mesmo Departamento pelo excelente ambiente de trabalho e camaradagem, sem o que nossa tarefa teria sido mais árdua.

Finalmente, ao Diretor do Instituto de Matemática da Universidade Federal do Ceará, Prof. Airton Fontenele Sampaio Xavier (por "acaso", meu marido), por todo o apôio e estímulo.

São Paulo, julho de 1972

Teresinha de Maria Beserra Sampaio Xavier

§ 1. *MATRIZES DE MARKOV*

As "matrizes de Markov" constituem importante instrumental matemático para a abordagem do caso discreto de problemas de programação dinâmica estocástica, especialmente na situação limite $\beta = 1$; por tal motivo, são elas previamente estudadas, nêste parágrafo. Referências bibliográficas para o assunto, constituem-se os livros de Bellman (3), Gantmacher (4) e Kemeny e Snell (5).

Definição 1.1

Uma matriz quadrada $Q = (q_{ij})$, de ordem $S \times S$, diz-se uma matriz de Markov quando:

(i) $0 \leq q_{ij} \leq 1, \ \forall\ (i,j) \in \{1, 2, \ldots, S\}^2$ e

(ii) $\sum_{j=1}^{S} q_{ij} = 1, \ \forall\ i \in \{1, 2, \ldots, S\}$

Exemplos: se o conjunto de estados $\mathcal{S}$ é finito, ou seja, $\mathcal{S} = \{1, 2, \ldots, S\}$, e é fixada uma ação $a \in A$ (vide § 0), então $Q_a = (q_{ss'})$ é uma matriz de Markov, com $q_{ss'} = q(s'|s,a)$; a matriz de identidade $I = (\delta_{ss'})$, onde os $\delta_{ss'}$ são deltas de Kronecker, é uma matriz de Markov.

Proposição 1.2

O produto Q_1Q_2 de duas matrizes de Markov Q_1 e Q_2, de mesma ordem $S \times S$, é também uma matriz de Markov.

Dem.

Sejam $Q_1 = (p_{ij})$ e $Q_2 = (q_{jk})$ matrizes de Markov; sejam $Q_1Q_2 = (\alpha_{ij})$.

Teremos:

(i) $0 \le \alpha_{ik} = \sum_{j=1}^{S} p_{ij}\, q_{jk} \le \sum_{j=1}^{S} p_{ij} = 1;$

(ii) $\sum_{k=1}^{S} \alpha_{ik} = \sum_{k=1}^{S} \left(\sum_{j=1}^{S} p_{ij} q_{jk} \right) = \sum_{j=1}^{S} \left[p_{ij} \left(\sum_{j=1}^{S} q_{jk} \right) \right] = \sum_{j=1}^{S} p_{ij} = 1$, logo,

$Q_1 Q_2$ é uma matriz de Markov.

C.Q.D.

Corolário

Se $Q_1, Q_2, \ldots, Q_n$ são matrizes de Markov, de mesma ordem S X S, também o é $Q_1, Q_2, \ldots, Q_n$; em particular, se Q é uma matriz de Markov, também o são as matrizes Q^n, $n = 0, 1, 2, \ldots$. (onde $Q^0 = I$).

Proposição 1.3

Se $Q_1, Q_2, \ldots, Q_n$ são matrizes de Markov, de mesma ordem S X S, também o é a matriz,

$$P = \frac{Q_1 + Q_2 + \ldots + Q_n}{n}$$

Dem.

Sejam $Q_k = (q_{ij}^{(k)})$, $k = 1,2,\ldots,n$; seja $P = (\alpha_{ij})$. Teremos:

(i) $0 \le \alpha_{ij} = \dfrac{\sum_{k=1}^{n} q_{ij}^{(k)}}{n} = \dfrac{\sum_{k=1}^{n} 1}{n} \le 1;$

(ii) $\sum_{j=1}^{S} \alpha_{ij} = \sum_{j=1}^{S} \left(\dfrac{\sum_{k=1}^{n} q_{ij}^{(k)}}{n} \right) = \dfrac{\sum_{k=1}^{n} \left(\sum_{j=1}^{S} q_{ij}^{(k)} \right)}{n} = \dfrac{\sum_{k=1}^{n} 1}{n} = 1;$

$i = 1, 2, \ldots, S$; logo, P é de Markov.

C.Q.D.

Proposição 1.4

Uma matriz $Q = (q_{ij})$, de ordem S X S, não negativa (isto é, $q_{ij} \ge 0$, $i,j = 1,2,\ldots,S$) é de Markov, se e só tem autovetor u, tal que $u^t = (1, 1, \ldots, 1)$ com autovetor correspondente igual a 1.

Dem.

Com efeito, se Q é não negativo, tem-se $Qu = u$, $u^t = (1,1,\ldots,1)$ se e só se a soma dos termos de cada linha é igual a unidade, isto é, quando Q é de Markov.

C.Q.D.

Proposição 1.5

Se λ é um autovalor (não trivial, isto é, correspondente a um autovetor não nulo), para uma matriz de Markov Q, então $|\lambda| \leq 1$.(*)

Dem.

Seja λ um autovalor, com autovetor $v = 0$; notemos $v^t = (v_1, \ldots, v_n)$, onde v^t é o transposto do vetor coluna v; então, $Qv = \lambda v$.

Se $m = \max |v_j|$ (é óbvio que $m > 0$), e notando $[Qv]_i$ a i-ésima componente de Qv, segue-se:

$$|\lambda| m = \max_i |\lambda v_i| = \max_i |[Qv]_i| = \max_i \left|\sum_{j=1}^{S} q_{ij} v_j\right| \leq \max_i \left(\sum_{j=1}^{S} m q_{ij}\right) =$$

$$= m\left(\max_i \sum_{j=1}^{S} q_{ij}\right) = m, \quad \text{donde } |\lambda| \leq 1.$$

C.Q.D.

Proposição 1.6

Se Q é uma matriz de Markov, então existe outra matriz de Markov Q^*, tal que:

$$Q^* = \lim_{n\to\infty} (Q + \ldots\ldots + Q^n)/ n. \quad *$$

Dem. Remetemos ao Apêndice I.

Lema 1.7

Se Q é uma matriz de Markov, de ordem $S \times S$; então:

(a) A sequência $\dfrac{I + Q + Q + \ldots\ldots + Q^n}{n+1}$ converge para uma matriz Q, quando $n \longrightarrow \infty$, tal que:

(*) - O autovalor $\lambda = 1$ é, portanto, dominante.

$$Q\,Q^* = Q^*Q = Q^*Q^* = Q^*;$$

Além disto, para um inteiro $n > 0$, se tem:

$$(Q - Q^*)^n = Q^n - Q^*.$$

(b) $I - (Q - Q^*)$ é não singular e satisfaz a

$$H(\beta) = \left[\sum_0^{\infty} \beta^n\,(Q^n - Q^*)\right] \longrightarrow H = (I - Q + Q^*)^{-1} - Q^*, \text{ quando}$$

$\beta \longrightarrow 1$ $(\beta < 1)$.

(c) $H(\beta)\,Q^* = Q^*H(\beta) = H\,Q^* = Q^*H = 0$

e

$$(I - Q)\,H = H\,(I - Q) = I - Q^*.$$

(d) posto $(I - Q)$ + posto $Q^* = S$

(e) Para todo vetor coluna c, de ordem $S \times 1$, o sistema:

$$\begin{cases} Q\,x = x \\ Q^*x = Q^*c \end{cases}$$

tem solução única.

<u>Dem. de (a)</u>

A existência da matriz limite Q^* já foi provada, conforme proposição, ver Apêndice I.

Resta finalmente mostrar:

(i) $\boxed{Q\,Q^* = Q^*\,Q}$

De fato,

$$Q\,\frac{I + Q + \ldots + Q^n}{n+1} = \frac{Q + Q^2 + \ldots + Q^{n+1}}{n+1} = \frac{I + Q + \ldots + Q^n}{n+1}\,Q,$$

donde,

$$Q\left(\lim_{n\to\infty} \frac{I + Q + \ldots + Q^n}{n+1}\right) = \left(\lim_{n\to\infty} \frac{I + Q + \ldots + Q^n}{n+1}\right) Q \quad \text{isto é,}$$

$$Q\,Q^* = Q^*\,Q$$

(ii) $\boxed{Q^*Q = Q^*}$

De fato,

$$Q^*Q = \lim_{n\to\infty} \frac{I + Q + \ldots + Q^n}{n+1} . Q = \lim_{n\to\infty}$$

$$= \lim_{n\to\infty} \frac{Q + Q^2 + \ldots + Q^{n+1}}{n+1} = \lim_{n\to\infty} \frac{I + Q + Q^2 + \ldots + Q^{n+1}}{n+1} =$$

$$= \lim_{n\to\infty} \frac{I + Q + Q^2 + \ldots + Q^{n+1}}{n+2} . \lim_{n\to\infty} \frac{n+2}{n+1} = Q^*$$

(iii) $\boxed{Q^*Q^* = Q^*}$

Observe-se, inicialmente, que em virtude do resultado precedente se tem $Q^nQ^* = Q^*$, para todo n (é uma verificação trivial, por indução finita); logo:

$$Q^*Q^* = (\lim_{n\to\infty} \frac{I + Q + \ldots + Q^n}{n+1})Q^* = \lim_{n\to\infty} \frac{Q^* + QQ^* + \ldots Q^nQ^*}{n+1} =$$

$$= \lim_{n\to\infty} \frac{(n+1)\, Q^*}{n+1} = Q^*$$

(iv) $\boxed{(Q - Q^*)^n = Q^n - Q, \text{ se } n > 0}$

Prova-se, sem dificuldades, por indução finita, usando o fato

$$Q^nQ^* = Q^*.$$

C.Q.D.

<u>Dem. de (b)</u>

$$\text{Tem-se } H(\beta) + Q^* = \left[\sum_{n=0}^{\infty} \beta^n (Q^n - Q^*)\right] + Q^* = (I - Q^*) +$$

$$+ \left[\sum_{n=1}^{\infty} \beta^n (Q - Q^*)^n\right] + Q^* = I + \sum_{n=1}^{\infty} \beta^n (Q - Q^*)^n =$$

$$= \sum_{n=0}^{\infty} \beta^n (Q - Q^*)^n;$$

segue-se:

$$(H(\beta)+Q^*) \left[I-\beta(Q-Q^*)\right] = \sum_{n=0}^{\infty} \beta^n (Q-Q^*)^n - \sum_{n=1}^{\infty} \beta^{n+1} (Q-Q^*)^{n+1} = I,$$

isto é, $\qquad (H(\beta) + Q^*) \left[I - \beta(Q - Q^*)\right] = I;$

em seguida, obtemos:

$(H(\beta) + Q^*)\ (I - Q + Q^*) =$

$= (H(\beta) + Q^*)\ \left[I - \beta(Q - Q^*) - (1 - \beta)(Q - Q^*)\right] = I - H(\beta)(1-\beta)(Q-Q^*) -$

$- (1 - \beta)(Q^* Q - Q^*Q^*) = I - H(\beta)(1 - \beta)(Q - Q^*)$, isto é:

$$\boxed{(H(\beta) + Q^*)(I - Q + Q^*) = I - H(\beta)(1 - \beta)(Q - Q^*)} \qquad (1)$$

Ora, a condição de <u>somabilidade a Cesaro</u>(*) de (Q^n) para Q^*, isto é, $\sum_{k=0}^{\infty} \frac{Q^k}{n+1} \longrightarrow Q^*$, quando $n \longrightarrow \infty$, implica a condição de <u>somabilidade a Abel</u>(**) de (Q_n) para Q^*; ou seja:

$(1 - \beta) \sum_{n=0}^{\infty} \beta^n Q^n \longrightarrow Q^*$, quando $\beta \longrightarrow 1$ $(\beta < 1)$, e como

$\sum_{n=0}^{\infty} \beta^n Q^* = \frac{Q^*}{1 - \beta}$, $\beta < 1$, segue-se que $(1 - \beta) \sum_{n=0}^{\infty} \beta^n (Q^n - Q^*) =$

$= (1 - \beta)\ H(\beta) \longrightarrow 0$, quando $\beta \longrightarrow 1$ $(\beta < 1)$.

Aplicando êste resultado em (1), obtemos:

$(H(\beta) + Q^*)(I - Q + Q^*) \longrightarrow I$, quando $\beta \longrightarrow 1$ $(\beta < 1)$;

analogamente,

$(I - Q + Q^*)\ (H(\beta) + Q^*) \longrightarrow I$, quando $\beta \longrightarrow 1$ $(\beta < 1)$;

em consequência, $I - Q + Q^*$ é uma matriz não singular, e se tem:

$$H(\beta) \longrightarrow H = (I - Q + Q^*)^{-1} - Q^*.$$

(*) Seja $\sum_{n=0}^{\infty} A_n$ uma série e notemos $S_n = \sum_{k=0}^{n} A_k$ (n-ésima soma parcial); seja (S_n') a sucessão das médias aritméticas $S_n' = (S_0+S_1+\ldots+S_n)/n+1$. Diz-se que a série ΣA_n é <u>somável a Cesaro</u> se S_n' converge para S_1, chamado <u>soma de Cesaro</u>.

(**) Diz-se que a série ΣA_n é <u>somável a Abel</u> quando existem as somas $S_\beta = (1-\beta \sum_{n=0}^{\infty} \beta^n A_n)$ para β suficientemente próximo de 1, e além disso, S_β converge para S_2, quando $\beta \longrightarrow 1 - 0$ (convergência à esquerda); S_2 é a <u>soma de Abel</u> da série.

Dem. de (c)

São verificações de rotina:

(i) $\boxed{H(\beta)\, Q^* = 0}$

De fato,

$$H(\beta)\, Q^* = \left[\sum_0^\infty \beta^n \,(Q^n - Q^*)\right] Q^* = \sum_0^\infty \beta^n (Q^n Q^* - Q^* Q^*) = \sum_0^\infty \beta^n (Q^* - Q^*) = 0;$$

(ii) $\boxed{Q^* H(\beta) = 0}$ (análogo);

(iii) $\boxed{H\, Q^* = 0}$

De fato,

$$H\, Q^* = (\lim_{\substack{\beta\to 1\\ \beta<1}} H(\beta))\, Q^* = \lim_{\substack{\beta\to 1\\ \beta<1}} (H(\beta)\, Q^*) = 0;$$

(iv) $\boxed{Q^* H = 0}$ (análogo);

(v) $\boxed{(I - Q)\, H = I - Q^*}$

Tem-se:

$$(I - Q) H(\beta) = (I - Q) \sum_0^\infty \beta^n (Q^n - Q^*) = \sum_0^\infty \beta^n \,(Q^n - Q^*) -$$

$$- \sum_0^\infty \beta^n \,(Q^{n+1} - Q^*) = \sum_0^\infty \beta^n \,(Q^n - Q^*) - \frac{1}{\beta} \sum_0^\infty \beta^n \,(Q^n - Q^*) +$$

$$+ \frac{I - Q}{\beta} = H(\beta) - \frac{H(\beta)}{\beta} + \frac{I - Q^*}{\beta};$$

segue-se:

$$(I - Q)\, H = (I - Q) \lim_{\substack{\beta\to 1\\ \beta<1}} h(\beta) = \lim_{\substack{\beta\to 1\\ \beta<1}} \left[(I - Q)\, H(\beta) = \right.$$

$$\left. = \lim_{\substack{\beta\to 1\\ \beta<1}} H(\beta) - \frac{H(\beta)}{\beta} + \frac{I - Q^*}{\beta} \right] = I - Q$$

(vi) $\boxed{M(I - Q) = I - Q^*}$ análogo;

<u>Dem. de (d)</u>

Decorre de (c) que a matriz $I - Q + Q^*$ é não singular, e logo,

$$\boxed{S = \text{posto } (I - Q + Q^*)} \;;$$

por outro lado, do conhecido fato posto $(A + B) \leq$ posto A + posto B, obtemos:

$$\boxed{\text{posto } (I - Q + Q^*) \leq \text{posto } (I - Q) + \text{posto } Q^*}$$

Além disso, se x é um vetor coluna de ordem $S \times 1$, não nulo, tem-se:

$$(I - Q)\, x = 0 \longrightarrow Q\, x = x \longrightarrow Q^n x = x \;\forall\, n = 1, 2, \ldots\ldots \longrightarrow$$
$$\longrightarrow \frac{I - Q + \ldots + Q^n}{n + 1} . x = x, \;\forall\, n = 1, 2, \ldots \longrightarrow Q^* x = x \neq 0;$$

ou seja, se $x \neq 0$ pertence ao "espaço solução" de $I - Q$, então não pertence ao "espaço solução" de Q^*; notando $< I - Q >$ e $< Q^* >$ êstes espaços, tem-se:

$$< I - Q > \cap < Q^* > \; = \{0\},$$

donde

$$\dim < I - Q > + \dim < Q^* > \; \leq \; S;$$

e como,

$$\begin{cases} \text{posto } (I - Q) + \dim < I - Q > = S \\ \text{posto } (Q^*) + \dim < Q^* > = S, \end{cases}$$

segue-se:

$$\boxed{\text{posto } (I - Q) + \text{posto } Q^* \leq S}$$

Finalmente, a igualdade e as desigualdades em cercaduras, garantem que

$$\text{posto } (I - P) + \text{posto } Q = S.$$

<u>Dem. de (e)</u>

<u>Unicidade</u>

Suponhamos que x e y sejam soluções; em consequência, teremos:

$$\begin{cases} (I - Q)\ (x - y) = 0 \\ Q^{*}(x - y) = 0 \end{cases}$$

ou seja, $x - y$ pertence a ambos os espaços soluções $<I - Q>$ e $<Q^{*}>$, como $<I - Q> \bigcap <Q^{*}> = \{0\}$, segue-se:

$$x - y = 0, \text{ isto é, } x = y.$$

<u>Existência</u>

A solução é $x = Q^{*}c$.

§ 2. CASO DISCRETO - PRELIMINARES

O caso discreto é caracterizado pelo fato de serem finitos, tanto o conjunto $\mathcal{S}$ de estados, como o conjunto A de ações; notaremos $\mathcal{S} = \{1, 2, \ldots\ldots, S\}$ (*)

As notações e definições introduzidas no § 0, são mantidas. O ganho (num dado instante), porém, dependerá apenas do estado $\underline{s}$ e da ação $\underline{a}$ empreendida, naquêle instante (sendo independente do estado s', no instante seguinte); tal motivo, notaremos $r(s, a)$ em vez de $r(s,a,s')$

Uma estratégia $\Pi = (f_1, f_2, \ldots.)$ será notada $\Pi = (f_n)$, subtendendo-se $n = 1, 2, \ldots\ldots$

<u>Definições 2.1</u>

Fixada uma estratégia $\Pi = (f_1, f_2, \ldots..)$ e dadas regras de ação $g_1, g_2, \ldots.., g_N \in F$, então:

(a) A estratégia $(g_1, g_2, \ldots.., g_N, \Pi) = (h_n)$, é tal que $h_n = g_n$, se $1 \leq n \leq N$; $h_n = f_{n-N}$, se $n > N$; isto é, $g_1, q_2, \ldots, g_N, \Pi =$ $= (g_1, g_2 m \ldots., g_N, f_1, f_2, \ldots..)$;

(b) $g^{(N)}, \Pi = \underbrace{g, g, \ldots\ldots\ldots, g}_{N \text{ vêzes}}, \Pi$;

(c) $g^{(\infty)} = (g, g, \ldots\ldots.)$;

(d) $T_\Pi = (f_2, f_3, \ldots..)$.

<u>Definições 2.2</u>

Dada uma regra de ação $f \in F$, então:

(*) A rigor, deveria designar-se "caso finito".

(a) $r(f)$ é o vetor coluna $S\times 1$ cuja s-ésima componente é o ganho $r(s,f(s))$, ou seja, o ganho que corresponde à passagem pelo estado $\underline{s}$ a escolha da ação $f(s)$; isto é, $[r(f)]_s = r(s,f(s))$.

(b) $Q(f)$ é a matriz quadrada $S \times S$, cujo termo de ordem (ss') é a probabilidade de transição $q(s'|s,f(s))$; isto é, $[Q(f)]_{ss'} = q(s'|s,f(s))$. Logo, $r(f)$ e $Q(f)$ dão o ganho e a lei da transição dependendo do estado atual $\underline{s}$ e de uma regra de ação f escolhida no momento.

Se usarmos a estratégia $\Pi = (f_n)$ e se o sistema se encontra inicialmente (isto é, no instante $\underline{0}$) no estado $\underline{s}$, a probabilidade do mesmo atingir o estado $\underline{s'}$, no instante n, é o termo de ordem (ss') da matriz produto:

$$Q_n(\Pi) = Q(f_1)\, Q(f_2) \ldots Q(f_n).$$

Fixando-se uma estratégia $\Pi = (f_n)$, e um fator desconto β, $0 \leq \beta <$ o ganho total esperado descontado é dado pelo vetor coluna $S \times 1$;

$$\boxed{V(\Pi) = \sum_{n=0}^{\infty} \beta^n\, Q_n(\Pi)\, r(f_{n+1})} \qquad (1)$$

Sua s-ésima componente $[V(\Pi)]_s$ se interpreta como o ganho esperado (corrigido), quando o estado inicial é s $(s = 1, 2, \ldots, S)$.

Com efeito, fixados um estado inicial s e uma estratégia $\Pi = (f_n)$ é óbvio que o ganho esperado, na transição do n-ésimo ao (n+1)-ésimo instante $(n \geq 1)$, é dado por:

$$\sum_{s'=1}^{S} [r(f_{n+1})]_{s'}\, [Q_n(\Pi)]_{ss'} = [Q_n(\Pi)\, r(f_{n+1})]_s;$$

nas mesmas condições, tomando $Q_0(\Pi) = I$ (matriz identidade) tem-se o ganho esperado, na transição do instante inicial ao instante $n = 1$, dado por: $[r(f_1)]_s = [Q_0(\Pi)\, r(f_1)]_s$; logo, o ganho total esperado (sem desconto) é dado por:

$$\sum_{n=0}^{\infty} [Q_n(\Pi)\, r(f_{n+1})]_s ;$$

se introduzirmos um fator de desconto β, $0 \leq \beta < 1$, de sorte que na transição do n-ésimo ao (n+1)-ésimo instante $(n \geq 0)$, o ganho seja corrigido através do fator β^n, tem-se o ganho total esperado:

$$[V(\Pi)]_s = \sum_{n=0}^{\infty} \beta^n [Q_n(\Pi)\, r(f_{n+1})]_s = \left[\sum_{n=0}^{\infty} \beta^n\, Q_n(\Pi)\, r(f_{n+1})\right]_s .$$

<u>Proposição 2.3</u>

A série que define cada componente de $V(\Pi)$ conforme (1), é vergente.

Como $\mathcal{S}$ e A são finitos, existe $\alpha = \max|r(s,a)|$, $0 \le \alpha < \infty$, logo, $0 \le \left|[r(f)]_s\right| = \left|r(s, f(s))\right| \le \alpha$, $\forall f \in F$, $\forall s \in \mathcal{S}$, donde:

$$0 \le \left|[Q_n(\Pi)\ r(f_{n+1})]_s\right| = \left|\sum_{s'=1}^{S} [r(f_{n+1})]_{s'}\ [Q_n(\Pi)]_{ss'}\right| \le$$

$$\le \alpha \sum_{s'=1}^{S} |Q_n(\Pi)|_{ss'} = \alpha\ ,\ n \ge 0$$ (lembrar que cada $Q_n(\Pi)$ é uma matriz de Markov).

Portanto,

$$0 \le \left|[V(\Pi)]_s\right| \equiv \left|\sum_{n=0}^{\infty} \beta^n [Q_n(\Pi)\ r(f_{n+1})]_s\right| \le \alpha \sum_{n=0}^{\infty} \beta^n = \frac{\alpha}{1-\beta} < \infty$$

<u>Preposição 2.4</u>

$$V(\Pi) = r(f_1) + \beta Q(f_1) \sum_{n=1}^{\infty} \beta^{n-1}\ Q_{n-1}(T\Pi)\ r(f_{n+1}) =$$

$$\equiv r(f_1) + \beta Q(f_1)\ V(T\Pi).$$

<u>Dem.</u>

Se $\Pi = (f_1, f_2, \ldots\ldots)$, então $T\Pi = (h_1, h_2, \ldots\ldots)$, com $h_1 = f_2$, $h_2 = f_3, \ldots\ldots$; segue-se:

$$V(\Pi) = \sum_{n=0}^{\infty} \beta^n\ Q_n(\Pi)\ r(f_{n+1}) =$$

$$= r(f_1) + \sum_{n=1}^{\infty} \beta^n\ Q_n(\Pi)\ r(f_{n+1}) =$$

$$= r(f_1) + \sum_{n=1}^{\infty} \beta^n [Q(f_1)Q(f_2)\ldots\ldots Q(f_n)]r\ (f_{n+1}) =$$

$$= r(f_1) + \beta Q(f_1) \sum_{n=1}^{\infty} \beta^{n-1} [Q(h_1)\ \ldots\ Q(h_{n-1})]r(h_n) =$$

$$= r(f_1) + \beta Q(f_1) \sum_{n=1}^{\infty} \beta^{n-1}\ Q_{n-1}(T\Pi)\ r(h_n) =$$

$$= r(f_1) + \beta Q(f_1) \sum_{m=0}^{\infty} \beta^m\ Q_m\ (T\Pi)\ r(h_{m+1}) =$$

$$= r(f_1) + \beta Q(f_1)\ V(T\Pi).$$

C.Q.D.

Definição 2.5

Dada a regra de ação $f \in F$, a transformação $L(f)$ aplica cada vetor coluna w, de ordem $S \times 1$, em $L(f)\, w = r(f) + \beta Q(f)\, w$.

Proposição 2.6

(a) $V(f,\Pi) = L(f)\, V(\Pi)$

(b) $V(f_1, \ldots, f_N, \Pi) = L(f_1) \ldots L(f_N) V(\Pi)$.

Dem.

(a) Se $\Pi = (f_1, f_2, \ldots)$, teremos: $(f, \Pi) = (f, f_1, f_2, \ldots)$
$T(f, \Pi) = \Pi$. Logo, aplicando a Prop. 2.4 e Def. 2.5,

$$V(f, \Pi) = r(f) + \beta Q(f)\, V(\Pi) = L(f)\, V(\Pi).$$

(b) Provaremos por indução finita.

1º) Vale para $N = 1$, conforme já demonstrada no item precedente.

2º) Valendo para 1 e N, vale para $N + 1$; com efeito:

$$V(f_1, \ldots f_N, f_{N+1}, \Pi) = V(f_1, \ldots, f_N; (f_{N+1}, \Pi)) =$$
$$= L(f_1) \ldots L(f_N)\, V(f_{N+1}, \Pi) = L(f_1) \ldots L(f_N)\, L(f_{N+1})\, V(\Pi).$$

Dados dois vetores colunas w_1 e w_2, de mesma ordem $S \times 1$, notaremos $w_1 \geq w_2$, quando $(w_1)_s \geq (w_2)_s$, $s = 1, 2, \ldots S$ (isto é, a desigualdade é válida para todas correspondentes componentes de w_1 e w_2). Notaremos $w_1 > w_2$, quando $w_1 \geq w_2$ e existe pelo menos um índice s tal que $(w_1)_s > (w_2)_s$.

Proposição 2.7

$L(f)$ é uma transformação monótona, ou seja, $w_1 \geq w_2$ implica $L(f)w_1 \geq L(f)w_2$.

Dem.

Se $w_1 \geq w_2$, então $Q(f)w_1 \geq Q(f)w_2$, pois $Q(f)$ é uma matriz cujos termos são todos não negativos; segue-se:

$$L(f)w_1 = r(f) + \beta Q(f)w_1 \geq r(f) + \beta Q(f)w_2 \geq L(f)w_2.$$

C.Q.D.

<u>Definição 2.8</u>

(a) Dadas duas estratégias Π_1 e Π_2, notaremos $\Pi_1 \geq \Pi_2$, quando $V(\Pi_1) \geq V(\Pi_2)$, ou seja, quando o ganho total esperado de Π_1 é maior ou igual ao ganho esperado de Π_2, qualquer que seja o estado inicial s. ($\Pi_1 \geq \Pi_2$ entende-se: 'Π_1 é melhor que Π_2').

(b) Uma estratégia Π^* diz-se <u>ótima</u> quando $\Pi^* \geq \Pi$ qualquer que seja a estratégia Π.

§ 3. CASO DISCRETO - ESTRATÉGIAS ÓTIMAS, NO CASO $\beta < 1$

Proposição 3.1

Se $\beta < 1$, e são dadas estratégias $\Pi_1 = (f_1, f_2, \ldots\ldots)$ e Π_2, então:

$$V(f_1, f_2, \ldots, f_N, \Pi_2) \longrightarrow V(\Pi_1), \text{ quando } N \longrightarrow \infty.$$

Dem.

(a) Inicialmente verificaremos (por indução finita) que para qualquer estratégia Π, se tem:

$$V(\Pi) = \left[\sum_{n=0}^{N-1} \beta^n\, Q_n\ (\Pi)\ r(f_{n+1})\right] + \beta^N Q_N\ (\Pi)\ V(T^N \Pi),$$

onde $T^N \Pi$ é a transformação T aplicada a Π, N vezes.

1º) Vale para $N = 1$ (conforme Proposição 2.4)

2º) Mostraremos que, sendo a expressão válida para N, será válida para $N + 1$; com efeito, lembrando $T^N \Pi = (f_{N+1}, f_{N+2}, \ldots\ldots)$ e $Q_N(\Pi)\ Q(f_{N+1}) = Q_{N+1}(\Pi)$, segue-se:

$$V(\Pi) = \left[\sum_{n=0}^{N-1} \beta^n\, Q_n\ (\Pi)\ r(f_{n+1})\right] + \beta^N\, Q_N(\Pi) V(T^N \Pi) = \left[\sum_{n=0}^{N} \beta^n Q_n(\Pi) r(f_{n+1})\right] -$$

$$- \beta^N\, Q_N(\Pi) r(f_{N+1}) + \beta^N Q_N(\Pi) \left[r(f_{N+1}) + \beta Q(f_{N+1}) V(T(T^N \Pi))\right] =$$

$$= \left[\sum_{n=0}^{N} \beta^n\, Q_n(\Pi)\ r(f_{n+1})\right] + \beta^{N+1}\, Q_{N+1}\ (\Pi)\ V(T^{N+1} \Pi).$$

(b) Se em lugar de Π, tivermos $(f_1, \ldots\ldots, f_N, \Pi_2)$, e observarmos que $T^N(f_1, \ldots, f_N, \Pi_2) = \Pi_2$, segue-se, aplicando a expressão precedentemente provada:

$$V(f_1, \ldots, f_N, \Pi_2) = \left[\sum_{n=0}^{N-1} \beta^n\, Q_n(f_1, \ldots, f_N, \Pi_2)\ r(f_{N+1})\right] +$$

$$+ \beta^N Q_N(f_1, \ldots, f_N, \Pi_2).V(\Pi_2) = \left[\sum_{n=0}^{N-1} \beta^n Q_n(\Pi) r(f_{n+1})\right] + \beta^N Q_n(\Pi)\ V(\Pi_2);$$

ora, o primeiro termo do segundo membro desta igualdade converge para $V(\Pi_1) = \sum_{n=0}^{\infty} \beta^n Q_n(\Pi)\ r(f_{n+1})$, e o segundo termo, para zero, quando $N \longrightarrow \infty$, ou seja:

$$V(f_1, \ldots, f_N, \Pi_2) \longrightarrow V(\Pi_1), \text{ onde } \Pi_1 = (f_1, f_2, \ldots).$$

<u>Teorema 3.2</u>

Se $\Pi^* \geq (f,\Pi^*)$, para todo $f \in F$, então Π^* é uma estratégia ótima.

<u>Dem.</u>

Suponhamos que $\Pi^* \geq (f,\Pi^*)$, isto é, $L(f)\ V(\Pi^*) = V(f,\Pi^*) \leq V(\Pi^*)$, $\forall f \in F$.

Então, para qualquer estratégia $\Pi = (f_n)$, segue-se $L(f_N)V(\Pi^*) \leq V(\Pi^*)$, para cada índice N, e usando a monotonicidade das $L(f_i)$ $(i = 1, 2, \ldots, N-1)$, obtemos:

$$L(f_1) \ldots L(f_N)\ V(\Pi^*) \leq L(f_1) \ldots L(f_{N-1})\ V(\Pi^*),$$

que é equivalente, conforme (b) da Proposição 2.8, a:

$$V(f_1, \ldots, f_N, \Pi^*) \leq V(f_1, \ldots, f_{N-1}, \Pi^*);$$

analogamente:

$$V(f_1, \ldots, f_{N-1}, \Pi^*) \leq V(f_1, \ldots, f_{N-2}, \Pi^*);$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots;$$

$$V(f_1,\Pi^*) \leq V(\Pi^*);$$

desta sequência de desigualdades, concluimos que:

$$V(f_1, \ldots, f_N,\Pi^*) \leq V(\Pi^*),$$

donde:

$$\lim_{N\to\infty} V(f_1, \ldots, f_N, \Pi^*) \leq V(\Pi^*)$$

e pela proposição precedente, segue-se:

$$V(\Pi) \leq V(\Pi^*);$$

como Π é uma estratégia arbitrária, Π^* é ótima.

C.Q.D.

Comentário: O valor dêste Teorema consiste em permitir que se reduza a verificação de Π^* ser uma estratégia ótima, a comparação entre Π^* e as (f,Π^*), $f \in F$ (tratando-se, portanto, de comparações em número finito, pois F é finito), em vez de se ter de efetuar tôdas as comparações entre Π^* e as demais sequências $\Pi = (f_1, f_2, \ldots\ldots)$.

Teorema 3.3

Se $(f,\Pi) > \Pi$, então $f^{(\infty)} > \Pi$.

Dem.

Nossa hipótese é que $(f,\Pi) > \Pi$, isto é, $L(f)\ V(\Pi) = V(f,\Pi)V(\Pi)$; aplicando o operador monótono $L^{N-1}(f)$, obtemos:

$$V(f^N,\Pi) = L^N(f)\ V(\Pi) \geq L^{N-1}(f)\ V(\Pi);$$

por outro lado, $L^{(N-1)}(f)\ V(\Pi) > V(\Pi)$, donde:

$$V(f^N,\Pi)\ V(\Pi),$$

e por passagem ao limite (conforme Proposição 3.1):

$$V(f) > V(\Pi)$$

isto é, $f^{\infty} > \Pi$.

C.Q.D.

Em seguida, passaremos a notar $V_s(\Pi)$, em vez de $[V_s(\Pi)]$ (s-ésima componente do vetor $V(\Pi)$; analogamente, dada $f \in F$, notaremos $r_s(f)$ em vez de $[r(f)]_s$. Por outro lado, $p(s,a)$ representará o vetor-linha cuja s'-ésima coordenada é $q(s'|s,a)$; logo, conforme Def. 2.2 (b),

$p(s,f(s))$ é a s-ésima linha matriz $Q(f)$.

Definição 3.4

Dados $f \in F$ e $s \in \mathcal{S}$, então $G(s,f)$ se define como o conjunto de tôdas as ações $a \in A$ tais que:

$$r(s,a) + \beta\, p(s,a)\, V(f^{(\infty)}) > V_s(f^{(\infty)}).$$

Teorema 3.5

(a) Se $G(s,f) = \emptyset$, para cada $s \in \mathcal{S}$, então $f^{(\infty)}$ é uma estratégia ótima.

(b) Se $g \in F$ é tal que:

i) $g(s) \in G(s,f)$, para algum s,

e

ii) $g(s) = f(s)$, quando $g(s) \notin G(s,f)$,

então: $g^{(\infty)} > f^{(\infty)}$.

Dem.

(a) Como:

$$V(g,f^{(\infty)}) = r(g) + \beta Q(g)\, V(T(g,f^{(\infty)}) = r(g) + \beta Q(g)\, V(f^{(\infty)}),$$

segue-se:

$$V_s(g,f^{(\infty)}) = r_s(g) + \beta p(s,g(s))\, V(f^{(\infty)})$$

Em virtude dêste resultado, e de acordo com a Def. 3.4, dada $g \in F$, tem-se:

$$g(s) \in G(s,f) \longleftrightarrow V_s(g,f^{(\infty)}) > V_s(f^{(\infty)}).$$

Em consequência:

$$G(s,f) = \emptyset,\ \forall s \in \mathcal{S} \Longrightarrow g(s) \notin G(s,f),\ \forall s \in \mathcal{S},\ \forall g \in F \longrightarrow$$
$$\longrightarrow V_s(g,f^{(\infty)}) \leq V_s(f^{(\infty)}),\ \forall s \in \mathcal{S},\ \forall g \in F,$$

e de acordo com o Teorema 3.2, $f^{(\infty)}$ é uma estratégia ótima.

(b) Agora, suponhamos que g satisfaça a i) e ii).

De i), conforme já verificado na parte precedente desta demonstração, obtem-se:

$V_s(g,f^{(\infty)}) > V_s(f^{(\infty)})$, para algum $s \in \mathcal{S}$; enquanto de ii) obtemos:

$$V_s(g,f^{(\infty)}) = V_s(f^{(\infty)}, \text{ quando } V_s(g,f^{(\infty)}) \leq V_s(f^{(\infty)}).$$

Logo, $V(g,f^{(\infty)}) > V(f^{(\infty)})$, e de acordo com o Teorema 3.3, tem-se $g^{(\infty)} > f^{(\infty)}$.

<u>Definição 3.6</u>

Uma estratégia Π diz-se estacionária quando é da fora $\Pi = f^{(\infty)}$.

<u>Corolário 3.7</u>

uma estratégia ótima que é estacionária.

<u>Dem.</u>

Pelo Teorema 3.5, se $\Pi = f^{(\infty)}$ (estratégia estacionária), só podem ocorrer duas alternativas:

1a.) $f^{(\infty)}$ é ótima (no caso em que $G(s,f) = \emptyset$, $\forall s \in \mathcal{S}$);

ou

2a.) $f^{(\infty)}$ admite uma estratégia estacionária $g^{(\infty)}$ melhor que ela, isto é, tal que $g^{(\infty)} > f^{(\infty)}$.

Como só existe um número finito de estratégias estacionárias (pois o conjunto F é finito), existe uma que não pode ser melhorada, logo esta é ótima.

§ 4. APLICAÇÕES

Nêste parágrafo, apresentaremos dois exemplos de situações que podem ser tratadas com os métodos desenvolvidos nos parágrafos precedentes.

Observemos que o Teorema 3.5 nos fornece um algorítimo para determinação de uma estratégia estacionária ótima (considerado um desconto $\beta < 1$); êste algorítimo é conhecido como "rotina de Howard"(*), e tem duas etapas básicas:

1a.) Escolhida uma regra de ação $f \in F$, notemos $v_s = V_s(f^{(\infty)})$; observemos que:

$$V(f^{(\infty)}) = r(f) + \beta\, Q(f)\, V(f^{(\infty)}),$$

donde se segue:

$$v_s = r(s,f(s)) + \beta \sum_{s'} q(s'|s,f(s)) v_{s'},$$

$$s = 1, 2, \ldots, S;$$

o sistema de equações, acima, permite determinar os v_s.

2a.) Usando os valores $v_s (s \in S)$, passaremos ao exame dos conjuntos $G(s,f)$ definidos em 3.4; se $G(s,f) = \emptyset$, para cada $s \in$, então $f^{(\infty)}$ é uma estratégia β-ótima (isto é, ótima para o valor $\beta < 1$ prefixado).

Mas, se $G(s,f) \neq \emptyset$, para algum $s \in \mathcal{S}$, definamos $g \in F$, tal que $g(s) \in G(s,f)$, enquanto $g(s) = f(s)$ se $G(s,f) = \emptyset$; pelo Teorema 3.5, é claro que $g^{(\infty)} > f^{(\infty)}$.

Por aplicações, tantas vezes quantas forem necessárias, do algorítimo acima, acabaremos por chegar a uma estratégia estacionária ótima. No exemplo . mostramos o emprego efetivo desta rotina de cálculo.

(*) "Howard policy improvement routines".

x

<u>Exemplo</u> (Problema do Taxi)

Êste exemplo é devido a Howard[6], pág. 44 e seguintes. Assim, consideremos um taxi cujo território engloba três cidades próximas: A, B e C. O conjunto de estados a considerar é $\mathcal{S} = \{1,2,3\}$, onde 1,2, 3, respectivamente, equivale ao taxi se encontrar em A, B, C. Instantes sucessivos correspondem a corridas sucessivas

Uma vez concluída uma corrida, o taxi pode se encontrar em qualquer uma das três cidades, e o motorista tem então à sua escolha ações seguintes:

<u>ação 1</u> = circular, até encontrar um passageiro que lhe dê sinal;

<u>ação 2</u> = procurar o ponto de taxi mais próximo e aguardar em fila a sua vez(*).

Uma vez escolhida a ação, o ganho na corrida a ser feita (correspondendo a uma transição de estados) é representado pelo faturamento líquido após dedução, não só das despesas da própria corrida, como também daquelas que são uma consequência direta do tipo de ação empreendida (no caso da escolha da <u>ação 1</u>, há despesas decorrentes do tempo de circulação até o motorista conseguir passageiro; no caso da <u>ação 2</u>, despesas com o deslocamento até o ponto de taxi mais próximo).

Consideremos os dados seguintes:

$$\begin{pmatrix} p(1,1) \\ p(2,1) \\ p(3,1) \end{pmatrix} = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{pmatrix}$$

e

(*) Aqui, fazemos pequena simplificação no problema original de Howard (não consideraremos um terceiro tipo de ação possível).

$$\begin{bmatrix} p(1,2) \\ p(2,2) \\ p(3,2) \end{bmatrix} = \begin{bmatrix} 1/16 & 3/4 & 3/16 \\ 1/16 & 7/8 & 1/16 \\ 1/8 & 3/4 & 1/8 \end{bmatrix}$$

$$\begin{bmatrix} r(1,1) \\ r(2,1) \\ r(3,1) \end{bmatrix} = \begin{bmatrix} 8 \\ 16 \\ 17 \end{bmatrix} \quad e \quad \begin{bmatrix} r(1,2) \\ r(2,2) \\ r(3,2) \end{bmatrix} = \begin{bmatrix} 2,75 \\ 15 \\ 4 \end{bmatrix} \quad ;$$

vamos supor um desconto dado por $\beta = 0,9$.

Comecemos por $f \equiv 1$; obtemos o sistema:

$$\begin{cases} v_1 = \ 8 + 0,9\ (1/2\ v_1 + 1/4\ v_2 + 1/4\ v_3) \\ v_2 = 16 + 0,9\ (1/2\ v_1 + 1/2\ v_3) \\ v_3 = 17 + 0,9\ (1/4\ v_1 + 1/4\ v_2 + 1/2\ v_3) \end{cases}$$

que tem solução:

$$\begin{cases} v_1 = 91,3 \\ v_2 = 97,6 \\ v_3 = 90,0 \end{cases}$$

Devemos, em seguida, considerar os conjuntos $G(1,f)$, $G(2,f)$ e $G(3,f)$; iniciemos por $G(1,f)$.

Por definição, $G(1,f) = \{a \in A;\ r(1,a) + (0,9)\ p(1,a)\ V(f^{(\infty)}) > V_1(f^{(\infty)})\}$; estamos notando $V(f^{(\infty)}) = (v_1\ ,\ v_2\ ,\ v_3)'$ (vetor coluna) e $v_s(f^{(\infty)}) = v_s$, $s = 1,2,3$. Para $a = 1$, fica:

$$r(1,1) + (0,9)\ p(1,1)\ .\ (v_1,v_2,v_3)' =$$
$$= 8 + (0,9)\ (1/2,\ 1/4,\ 1/4).(91,3;\ 97,6;\ 90,0)' =$$
$$= 8 + (0,9)\ (91,3)/2 + (97,6)/4 + (90,0)/4 = 81,295 \leq 91,3 = v_1,$$

e portanto, $1 \notin G(1,f)$; para $a = 2$, fica:

$$r(1,2) + (0,9)\ p(1,2)\ .\ (v_1,v_2,v_3)' =$$
$$= 2,75 + (0,9)\ (1/16,\ 3/4,\ 3/16)\ .\ (91,3;\ 97,6;\ 90,0)' =$$
$$= 88,948 < 91,3 = v_1,$$

e portanto, $2 \notin G(1,f)$; como $A = \{1, 2\}$, segue-se que $G(1,f) = \emptyset$.

Por outro lado, $2 \in G(2,f) \neq \emptyset$ e $2 \in G(3,f) \neq \emptyset$, sendo verificações análogas às precedentes; mas a título de exemplificação extra, se tem:

$$r(2,2) + (0,9)\ p(2,2)\ .\ (v_1, v_2, v_3)' =$$
$$= 15 + (0,9)\ (1/16,\ 7/8,\ 1/16).(91,3;\ 97,6;\ 90,0)' =$$
$$= 102,056 > 97,6 = v_2,$$

isto é, de fato $2 \in G(2,f)$.

Ora, como $G(1,f) = \emptyset$, $2 \in G(2,f) \neq \emptyset$ e $2 \in G(3,f) \neq \emptyset$, é claro que não podemos garantir que $f^{(\infty)}$ seja uma estratégia ótima, e na verdade, não é; com efeito, se definirmos $g(2) = g(3) = 2$ e $g(1) = f(1) = 1$, a função g é tal que $g(s) \in G(s,f)$, se $G(s,f) \neq \emptyset$ (no caso, se $s = 2,3$), enquanto $g(s) = f(s)$, se $G(s,f) = \emptyset$ (no caso, se $s = 1$), e então podemos garantir que $g^{(\infty)} > f^{(\infty)}$, conforme o Teorema 3.5 (vide a segunda etapa na "rotina de Howard", descrita no início dêste parágrafo).

Retornando ao cálculo de v_1, v_2, v_3, agora para o caso de $g^{(\infty)}$, obtemos:

$$\begin{cases} v_1 = 119,4 \\ v_2 = 134,5 \\ v_3 = 121,9. \end{cases}$$

Como $2 \in G(1,g) \neq \emptyset$ e $G(2,g) = G(3,g) = \emptyset$ (agora, omitimos os cálculos), é claro que $g^{(\infty)}$ não é uma estratégia ótima, e além disto, se considerarmos $h \in F$, tal que $h(1) = 2$, $h(2) = 2$ e $h(3) = g(3) = 2$, teremos $h^{(\infty)} > g^{(\infty)}$.

Mais uma vez retornamos ao cálculo de v_1, v_2, v_3, agora para o caso de $h^{(\infty)}$, e obtemos:

$$\begin{cases} v_1 = 121,6 \\ v_2 = 153,3 \\ v_3 = 122,8 \end{cases}$$

Como $G(1,h) = G(2,h) = G(3,h) = \emptyset$ (novamente, omitimos os cálculos), segue-se que $h^{(\infty)}$ é uma estratégia estacionária ótima e o ganho esperado máximo (corrigido por $\beta = 0,9$) é dado por:

$$V(h^{(\infty)}) = \begin{pmatrix} 121,6 \\ 153,3 \\ 122,8 \end{pmatrix} .$$

Cabe apenas algumas palavras sôbre a interpretação do fator de desconto $\beta = 0,9$; no caso, $1 - \beta = 0,1$, pode se interpretar com a probabilidade do taxi ter de desistir da próxima corrida (por defeito mecânico, etc.); assim na n-ésima corrida, a probabilidade do mesmo realizá-la é dada por β^n, sendo pois razoável a consideração do desconto β da forma como o foi em (2.4).

§ 5. CASO DISCRETO. ESTRATÉGIAS ÓTIMAS PARA $\beta = 1$

Quando $\beta = 1$ (inexistência de desconto) o ganho total esperado, para uma dada estratégia, é tipicamente infinito (mais exatamente, não se poder garantir a convergência de $V(\Pi)$, pois, para certas estratégias se verifica $[V(\Pi)]_s = \infty$ para cada $s \in$, enquanto para outras o ganho $V(\Pi)$ é indeterminado).

<u>Exemplos</u>

1º) $[V(\Pi)]_s = \infty$ para cada $s \in \mathcal{A}$ quando, por exemplo, se tem $r = 1$.

2º) O ganho esperado $V(\Pi)$ é indeterminado no seguinte caso: $S = \{0, 1\}$; $\mathcal{A}$ tem um único elemento; q tem caráter determinístico, com $0 \longrightarrow 1$ e $1 \longrightarrow 0$ (*); a transição $0 \longrightarrow 1$ dá um ganho de Cr\$1,00 (1), e a transição $1 \longrightarrow 0$ custa Cr\$1,00, ou seja, dá um ganho de - Cr\$1,00 (-1). De fato, nêste exemplo, se partirmos do estado $s = 0$, teremos as transições:

$$0 \longrightarrow 1 \longrightarrow 0 \longrightarrow 1 \quad \ldots\ldots\ldots\ldots\ldots$$

com ganhos sucessivos $+1, -1, +1, -1, \ldots\ldots\ldots$, o que nos dá:

$$[V(\Pi)]_{s=0} = 1 - 1 + 1 - 1 + \ldots\ldots\ldots,$$

que é uma série alternada com soma indeterminada.

(*) ou seja,

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} q(0|0) & q(1|0) \\ q(0|1) & q(1|1) \end{pmatrix}$$

Há duas maneiras de abordar tais situações, decorrentes de não consideração de desconto: (i) tratar a situação $\beta = 1$ como caso limite da situação $\beta < 1$, ou seja, procurar estratégias ótimas para "β suficientemente próximo de 1"; (ii) analisar o ganho médio, por unidade de tempo, quando êste se torna infinito.

Iremos considerar apenas a primeira forma de abordagem, que é a utilizada por Blackwell, em (1).

Como β agora é uma variável, será necessário exibir a dependência de $V(\Pi)$ em relação a β, pelo que passaremos a notar $V_\beta(\Pi)$ em vez de $V(\Pi)$, e também passaremos a nos referir a estratégias β-ótimas; além disso, se $\Pi = f^{(\infty)}$, notaremos $V_\beta(f)$ por $V_\beta(f^{(\infty)})$. Representaremos por $U(\beta)$, para um dado $\beta < 1$, o ganho total esperado (corrigido por β para uma estratégia β-ótima.

Definição 5.1

a) Uma estratégia Π é ótima se for β-ótima para β suficientemente próximo de 1, isto é, quando $V_\beta(\Pi) = U(\beta)$ para β suficientemente próximo de $1^{(*)}$.

b) Uma estratégia Π é quase-ótima se $U(\beta) - V_\beta(\Pi) \longrightarrow 0$, quando $\beta \longrightarrow 1^{(**)}$.

Observemos que:

Se Π é uma estratégia ótima, então também é quase-ótima.

A seguir, nosso problema será a determinação de estratégias ótimas ou quase-ótimas; os resultados são reunidos no seguinte Teorema.

(*) ou seja, existe δ, $0 < \delta < 1$, tal que $V_\beta(\Pi) = U(\beta)$, $\delta < \beta < 1$.

(**) ou seja, dado $\varepsilon > 0$, existe δ, $0 < \delta < 1$, tal que $|U(\beta) - V_\beta(\Pi)| < \varepsilon$, $\delta < \beta < 1$.

<u>Teorema 5.2</u>

Seja $f \in F$; representemos por $Q^*(f)$ a matriz Q^* (conforme Lema 1.7) associada a $Q(f)$, isto é, $Q^*(f) = [Q(f)]^*$.

Então:

(a) $$V_\beta(f) = [x(f) / (1-\beta)] + y(f) + \varepsilon(\beta, f) ,$$

onde $x(f)$ é a única solução de

$$(I - Q(f))x = 0 \quad \underline{e} \quad Q^*(f)\, x = Q^*(f)\, r(f) ,$$

$y(f)$ é a única solução de

$$(I - Q(f))y = r(f) - x(f) \quad \underline{e} \quad Q^*(f)\, y = 0.$$

e onde $\varepsilon(\beta, f) \longrightarrow 0$, quando $\beta \longrightarrow 1$.

(b) Para cada $s \in \mathcal{S}$, seja $G(s, f)$ o conjunto de tôdas as ações a, tais que:

$$p(s,a)\, x(f) > x_s(f)$$

<u>ou</u>

$$\begin{cases} p(s,a)\, x(f) = x_s(f) \\ r(s,a) + p(s,a)\, y(f) > x_s(f) + y_s(f) , \end{cases}$$

onde $x_s(f)$ e $y_s(f)$ são as componentes de ordem s de $x(f)$ e $y(f)$; então, para cada $g \in F$ tal que: $g(s) \in G(s,f)$ para algum estado s, e $g(s) = f(s)$ quando $g(s) \notin G(s,f)$, tem-se:

$$g > f \quad (\text{isto é, } g^{(\infty)} > f^{(\infty)})$$

para todo $\beta < 1$ suficientemente próximo de 1.

(c) Para cada $s \in \mathcal{S}$ seja $E(s,f)$ o conjunto das ações a tais que:

$$\begin{cases} p(s,a)\, x(f) = x_s(f) \\ r(s,a) + p(s,a)\, y(f) = x_s(f) + y_s(f) ; \end{cases}$$

sempre se tem, observe-se $f(s) \in E(s,f)$. Então, se para cada estado s tivermos $G(s,f) = \emptyset$, e se $E(s,f)$ contém apenas o ponto $f(s)$, segue-se que f é ótima, ou mais precisamente, $f^{(\infty)} = (f, f, \ldots\ldots)$ é ótima.

(d) Se cada $s \in \mathcal{S}$ tivermos $G(s,f) = \emptyset$, e se $g(s) \in E(s,f)$ para todo s implicar $Q^*(g)\,Q^*(f) = Q^*(g)$, então f é quase-ótima.

(e) Para f_o tal que $G(s,f_o)$ é vazio, para todo s, temos $x(f_o) \geq x(g)$, para todo $g \in F$. Representemos por F^* o conjunto das g tais que $x(g) = x(f_o)$. Então existe $f^* \in F^*$, com $y(f^*) \geq y(g)$, para toda $g \in F^*$. As g quase-ótimas são exatamente para as quais $x(g) = x(f^*)$ e $y(g) = y(f^*)$.

Dem. de (a)

1º) Da mesma forma como se decidiu representar por $Q^*(f)$ a matriz Q^* associada a $Q(f)$, representaremos por $H(\beta, f)$ e $H(f)$, respectivamente, as matrizes $H(\beta)$ e H, associadas a $Q(f)$, bem como notaremos: $Q^n(f) = [Q(f)]^n$; conforme (c) do Lema 1.7, tem-se:

$$H(\beta, f) = \sum_{n=0}^{\infty} \beta^n \left(Q^n(f) - Q^*(f)\right). \text{ portanto, segue-se:}$$

$$V_\beta(f) = \left[\sum_{n=0}^{\infty} \beta^n\, Q^n(f)\right] r(f) = \left[\sum_{n=0}^{\infty} \beta^n\, Q^*(f) +\right.$$

$$\left. + \sum_{n=0}^{\infty} \beta^n (Q^n(f) - Q^*(f))\right] r(f) = \frac{Q^*(f)\, r(f)}{1-\beta} +$$

$$+ H(\beta, f)\, r(f) = \frac{Q^*(f)\, r(f)}{1-\beta} + H(f)\, r(f) +$$

$$+ \left[H(\beta, f) - H(f)\right] r(f);$$

fazendo,

$$x(f) = Q^*(f)\, r(f),$$

$$y(f) = H(f)\, r(f),$$

$$\varepsilon(\beta, f) = \left[H(\beta, f) - H(f)\right] r(f),$$

obtemos:

$$\underset{\beta}{V}(f) = \frac{x(f)}{1-\beta} + y(f) + \varepsilon(\beta, f) , \qquad (1)$$

Onde $\varepsilon(\beta, f) \longrightarrow 0$, quando $\beta \longrightarrow 1$, pois, $H(\beta, f) \longrightarrow$ $\longrightarrow H(f)$, quando $\beta \longrightarrow 1$.

2º) Observe-se que:

$$(I - Q(f))\, x(f) = (I -(f))(Q^*(f)\, r(f)) =$$

$$= [Q^*(f) - Q(f)\, Q^*(f)]\, r(f) = [Q^*(f) - Q^*(f)]\, r(f) = 0$$

e

$$Q^*(f)\, x(f) = Q^*(f)\, (Q^*(f)\, r(f)) = Q^*(f)\, r(f) \qquad ,$$

ou seja, $x(f) = Q^*(f)\, r(f)$ é a solução (única, conforme Lema 1.7 (e)), de:

$$(I - Q(f))\, x = 0 \qquad \underline{e} \qquad Q^*(f)x = Q^*(f)\, r(f).$$

Por outro lado, a verificação de $y(f) = H(f)\, r(f)$ ser solução de:

$$(I - Q(f)\, y = r(f) - x(f) \quad \underline{e} \quad Q^*(f)\, y = 0.$$

é decorrência imediata usando-se, mais uma vez o fato $x(f) = Q^*(f)\, r(f)$ das identidades $(I - Q(f))H(f) = I - Q^*(f)$ e $Q^*(f)\, H(f) = 0$ (verificadas em 1.7 (c)).

Resta a unicidade da solução $y = y(f)$, de fato, se existirem duas soluções y e y', isto é, se tivermos $(I - Q(f))\, y = (I - Q(f))y' = r(f) - x(f)$ e $Q^*(f)\, y = Q^*(f)\, y' = 0$, então,

$$(I - Q(f))(y - y') = 0 \quad e \quad Q^*(f)(y - y') = o;$$

a primeira igualdade fornece, sucessivamente, $Q(f)\,(y - y') = y - y'$; $Q^N(f)\,(y - y') = y - y'$, $n = 0,1,2,........$

(por indução finita); $(I + Q(f)(\, y - y') + \ldots\, Q^n(f)\,(y - y'))/(n+1)$

$= y - y'$, $n = 0, 1, 2, \ldots\ldots$; e finalmente, por passagem ao limite, $Q^*(f)\ (y - y') = y - y'$, que ao se comparar com $Q^*(f)\ (y - y') = 0$, permite concluir $y - y' = 0$, isto é, $y = y'$.

Dem. de (b)

1º) De acordo com:

$$V_\beta(g, f^{(\infty)}) = r(g) + \beta\, Q(g)\, V_\beta(f^{(\infty)}) = r(g) + \beta Q(g) \left[\frac{x(f)}{1-\beta} + \right.$$

$$\left. + y(f) + \varepsilon(\beta, f)\right] = r(g) + \frac{\beta}{1-\beta} Q(g)\, x(f) + \beta Q(g)\, y(f) +$$

$$+ \beta Q(g)\, \varepsilon(\beta, f) = r(g) + \left[\frac{Q(g)\, x(f)}{1-\beta} - Q(g)\, x(f)\right] + \left[Q(g)\, y(f) - \right.$$

$$\left. - (1-\beta)\, Q(g)\, y(f)\right] + \beta Q(g)\, \varepsilon(\beta, f),$$

e fazendo $\varepsilon_1(\beta, f, g) = -(1-\beta)\, Q(g)\, y(f) + Q(g)\, \varepsilon(\beta, f)$,

obtemos:

$$V_\beta(g, f^{(\infty)}) = \frac{Q(g)\, x(f)}{1-\beta} + r(g) - Q(g)\, x(f) +$$

$$+ Q(g)\, y(f) + \varepsilon_1(\beta, f, g) \qquad (2)$$

onde $\varepsilon_1(\beta, f, g) \longrightarrow 0$, quando $\beta \longrightarrow 1$ (*).

2º) Agora, passaremos a comparar $V_\beta(f)$ e $V_\beta(g, f^{(\infty)})$, através de suas expressões (1) e (2), respectivamente.

Inicialmente, provaremos que se $g(s) \in G(s,f)$ e $\beta < 1$ é tomado suficientemente próximo de 1, então a s-ésima coordenada de $V_\beta(g, f^{(\infty)})$ excede a coordenada de mesmo índice, de $V_\beta(f)$, ou seja, $[V_\beta(g, f^{(\infty)})]_s > [V_\beta(f)]_s$.

Ora, por definição de $G(s,f)$ (tomando $a = g(s)$), a hipótese $g(s) \in G(s,f)$ significa que:

(*) Com efeito, $\lim\limits_{\substack{\beta\to 1\\ \beta<1}} \varepsilon_1(\beta,f,g) = -\left[\lim\limits_{\beta\to 1}(1-\beta)\right] Q(g)\, y(f) +$

$+ (\lim\limits_{\beta\to 1}\beta)\, Q(g) \left[\lim\limits_{\beta\to 1} (\beta,f)\right] = 0.\ Q(g)\, y(f) + Q(g).\ o = 0.$

$$p(s,\ g(s))\ x(f) > x_s(f)$$

<u>ou</u>

$$\begin{cases} p(s,\ g(s))\ x(f) = x_s(f) \\ r(s,g(s)) + p(s,\ g(s))\ y(f) > x_s(f) + y_s(f); \end{cases}$$

mas, $p(s,\ g(s)) = [Q(g)]_s$ (vide observações no § 3, precedendo a definição 3.4), donde $p(s,\ g(s))\ x(f) = [Q(g)\ x(f)]_s$; analogamente, $p(s,\ g(s))\ y(f) = [Q(g)\ y(f)]_s$; além disso, já sabemos que $r(s,\ g(s)) = r_s(g)$ (isto é, a s-ésima coordenada do vetor $r(g)$); logo, o sistema precedente poderá ser escrito:

$$[Q(g)\ x(f)]_s > x_s(f)$$

<u>ou</u>

$$\begin{cases} [Q(g)\ x(f)]_s = x_s(f) \\ r_s(g) - [Q(g)\ x(f)]_s + [Q(g)\ y(f)]_s > y_s(f). \end{cases}$$

Assim, no caso de valer a alternativa $[Q(g)\ x(f)]_s > x_s(f)$, segue-se que a expressão $[Q(g)\ x(f)]_s\ /\ (1 - \beta)$ se tornará muito maior maior que $x_s(f)\ /\ (1 - \beta)$, para β suficientemente próximo de 1, donde (levando em conta que $\varepsilon(\beta,\ f) \longrightarrow 0$ e $\varepsilon_1(\beta, f,\ g) \longrightarrow 0$, quando $\beta \longrightarrow 1$) poderemos concluir que, independentemente dos valores que assumirem $y_s(f)$ e $r_s(g) - [Q(g)\ x(f)]_s + [Q(g)\ y(f)]_s$, será:

$$[V_\beta(g,\ f^{(\infty)})]_s > [V_\beta(f)]_s\ ;$$

a esta mesma conclusão chegaremos obviamente, no caso da outra alternativa: $[Q(g)\ x(f)]_s = x_s(f)$ <u>e</u> $r_s(g) - [Q(g)\ x(f)]_s + [Q(g)\ y(f)]_s > y_s(f)$.

3º) Por outro lado, se for $g(s) \notin G(s,f)$, temos $g(s) = f(s)$, e então, mostraremos que $[V_\beta(g,\ f^{(\infty)})]_s = [V_\beta(f)]_s$.

Observemos que a hipótese $g(s) = f(s)$ implica $r(s, g(s)) = r(s, f(s))$, ou seja:

$$\boxed{r_s(g) = r_s(f)} \quad ,$$

enquanto $[Q(f)\ V_\beta(f)]_s = \sum_{s'} q(s' | s, f(s))\ [V_\beta(f)]_{s'}$ $= \sum_{s'} q(s'\ s, | g(s))\ [V_\beta(f)]_{s'} = [Q(g)\ V_\beta(f)]_s$, em resumo:

$$\boxed{[Q(f)\ V\ (f)]_s = [Q(g)\ V\ (f)]_s}$$

tem-se, ainda, usando-se a Proposição 2.6, que:

$$\boxed{[V_\beta(f)]_s = r_s(f) + [Q(f)\ V_\beta(f)]_s} \quad ,$$

enquanto de $V_\beta(g, f^{(\infty)}) = r(g) + Q(g)\ V_\beta(f)$, obtemos:

$$\boxed{[V_\beta(g, f^{(\infty)}))]_s = r_s(g) + \beta[Q(g)\ V_\beta(f)]_s} \quad ;$$

das quatro últimas igualdades postas em destaque, concluimos, então, que:

$$[V_\beta(g, f^{(\infty)})]_s = [V_\beta(f)]_s \ .$$

4º) Em suma, as condições impostas no enunciado de (b). conduzem a:

$$[V_\beta(g, f^{(\infty)})]_s \geq [V_\beta(f)]_s \ ,$$

para cada $s \in \mathcal{S}$ e β suficientemente próximo de 1, sendo ainda certo que existe pelo menos um "índice" s para o qual vale a desigualdade restrita ($>$), e assim,

$$(g, f^{(\infty)}) > f^{(\infty)}$$

que então nos permite afirmar (aplicando o Teorema 3.3) que:

$$g^{(\infty)} > g^{(\infty)} \text{ , isto é,}$$

$$g > f$$

(para β suficientemente próximo de 1).

Dem. de (c)

1º) Inicialmente, mostraremos que $f(s) \in E(s,f)$, para todo $s \in$ e toda $f \in F$.

De fato, por um lado, temos:

$p(s, f(s))\ x(f) = [Q(f)\ x(f)]_s = [Q(f)\ Q^*(f)\ r(f)]_s =$

$= [Q^*(f)\ r(f)]_s = [x(f)]_s = x_s(f)$

Por outro lado, como $(I - Q(f))\ H(f) = I - Q^*(f)$, segue-se

$p(s, f(s))\ y(f) = [Q(f)\ y(f)]_s = [Q(f)\ H(f)\ r(f)]_s =$

$= [H(f)\ f(f) - r(f) + Q^*(f)\ r(f)]_s = [y(f) - r(f) + x(f)]_s =$

$= y_s(f) - r_s(f) + x_s(f)$, e uma vez que $r(s, f(s)) = r_s(f)$, obtemos:

$$r(s, f(s)) + p(s, f(s))\ y(f) = y_s(f) + x_s(f).$$

Com isso, concluimos a verificação de $f(s) \in E(s, f)$.

2º) Agora, devemos provar que se $G(s,f) = \emptyset$ e $E(s,f) = \{f(s)\}$ para cada $s \in \mathcal{S}$, então $f^{(\infty)}$ é ótima (ou seja, é β-ótima para todo $\beta < 1$ suficientemente próxima de 1); de acôrdo com o Teorema 3.2, bastará provar que $f^{(\infty)} \geq (g,f^{(\infty)})$, $\forall\ g \in F$, isto é, que:

$$[V_\beta(f)]_s \geq [V_\beta(g,f^{(\infty)})]\ , \forall\ g \in F,\ \forall\ s \in \mathcal{S},$$

e todo β suficientemente próximo de 1.

3º) Ora, dados quaisquer $g \in F$ e $s \in \mathcal{S}$, como por hipótese $E(s,f) = \{f(s)\}$, só há duas alternativas: $g(s) = f(s)$ ou $g(s) \notin E(s,f)$; mas $g(s) = f(s)$, já conduz como se viu na prova de (b), a:

$$[V_\beta(g,f^{(\infty)})]_s = [V_\beta(f)]_s$$

que satisfaz a $[V_\beta(f)]_s \geq [V_\beta(g,f^{(\infty)}]_s$.

Assim, resta-nos considerar a alternativa $g(s) \notin E(s,f)$, juntamente com a hipótese $G(s,f) = \emptyset$ (caso em que $g \neq f$).

4º) Por definição, $g(s) \notin E(s,f)$ significa:

$$[Q(g)\ x(f)]_s \neq x_s(f)$$

ou

$$r_s(g) - [Q(g)\ x(f)]_s + [Q(g)\ y(f)]_s \neq y_s(f);$$

por outro lado $G(s,f) = \emptyset$, em particular, implica que $g(s) \notin G(s,f)$, e também por definição(*), se tem:

$$[Q(g)\ x(f)]_s \leq x_s(f)$$

e

$$\left\{\begin{array}{l} [Q(g)\ x(f)]_s \neq x_s(f) \\ \qquad\qquad \text{ou} \\ r_s(g) + [Q(g)\ y(f)]_s \leq x_s(f) + y_s(f), \end{array}\right. \tag{3}$$

que é o mesmo de:

$$[Q(g)\ x(f)]_s < x_s(f)$$

<u>ou</u>

$$\left\{\begin{array}{l} [Q(g)\ x(f)]_s \leq x_s(f) \\ r_s(g) + [Q(g)\ y(f)]_s \leq x_s(f) + y_s(f), \end{array}\right.$$

que por sua vez implica(**)

(*) negação de $g(s) \in G(s,f)$.

(**) em virtude de: $(A < B)$ ou $\{(A \leq B)$ e $(C \leq D)\} \longrightarrow$
$\longrightarrow (A < D)\ \vee \{(A = B) \wedge (C \leq D)\}$.

$$[Q(g)\ x(f)]_s < x_s(f)$$

<u>ou</u>

$$\begin{cases} [Q(g)\ x(f)]_s = x_s(f) \\ r_s(g) - [Q(g)\ x(f)]_s + [Q(g)\ y(f)]_s \leq y_s(f). \end{cases} \qquad (4)$$

Portanto, temos de considerar os sistemas (3) e (4), simultaneamente. Observe-se que em (4), a condição $[Q(g)\ x(f)]_s < x_s(f)$ já implicaria sôzinha em $[V_\beta(g,f^{(\infty)})]_s < [V_\beta(f)]_s$, também satisfazendo a $[V_\beta(f)]_s \geq [V_\beta(g,f^{(\infty)})]_s$, e portanto, resta-nos considerar, simultaneamente com (3), o sistema:

$$\begin{cases} [Q(g)\ x(f)]_s = x_s(f) \\ r_s(g) - [Q(g)\ x(f)]_s + [Q(g)\ y(f)]_s \leq y_s(f); \end{cases} \qquad (4')$$

ora, (3) e (4') implicam

$$\begin{cases} [Q(g)\ x(f)]_s = x_s(f) \\ r_s(g) - [Q(g)\ x(f)]_s + [Q(g)\ y(f)]_s < y_s(f), \end{cases}$$

donde se conclui da comparação de (1) e (2), como em (b), que

$$[V_\beta(g,f^{(\infty)})]_s < [V_\beta(f)]_s$$

Portanto, consideradas todas as possíveis alternativas, tem-se sempre:

$$[V_\beta(g),f^{(\infty)})]_s \leq [V_\beta(f)]_s,\ \forall g,\ \forall\ s \in \mathcal{S},$$

isto é, $f^{(\infty)} \geq (g,f^{(\infty)})$, como requerido.

Para continuar a demonstração do Teorema 5.3, teremos necessidade de enunciar e provar o seguinte lema:

<u>Lema 5.4</u>

Se f, $g \in F$ são tais que $g(s) \in E(s,f)$, $\forall\ s \in$, então $x(f) = x(g)$. Se além disso, $Q^*(g)Q^*(f) = Q^*(g)$, então $y(f) = y(g)$.

Dem. do Lema

1º) $g(s) \in E(s,f)$, $\forall s \in \mathcal{S}$, equivale a:

$$\begin{cases} p(s,g(s))\ x(f) = x_s(f) \\ r(s,g(s)) + p(s,g(s))\ y(f) = x_s(f) + y_s(f), \ \forall s \in \mathcal{S}, \end{cases}$$

ou seja:

$$\begin{cases} Q(g)\ x(f) = x(f) \\ r(g) + Q(g)\ y(f) = x(f) + y(f) \end{cases}$$

premultiplicando a segunda igualdade por $Q^*(g)$, obtemos:

$$Q^*(g)\ r(g) + Q^*(g)\ Q(g)\ y(f) = Q^*(g)\ x(f) + Q^*(g)\ y(f),$$

e como $Q^*(g)\ Q(g) = Q^*(g)$, segue-se:

$$Q^*(g)\ r(g) = Q^*(g)\ y(f) \ ;$$

ficamos então com:

$$\begin{cases} Q(g)\ x(f) = x(f) \\ Q^*(g)\ r(g) = Q^*(g)\ x(f) \end{cases}$$

mostrando-nos que $x(f)$ é solução do sistema:

$$\begin{cases} Q(g)\ x = x \\ Q^*(g)\ x = Q^*(g)\ c \\ c = r(g) \end{cases}$$

Ora, conforme (e) do Lema 1.7, êste sistema possui a solução única $x(g) = Q^*(g)\ r(g)$, e portanto, $x(f) = x(g)$.

2º) Consideremos, em seguida, a condição suplementar $Q^*(g)Q^*(f) = Q^*(g)$; como $y(f) = H(f)\ r(f)$ e $Q^*(f)\ H(f) = 0$, obtemos:

$$Q^*(f)\ y(f) = Q^*(f)\ (H(f)\ r(f)) = (Q^*(f)\ H(f))\ r(f) = 0$$

donde:

$$Q^*(g)\ y(f) = (Q^*(g)\ Q^*(f))\ y(f) = Q^*(g)(Q^*(f)\ y(f)) = 0;$$

assim, $y(f)$ é solução do sistema:

$$\begin{cases} r(g) + Q(g)\, y = x(f) + y \\ Q^*(f)\, y = 0 \end{cases}$$

Por outro lado, uma vez que $g(s) \in E(s,g)$, obtemos: $r(g) + Q(g)\, y(g) = x(g) + y(g)$, e como já se provou que $x(g) = x(f)$, fica:

$$r(g) + Q(g)\, y(g) = x(f) + y(g);$$

ainda mais, como $y(g) = H(g)\, r(g)$ e $Q^*(g)\, H(g) = 0$, teremos:

$$Q^*(g)\, y(g) = Q^*(g)\, H(g)\, r(g) = 0;$$

logo, $y(g)$ também é a solução do sistema precedente. E como êste sistema tem solução única (pois se y e y' forem soluções, partindo de $Q^*(g)(y - y')$ e $Q(g)(y - y') = y - y'$, chegaremos a $y = y'$ (*)) concluimos que $y(g) = y(f)$.

Dem. de (d)

1º) Observemos, inicialmente, que é possível escolher β próximo de 1, de sorte que para qualquer par $(g,h) \in F^2$, tal que $V_\beta(g,h^{(\infty)}) \geq V_\beta(h)$, decorre $g(s) \in G(s,h) \cup E(s,h)$, $\forall\, s \in \mathcal{S}$.

Com efeito, dado $(g,h) \in F^2$, a condição: "existe β suficientemente próximo de 1, tal que $V_\beta(g,h^{(\infty)}) \geq V_\beta(h)$, implica $g(s) \in G(s,h) \cup E(s,h)$, $\forall\, s \in \mathcal{S}$", é equivalente à condição "$g(c) \notin G(s,h) \cup E(s,h)$, $\forall\, s \in \mathcal{S}$, implica $V_\beta(g,h^{(\infty)}) < V_\beta(h)$, para todo β suficientemente próximo de 1"; mas foi esta última condição provada em (c) (com g e f, no lugar de g e h).

Note-se que, acima, tomar β suficientemente próximo de 1, é escolher um $\delta(g,h) > 0$ e trabalhar somente com os β's tais que $1 - \delta(g,h) < \beta < 1$; como os conjuntos $\mathcal{S}$ e A são finitos, também o é o conjunto F das regras de ação, donde existe $\delta = \inf \delta(g,h)$ para $(g,h) \in F$, e assim, trabalhando com os β's tais que $1-\delta < \beta < 1$, concluimos que em geral vale a propriedade enunciada, isto é, exis

(*) Trata-se da mesma técnica utilizada no final da prova de (a).

te β suficientemente próximo de 1, tal que $V_\beta(g,h^{(\infty)}) \geq V_\beta(h) \longrightarrow$ $\longrightarrow g(s) \in G(s,h) \cup E(s,h)$, $\forall\, s \in \mathcal{S}$.

2º) Escolhamos um β tal que $1-\delta < \beta < 1$; há duas alternativas possíveis. A primeira deslas consiste em f ser β-ótima, e nêste caso $U(\beta) = V_\beta(f)$ ou seja,

$$U(\beta) - V_\beta(f) = 0 \qquad (5)$$

A segunda alternativa consiste em f não ser β-ótima, mas nêste caso, conforme o Teorema 3.5, existe uma sequência $f_0=f, f_1, f_2 \ldots f_k$ (onde $k = k(\beta)$, isto é, depende de β), tal que $f < f_1 < \ldots < f_k$, ou seja, $U_\beta(f) < V_\beta(f_1) < \ldots < V_\beta(f_k)$ e $f_k^{(\infty)}$ é uma estratégia β-ótima.

Como $V_\beta(g,h^{(\infty)}) \geq V_\beta(h) \longleftrightarrow V_\beta(g) \geq V_\beta(h)$ (*), segue-se:

$$V_\beta(f_j, f_{j-1}^{(\infty)}) \geq V_\beta(f_{j-1}), \quad j = 1, \ldots\ldots, k,$$

e em virtude de ser $1-\delta < \beta < 1$, $\delta = \inf \delta(g,h,s)$, $(g,h,s) \in F^2 x$, e ainda em virtude do resultado verificado no item precedente, segue-se que:

$$f_j(s) \in G(s,f_{j-1}(s)) \cup E(s,f_{j-1}(s)), \forall \in \mathcal{S}, \quad j = 1, \ldots, k.$$

<u>Ora, por hipótese</u>, $G(s,f) = \emptyset$, $\forall\, s \in \mathcal{S}$, donde $f_1(s) \in G(s,f) \cup E(s,f) = E(s,f)$, $\forall\, s \in \mathcal{S}$, e portanto (conforme o enunciado de 5.3 (d)) se tem $Q^*(f_1)\, Q^*(f) = Q^*(f_1)$; e de acordo com o Lema 5.4, se pode concluir que $x(f_1) = x(f)$ e $y(f_1) = y(f)$.

Por outro lado, $G(s,f_1)$ depende só de $x(f_1)$, e $E(s,f_1)$ depende só de $x(f_1)$ e $y(f_1)$; uma vez que $x(f_1) = x(f)$ e $y(f_1) = y(f)$, segue-se $G(s,f_1) = G(s,f) = \emptyset$ e $E(s,f_1) = E(s,f)$, donde $f_2(s) \in$

(*) $V_\beta(h) = \Sigma \beta^n Q^n(h) r(h) = [I - Q(h)]^{-1} r(h)$ e $B_\beta(g,h^{(\infty)}) = r(g) + \beta Q(g) V_\beta(h)$, donde $V_\beta(g,h^{(\infty)}) \geq V_\beta(h) \longleftrightarrow r(g) + \beta Q(g)\, V_\beta(h) - V_\beta(h) = r(g) + [\beta Q(g) - I][I - \beta Q(h)]^{-1} r(h) \geq 0 \longleftrightarrow [I-\beta Q(g)]^{-1} \times r(g) = [I - \beta Q(h)]^{-1} r(h) \longleftrightarrow V_\beta(g) \geq V_\beta(h)$.

e $G(s,f_1) \cup E(s,f_1) = E(s,f)$, $\forall s \in$, e analogamente, se pode concluir que $x(f_2) = x(f)$ e $y(f_2) = y(f)$.

Procedendo desta forma, sucessivamente, obtemos:

$$x(f) = x(f_1) = \ldots\ldots\ldots\ldots = x(f_k)$$

e $$y(f) = y(f_1) = \ldots\ldots\ldots\ldots = y(f_k);$$

e como $f_k^{(\infty)}$ é uma estratégia β-ótima, tem-se $U(\beta) = V_\beta(f_k)$, donde se segue:

$$U(\beta) = \frac{x(f_k)}{1-\beta} + y(f) + \varepsilon(\beta, f_k) =$$

$$= \frac{x(f)}{1-\beta} + y(f) + \varepsilon(\beta, f_k),$$

e uma vez que $V_\beta(f) = x(f) / (1-\beta) + y(f) + \varepsilon(\beta, f)$, obtemos:

$$U(\beta) - V_\beta(f) = \varepsilon(\beta, f_k) - \varepsilon(\beta, f) \qquad (6)$$

(note-se que f_k depende de β, isto é, $f_k = f_{k(\beta)}$).

Porém, como qualquer que seja a regra de ação g, tem-se: $\varepsilon(\beta, g) \longrightarrow 0$, quando $\beta \longrightarrow 1$, e além disto tais $g \in F$ são em número finito, segue-se que:

$$U(\beta) - V_\beta(f) \longrightarrow 0, \text{ quando } \beta \longrightarrow 1 \qquad (7)$$

Logo, pois, (5) ou (7), f é "quase-ótima".

<u>Dem. de (e)</u>

1º) Se for $G(s,f_o) \neq \emptyset$, $\forall s \in \mathcal{S}$, tem-se para qualquer $g \in F$:

$$[Q(g)\ x(f_o)]_s < x_s(f_o)$$

ou

$$[Q(g)\ x(f_o)]_s = x_s(f_o)$$

$$r_s(g) - [Q(g)\ x(f_o)]_s + [Q(g)\ y(f_o)]_s \leq y_s(f_o),$$

$$\forall s \in \mathcal{S};$$

por outro lado,

$$[V_\beta(g,f_o^{(\infty)}) - V_\beta(f_o)]_s = \frac{[Q(g)x(f_o)]_s - x_s(f_o)}{1-\beta} +$$

$$+ \{r_s(g) - [Q(g)x(f_o)]_s + [Q(g)y(f_o)]_s - y_s(f_o)\} +$$

$$+ [\varepsilon_1(\beta, f_o, g) - \varepsilon(\beta, f_o)]_s, \ \forall\, s \in \mathcal{S}.$$

Ora, se fôr o caso da primeira aleternativa $[Q(g)x(f_o)]_s <$ $< x_s(f)$, teremos:

$$\frac{[Q(g)\ x(f_o)]_s - x_s(f_o)}{1-\beta} + \{r_s(g) - \ldots\ldots\ldots\} < 0,$$

para β suficientemente próximo de 1, donde:

$[V(g,f_o^{(\infty)}) - V(f_o)]_s \leq [\varepsilon_1(\beta, f_o, g) - \varepsilon(\beta, f_o)]_s$, para $1-\delta < \beta < 1$, $\delta = \delta(s)$;

se fôr o caso de se ter a outra alternativa, isto é, $[Q(g)x(f_o)]_s = x_s(f)$ e $r_s(f) + [Q(g)y(f_o)] - [Q(g)y(f_o)]_s \leq$ $\leq y_s(f_o)$, obtemos o mesmo resultado acima.

Portanto, se tomarmos $\delta = \inf\limits_s \delta(s)$, teremos:

$$V_\beta(g,f_o^{(\infty)}) - V_\beta(f_o) \leq \varepsilon_1(\beta, f_o, (g) - \varepsilon(\beta, f_o),$$

para $1-\delta < \beta < 1$ (isto é, para βsuficientemente próximo de 1).

2º) Sendo $\tau(\beta)$ a máxima coordenada, em valor absoluto, de $\varepsilon_1(\beta,f_o,g) -$ $- \varepsilon(\beta,f_o)$, teremos $\varepsilon_1(\beta,f_o,g) - \varepsilon(\beta,f_o) \leq \tau(\beta)u$, onde u é o vetor coluna cujas coordenadas são todas iguais a 1(*); logo:

$$V_\beta(g,f_o^{(\infty)}) \leq V_\beta(f_o) + \tau(\beta)\ u,$$

para β suficientemente próximo de 1.

Por outro lado, notando $L_\beta(g) = L_\beta$ (vide Definição 2.7), esta última desigualdade passa a se escrever:

(*) Se α é um vetor coluna, com $\alpha^t = (\alpha_1, \ldots, \alpha_S)$, é claro que $\alpha \leq \tau u$, se $\tau = \max\{|\alpha_1|, \ldots, |\alpha_S|\}$ e $u^t = (1, \ldots, 1)$.

$$L_\beta V_\beta(f_o) \leq V_\beta(f_o) + \tau(\beta)u,$$

e por indução finita, obtemos(*):

$L_\beta^n V_\beta(f_o) \leq V_\beta(f_o) + (1 + \beta + \ldots + \beta^{n-1})\ \tau(\beta)u$, para β suficientemente próximo de 1, e $n = 1, 2, \ldots\ldots$.

Além disso, como $1 + \beta + \ldots + \beta^{n-1} \leq 1 + \beta + \ldots\ldots = 1/(1-\beta)$, segue-se:

$$L_\beta^n\ V_\beta(f_o) \leq V_\beta(f_o) + \frac{\tau(\beta)u}{1 - \beta}\ ,$$

para β suficientemente próximo de 1, e $n = 1, 2, \ldots.$; e uma vez que $\lim\limits_{n\to\infty} L_\beta^n\ V_\beta(f_o) = V_\beta(g)$ (conforme Proposição 3.1), segue-se ainda:

$$V_\beta\ (g) - V_\beta\ (f_o) \leq \tau(\beta)\ u\ /\ (1 - \beta), \tag{8}$$

para β suficientemente próximo de 1.

Por outro lado,

$$V_\beta(g) - V_\beta(f_o) = \frac{x(g) - x(f_o)}{1 - \beta} + y(g) - y(f_o) + \varepsilon(\beta,g) - \varepsilon(\beta,f_o)$$

$$\forall\ \beta < 1. \tag{9}$$

Ora, (8) e (9) implicam em $x(g) \leq x(f_o)$, para todo $g \in F$(**).

(*) i) Vale, para n=1 (já verificado); ii) Supomos que vale para n arbitrário; segue-se (escrevendo L e τ, em lugar de L_β e $\tau(\beta)$, para aliviar a notação; idem $V(\ )$ em lugar de $V\ (\ )$):

$L^{n+1}V(f_o) = L\ (L^n\ V(f_o)) \leq L\left[V(f_o) + (1 + \beta + \ldots + \beta^{n-1})\ \tau u\right] =$
$= r(g) + \beta Q(g)\ \left[V(f_o) + (1 + \beta + \ldots + \beta^{n-1})\tau u\right] = \left[\ r(g)\ +\right.$
$\left. + \beta Q(g)\ V(f_o)\right] + \left[Q(g)(\beta + \beta^2 + \ldots + \beta^{n-1})\ \tau u\right] = L\ V(f_o)\ +$
$+ \left[Q(g)(\beta + \beta^2 + \ldots + \beta^{n-1})\ \tau u\right] \leq \left[L\ V(f_o) + (\beta + \beta^2 + \ldots + \beta^{n-1})\tau u\right] \leq$
$\leq L\ V(f_o) + (1 \pm \beta + \ldots + \beta^n)\ \tau u$, isto é, vale para $n + 1$.

(**) Com efeito, se fosse $x(g) > x(f_o)$, teríamos $x_s(g) > x_s(f_o)$ para algum s, donde (conforme técnica já nossa conhecida) obteríamos que $[V_\beta(g) - V_\beta(f_o)]_s$ seria arbitrariamente grande, para β suficientemente próximo de 1, que é uma contradição com respeito a:

$$[V_\beta(g) - V_\beta(f_o)]_s \leq \tau(\beta)\ /\ (1 - \beta).$$

3º) Em seguida, tomaremos uma $f^* \in F^*$ que seja β-ótima para um conjunto de β's, tendo 1 como ponto limite.

Uma tal f^* existe, de fato. Com efeito, consideremos uma sequência (γ_n) de números, tal que $\gamma_1 < \gamma_2 < \dots < \gamma_n < \dots < 1$ e $\gamma_n \longrightarrow 1$, quando $n \longrightarrow \infty$. Para cada γ_n existe uma $f_n \in F$, tal que $V_{\gamma_n}(f) = V_{\gamma_n}(f)$, $\forall f \in F$, ou seja, tal que f_n seja γ_n-ótima. Ora, F é um conjunto finito, ao passo que os índices γ_n constituem um conjunto infinito, e portanto, pelo menos uma das funções f_n deve se repetir infinitas vezes; assim, existe uma subsucessão $\gamma_{\alpha_1} = \beta_1$, $\gamma_{\alpha_2} = \beta_2$, ..., $\gamma_{\alpha_n} = \beta_n$, ..., tal que $f_{\beta_1} = f_{\beta_2} = \dots f_{\beta_n} = \dots$; chamando de f^* esta função, tem-se que f^* é β_n-ótima, com $\beta_n \longrightarrow 1$, quando $n \longrightarrow \infty$.

Para $g = f^*$, segue-se:

$$V_\beta(f^*) - V_\beta(f_o) = \frac{x(f^*) - x(f_o)}{1-\beta} + y(f^*) - y(f_o) +$$

$$+ \varepsilon(\beta, f^*) - \varepsilon(\beta, f_o) \ , \ \forall \beta < 1;$$

e como f^* é β-ótima, para β assumindo os valores $\beta_1, \beta_2, \dots\dots$, tem-se também:

$$V_{\beta_n}(f^*) - V_{\beta_n}(f_o) \geq 0, \quad n = 1, 2, \dots\dots\dots;$$

nestas circunstâncias, uma vez que $\varepsilon(\beta_n, f^*) - \varepsilon(\beta_n, f_o) \longrightarrow 0$, quando $\beta_n \longrightarrow 1$, é claro que não se poderia ter $x_s(f^*) - x_s(f_o) < 0$, para nenhum $s \in \mathcal{S}$, ou seja, temos $x_s(f^*) - x_s(f_o) \geq 0$, $s \in$, isto é,

$$x(f^*) - x(f_o) \geq 0, \qquad \text{e}$$

combinando êste resultado com $x(f^*) - x(f_o) \leq 0$ (já provado), concluimos que:

$$x(f^*) - x(f_o) = 0,$$

ou seja,

$$f^* \in F^*.$$

4º) Agora, para qualquer $g \in F^*$, temos:

$$V_\beta(f^*) - V_\beta(g) = \frac{x(f^*) - x(g)}{1 - \beta} + y(f^*) - y(g) + \varepsilon(\beta, f^*) - \varepsilon(\beta, g);$$

como $x(f^*) = x(f_o)$ e $x(g) = x(f_o)$, segue-se:

$$V_\beta(f^*) - V_\beta(g) = y(f^*) - y(g) + \varepsilon(\beta, f^*) - \varepsilon(\beta, g),$$

donde:

$$V_{\beta_n}(f^*) - V_{\beta_n}(g) \longrightarrow y(f^*) - y(g),$$

quando $n \longrightarrow \infty$;

e uma vez que f^* é β_n-ótima, para $n = 1, 2, \ldots$, tem-se ainda:

$$V_{\beta_n}(f^*) - V_{\beta_n}(g) \geq 0, \quad n = 1, 2, \ldots,$$

que combinando com a convergência indicada linhas atrás, implica:

$$y(f^*) \geq y(g), \quad \forall g \in F^*$$

Finalmente, é óbvio que as g "quase-ótimas" são aquelas para as quais $x(g) = x(f^*)$ e $y(g) = y(f^*)$.

<u>Observação</u>: Êste Teorema não descreve um algorítmo que conduza a estratégias ótimas ou mesmo quase-ótimas, comparável em simplicidade ao algorítimo descrito pelo Teorema 3.5, para a situação $\beta < 1$.

Com efeito, o algorítimo é simples até se obter uma f tal que $G(s,f) = \emptyset$, para cada $s \in \mathcal{S}$; a esta altura, se $E(s,f) = \{f(s)\}$, para cada $s \in \mathcal{S}$, então f é ótima (conforme 5.3.(c)); mas se tal não ocorre , apenas podemos garantir que $x(f) \geq x(g)$, para cada $g \in F$ (conforme 5.3 $\underline{d}$), que por sua vez só garante maximizar o "ganho médio"(*)

Em um caso, a verificação descrita em 5.3 $\underline{d}$ é simples: aquêle em que existe um estado terminal s^*, que é certo ser eventualmente alcançado, quaisquer que sejam o estado inicial e a estratégia usada, nêle permanecendo o sistema, indefinidamente, uma vez alcançado(**). Neste caso para cada g, $Q^*(g)$ é a matriz em que todas as linhas são iguais ao vetor unitário $(0, \ldots 0, \underset{s^*}{1}, 0, \ldots, 0)$, isto é, $Q^*_{ss'}(g) = 0$ ou 1, conforme $s' \neq s^*$ ou $s' = s^*$, respectivamente; em consequência, $Q^*(g)Q^*(f) = Q^*(g)$, e portanto, f é "quase-ótima".

Em geral, a verificação das condições exigidas em 5.3 $\underline{d}$ é tediosa; se não valem as condições, teremos de determinar o conjunto F^* e calcular $y(g)$ para cada $g \in F^*$; em seguida selecionaremos uma g tal que $y(g)$ seja maximal (isto é, escolheremos uma $g = f^*$, tal que $y(f^*) \geq y(g)$, $\forall g \in F$.

(*) Se $\Pi = (f_1, f_2, \ldots\ldots)$ é uma estratégia qualquer, o "ganho médio até o instante N", a ela devido, se define de uma maneira natural como:

$$V_N(\Pi) = \frac{\sum_{n=0}^{N} Q_n(\Pi)\, r(f_{n+1})}{N+1} \;;$$

no caso em que Π é uma estratégia estacionária, isto é $\Pi = (f^{(\infty)})$,

tem-se: $$V_N(f) = \frac{\sum_{n=0}^{N} Q^n(f)\, r(f)}{N+1}\,,$$ donde:

$$\lim_{N\to\infty} V_N(f) = \left| \lim_{N\to\infty} \frac{\sum_{n=0}^{N} Q^n(f)}{N+1} \right| r(f) = Q^*(f)\, r(f) =$$

$= z(f)$; isto é, o "ganho médio", por unidade de tempo, quando êste se torna infinito, é dado por $z(f)$.

(**) isto é, $q(s^*|s^*,a) = 1$, $a \in A$.

Teorema 5.5

Há uma estratégia ótima que é estacionária.

Dem.

Já sabemos que $V(f) = [I - \beta Q(f)]^{-1} r(f)$, donde se vê que $V_\beta(f)(s)$ é uma função racional de β.

Suponhamos, então, que f^* é ótima para uma coleção de β's, cujo limite é 1, isto é, existe $\beta_1 < \beta_2 < \ldots < \beta_n < \ldots < 1$, tal que $\beta_n \longrightarrow 1$, quando $n \longrightarrow \infty$, e $V(f^*) \geq V_{\beta_n}(g)$, $\forall\, g \in F$.

Como tôdas as coordenadas de $V_\beta(f^*)$ são funções racionais de β, segue-se que:

$$V_\beta(f^*) \geq V_\beta(g),$$

e para cada β suficientemente próximo de 1; $\forall\, g \in F$, isto é, é ótima.

C.Q.D.

Os exemplos seguintes são instrutivos:

Exemplo 1:

Uma f que satisfaz às hipóteses de (d) do Teorema 5.3, mas que não é ótima.

Tem-se $\mathcal{S} = \{1, 2\}$ e $A = \{1, 2\}$. No estado 1 a ação 1 dá um ganho igual a 1, e o sistema se move para o estado 2 com probabilidade de 1/2; enquanto a ação 2 dá um ganho igual a 2, e o sistema passa obrigatoriamente para o estado 2 (com probabilidade igual a 1). No estado 2, cada ação dá um ganho nulo e o sistema permanece obrigatoriamente no mesmo estado 2.

Efetivamente, só há duas regras de ação a considerar, f e g, tai que: $f(1) = 1$, $g(1) = 2$ e $f(2) = g(2) = 2$ (com efeito, se o sistema se encontra no estado 2, é indiferente tomar $f(2) = 1$ ou $f(2) = 2$, bem como $g(2) = 1$ ou $g(2) = 2$, pois o ganho é sempre nulo, e além disso, o sistema permanece obrigatoriamente no mesmo estado).

Para essas regras de ação f e g, acima consideradas, tem-se:

$$Q(f) = \begin{bmatrix} 1/2 & 1/2 \\ 0 & 1 \end{bmatrix} , \qquad r(f) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} ;$$

$$Q(g) = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} , \qquad r(g) = \begin{bmatrix} 2 \\ 0 \end{bmatrix} ;$$

logo calculamos(*)

$$V_\beta(f) = \begin{bmatrix} 2/(2-\beta) \\ 0 \end{bmatrix}$$

$$V_\beta(g) = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

Ora, como $U(\beta) = 2$, vemos que $f^{(\infty)}$ é "quase-ótima"(**), mas $f^{(\infty)}$ não é ótima, pois, $V_\beta(g) > V_\beta(f)$, para todo $\beta < 1$.

OBS.: A verificação de que f satisfaz às condições do Teorema 5.3 (d) são imediatas.

<u>Exemplo 2</u>

<u>Uma f tal que $G(s,f) = \emptyset$, para todo $s \in \mathcal{S}$, mas que não é "qua-</u>

(*) Inicialmente com o "estado 1", teremos $[V(f)]_s = 1 + \frac{1}{2}\beta + \frac{1}{4}\beta^2 + \ldots$ ...... $= 2/(2 - \beta)$ iniciando com o "estado 2", o sistema permanecerá indefinidamente nêste mesmo estado, com ganhos sucessivos nulos, isto é $[V(f)]_{s=2} = 0$.

(**) pois: $\lim\limits_{\beta \to 1} [U(\beta) - V_\beta(f)] = 0$.

se-ótima".

Ainda $\mathcal{S} = \{1, 2\}$ e $A = \{1, 2\}$. No estado 1, a ação 1 dá um ganho igual a 3, ficando o sistema no mesmo estado com probabilidade 1/2; quanto à ação 2, dá um ganho igual a 6, e o sistema obrigatoriamente se move para o estado 2. No estado 2, cada ação dá uma perda igual a 3 (isto é, um ganho igual a -3) e o sistema permanece no mesmo estado 2 com probabilidade 1/2.

Nêste cado, ainda basta considerar duas regras de ação f e g, tais que: $f(1) = 1$, $g(1) = 2$ e $f(2) = g(2) = 1$ (pois se o sistema se encontra no estado 2, é indiferente escolher a ação 1 ou a).

Tem-se, então:

$$Q(f) = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} , \qquad r(f) = \begin{pmatrix} 3 \\ -3 \end{pmatrix} ;$$

$$Q(g) = \begin{pmatrix} 0 & 1 \\ 1/2 & 1/2 \end{pmatrix} , \qquad r(g) = \begin{pmatrix} 6 \\ -3 \end{pmatrix} .$$

Calcularemos, em seguida, $x(f)$, $x(g)$, $y(f)$, $y(g)$.

Como $Q^n(f) = Q(f)$, $n = 1, 2, \ldots.$ (basta verificar $Q^2(f) = Q(f)$), obtemos $Q^*(f) = Q(f)$, donde:

$$x(f) = Q(f)\, r(f) = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} \cdot \begin{pmatrix} 3 \\ -3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} .$$

Por outro lado, $H(f) = [I - Q(f) + Q^*(f)]^{-1} - Q^*(f) = I - Q(f)$, donde:

$$y(f) = H(f)\, r(f) = [I - Q(f)]\, r(f) - Q(f)\, r(f) =$$

$$= r(f) = \begin{pmatrix} 3 \\ -3 \end{pmatrix} .$$

Quanto a $Q^*(g) = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ a relação $Q(g)Q^*(g) = Q^*(g)$, fornece $a = c$ e $b = d$; além disso, sendo $v = (1/3,\ 2/3)$ tal que $vQ(g) = v$, também deveremos ter $vQ^*(g) = v$, ou seja:

$$(1/3,\ 2/3)\begin{pmatrix} a & b \\ a & b \end{pmatrix} = (1/3,\ 2/3),$$

que fornece $a = 1/3$ e $b = 1/3$; segue-se:

$$x(g) = Q^*(g)\ r(g) = \begin{pmatrix} 1/3 & 2/3 \\ 1/3 & 2/3 \end{pmatrix} \cdot \begin{pmatrix} 6 \\ -3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Por outro lado, temos:

$$\left[I - Q(g) + Q^*(g)\right]^{-1} = \begin{pmatrix} 4/3 & -1/3 \\ -1/6 & 7/6 \end{pmatrix}^{-1} = \begin{pmatrix} 7/9 & 2/9 \\ 1/9 & 8/9 \end{pmatrix},$$

donde:

$$y(g) = H(g)\ r(g) = \{[I-Q(g)+Q^*(g)]^{-1} - Q^*(g)\}r(g) = [I-Q(g)+Q^*(g)]^{-1}r(g) =$$

$$= \begin{pmatrix} 7/9 & 2/9 \\ 1/9 & 8/9 \end{pmatrix} \cdot \begin{pmatrix} 6 \\ -3 \end{pmatrix} = \begin{pmatrix} 4 \\ -2 \end{pmatrix} =.$$

Logo:

$$V_\beta(g) - V_\beta(f) = \left[\frac{x(g)}{1-\beta} + y(g) + \varepsilon(\beta,\ g)\right] -$$

$$- \left[\frac{x(f)}{1-\beta} + y(f) + \varepsilon(\beta, f)\right] = \begin{pmatrix} 4 \\ -2 \end{pmatrix} - \begin{pmatrix} 3 \\ -3 \end{pmatrix} +$$

$$+ \varepsilon(\beta,\ g) - \varepsilon(\beta,\ f) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \varepsilon(\beta, g) - \varepsilon(\beta, f) \longrightarrow \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$ quando $\beta \longrightarrow 1$, e portanto, f não é "quase-ótima".

Não obstante, $G(s,f) = \emptyset$, $s = 1, 2$ (omitimos os cálculos.)

§ 6 - CONJUNTOS DE BOREL E FUNÇÕES MENSURÁVEIS A BOREL

A partir de agora, encaminharemo-nos para uma generalização dos processos de decisão estudados nos parágrafos precedentes. Aqui os conjuntos $\mathcal{B}$ e A de estados e de ações são conjuntos não necessariamente enumeráveis, um grau de generalidade satisfatório é obrido considerando êstes conjuntos como conjuntos de Borel de um espaço métrico completo e separável.

Seja E um espaço métrico; lembramos que o espaço métrico E diz-se <u>completo</u> quando tôda sucessão de Cauchy tem um único ponto limite no espaço; <u>separável</u>, quando existe um subconjunto C de E, enumerável e denso em E, ou seja, $C = \{c_1, c_2, \ldots\ldots\}$ é tal que seu fêcho $\bar{C} = E$.

<u>Definição 6.1</u>

<u>A σ-álgebra de Borel $\mathcal{B}(E)$</u>, num espaço métrico E, é a σ-álgebra gerada pelos abertos de E(*); todo membro desta σ-álgebra diz-se <u>um conjunto de Borel</u> em E.

Observemos que se X é um conjunto de Borel em E, podemos considerar em X a Topologia induzida pela Topologia de E, sendo os "abertos em X" as intersecções dos abertos em E, com X. Em consequência, os "conjuntos de Borel em X", são as intersecções dos conjuntos de Borel em E, com X.

<u>Definição 6.2</u>

Seja X um conjunto de Borel; $f: X \longrightarrow \mathbb{R}$ diz-se mensurável a Borel ou $\mathcal{B}(X)$ - mensurável, quando $f^{-1}(B) \in \mathcal{B}(X)$, para cada conjunto de Borel B, na reta $\mathbb{R}$.

<u>Definição 6.3</u>

Se X e Y são conjuntos de Borel, uma <u>probabilidade condicional em Y, dado X</u>, é uma função $q(.|.)$, tal que:

(*) Para a noção de σ-álgebra gerada por um conjunto de partes de um conjunto dado, remetemos a Bartle, R.G.([illegible]) Cap.2, ou a Neveu,J. (11), Cap.1.

(i) para cada $\in X$, $q(.|x)$ é uma probabilidade em $Y^{(*)}$;

(ii) Para cada subconjunto de Borel B de Y, isto é, para cada $B \in \mathcal{B}(Y)$, então $q(B|.)$ é uma função mensurável a Borel definida em X.

<u>Notações</u>:

$P(X)$ = classe de todas as probabilidades em X, ou mais exatamente, no espaço probabilizável $(X, \mathcal{B}(X))$;

$Q(Y|X)$= classe de todas as probabilidades condicionais em Y, dado X;

$M(X)$ = classe das funções de Baire, limitadas em X.

XY = espaço produto X po Y.

<u>Definição 6.4</u>

(a) Se $p \in P(X)$ e $u \in M(X)$, então pu é definida como $\int u\, dp$.

(b) Se $u \in M(XY)$ e $q \in Q(Y|X)$, então podemos definir $qu\ M(X)$, tal que:

$$qu(x_o) = \int u(x_o, y)\, dq(y|x_o), \ \forall\, x_o \in X.$$

(c) Se $p \in P(X)$ e $q \in Q(Y|X)$, então podemos definir $pq \in P(XY)$ (isto é, uma probabilidade em XY), tal que:

$$pq(u) = p(qu), \ \forall\, u \in M(XY).$$

As definições dadas acima exigem alguns comentários.

1) Em (b), observemos que uma vez fixado x_o, $q(.|x_o)$ é uma probabilidade definida no espaço Y e tem sentido considerar a integral $\int u(x_o,y)\, dq(y|x_o)$; ao variarmos x_o, esta integral aparece como uma função de $x \in X$ real desta última variável, mensurável a Borel.

2) Agora vamos examinar os fatos por trás de (c). Ora, conforme (b), dados $q \in Q(Y|X)$ e $u \in M(XY)$, fica definido $qu \in M(X)$; por sua vez, conforme (a), dados $qu \in M(X)$ e $p \in P(X)$, fica

(*) isto é, $q(B|x) \in [0,1]$, $\forall\, B \in \mathcal{B}(Y)$; $q(\phi|x) = 0$ e $q(Y|x) = 1$; se $B_j \in \mathcal{B}(Y)$, $j = 1, 2, \ldots$, são disjuntos doic a dois, então:

$$q(B_1 + B_2 + \ldots |x) = \sum_{j=1}^{\infty} q(B_j|x).$$

definida $p(qu)$, tal que:

$$p(qu) = \int qu \, dp = \int \left[\int u(x,y) \, dq(y|x) \right] dp(x);$$

então, definimos $pq(u) = p(qu)$.

Observação

Toda probabilidade $m \in P(XY)$, admite uma fatorização $m = pq$, tal que $p \in P(X)$ e $q \in Q(Y|X)$; com efeito, tomemos $p \in P(X)$, tal que $p(B) = m(BY)$, $\forall\, B \in \mathcal{B}(X)$, isto é, p é a distribuição marginal de $\underline{m}$ relativamente à primeira coordenada; quanto a q, tomemos como qualquer versão da distribuição condicional da segunda coordenada, dada a primeira; tem-se, dados $B \in \mathcal{B}(X)$ e $C \in \mathcal{B}(Y)$:

$$pq(BC) = pq(u_{BC}) = p(qu_{BC}) = \int \left[\int u_{BC}(x,y) \, dq(y|x) \right] dp(x) =$$

$$= \int u_{BC}(x,y) \, dq(y|x) \, dp(x) = \int u_{BC}(x,y) \, dm(x,y) = m(BC),$$

isto é, $pq = m$.

Generalizações:

1º) Sejam X_1, X_2, ... conjuntos de Borel, não vazios. Se $p \in P(X_1)$ e $q_n \in Q(X_{n+1}|X_1, \ldots, X_n)$, $n \geq 1$, então podemos definir $pq_1, \ldots, q_n$, que é uma probabilidade em $X_1, X_2 \ldots, X_{n+1} = X_1 \times X_2 \times \ldots \times X_{n+1}$. (verifica-se por indução) isto é, $pq_1 \ldots p_n \in$ $\in P(X_1X_2 \ldots X_{n+1})$.

2º) Uma vez definidas $pq_1 \ldots q_n$, para cada $n \geq 1$, existe uma extensão $pq_1q_2 \ldots$.., que é uma probabilidade em $X_1X_2, \ldots = X_1 \times X_2 \times \ldots$, isto é, $pq_1q_2 \ldots \in P(X_1X_2 \ldots)$; com efeito, observe-se: $(pq_1q_2 \ldots q_{n+1})(B_1B_2 \ldots B_nX_{n+1}) = (p,q_1,q_2 \ldots q_n)(B_1B_2 \ldots B_n)$, $\forall\, B_i \in \mathcal{B}(X_i)$, $i = 1,2,\ldots,n$, ou seja, $pq_1q_2 \ldots q_{n+1}$ é uma extensão de $pq_1q_2 \ldots q_n$, para cada $n \geq 1$, portanto conforme o Teorema da extensão de Kolmogorov), podemos considerar $pq_1q_2 \ldots$, tal que:

$pq_1q_2 \ldots (B_1B_2 \ldots B_nX_{n+1}X_{n+2} \ldots) = pq_1 \ldots q_n(B_1B_2 \ldots B_n$, $\forall\, B_i \in \mathcal{B}(X_i)$ $i = 1,2,\ldots n$, $n \geq 1$.

3º) Dados $q_2 \in Q(X_3|X_1X_2)$ e $q_3 \in Q(X_4|X_1X_2X_3)$, podemos considerar $p_{(x_1x_2)} = q(.|x_1,x_2) \in P(X_3)$ e $q_{(x_1x_2)} = q_3(.|x_1,x_2,..) \in$

$\in Q(X_4|X_3$, para cada $(x_1,x_2) \in X_1X_2$; logo, por (c) da Def. 6.4, temos $p_{(x_1,x_2)}\, q_{(x_1,x_2)} \in P(X_3X_4)$, $\forall\ (x_1,x_2) \in X_1X_2$, tal que:

$$(p_{(x_1,x_2)}\, q_{(x_1,x_2)})\,(B) = (p_{(x_1,x_2)}\, q_{(x_1,x_2)})\,(\mu_B) =$$

$$= \int \Big[\int \mu_B(x_3,x_4)\, dq_{(x_1,x_2)}\,(x_4|x_3) \Big]\, dp_{(x_1,x_2)}(x_3) =$$

$$= \int \Big[\int \mu_B(x_3,x_4)\, dq_3(x_4|x_1,x_2,x_3) \Big]\, dq_2(x_3|x_1,x_2);\ \forall\ B \in \mathcal{B}(X_3X_4);$$

logo, se definirmos q_2q_3, tal que: $(q_2q_3)\,(B|x_1,x_2) =$

$= (p_{(x_1,x_2)}\, q_{(x_1,x_2)})\,(B)$, $\forall\ B \in \mathcal{B}(X_3,X_4)$, $\forall\ (x_1,x_2) \in X_1X_2$, tem-se

$$q_2q_3 \in Q(X_3X_4|X_1X_2).$$

Se $u \in M(X_1,X_2,X_3,X_4)$, podemos definir $q_2q_3u \in M(X_1X_2)$, como se segue:

$$(q_2q_3u)(x_1,x_2) = \int u(x_1,x_2,x_3,x_4)\, d(q_2q_3(x_3,x_4|x_1,x_2)),$$

$\forall\ (x_1,x_2) \in X_1X_2$.

4º) Se $qu \in Q(X_{n+1}|X_1\ldots.X_n)$, para $n \geq 1$, temos:

$q_m\ldots.q_n \in Q(X_{m+1}\ldots.X_{n+1}|X_1\ldots.X_n$, para $m \leq n$;

logo, dado $u \in M(X_1X_2\ldots.X_{n+1})$, temos: $q_m\ldots q_nu \in M(X_1X_2\ldots.X_m)$.

Notação:

Dada uma função u em Y, usaremos o mesmo símbolo para representar a função v em XY, tal que $v(x,y) = u(y)$, $\forall$ y; por exemplo, qualquer $q \in Q(Y|X)$ serve para representar $q' \in Q(Y|ZX)$, tal que $q'(.|z,.) = q(.|.)$.

Definição 6.5

Uma probabilidade diz-se degenerada se é concentrada em algum ponto $x \in X$; uma probabilidade condicional $q \in Q(Y|X)$ diz-se degenerada quando cada $q(.|x)$ é degenerada.

Proposição 6.6

As probabilidade condicionais $q \in Q(Y|X)$ degeneradas são aque las para as quais existe uma função mensurável a Borel $f: X \longrightarrow Y$, tal que $q(\{f(x)\}|x) = 1$, $\forall x \in X$.

Dem.

1º) Seja $f: X \longrightarrow Y$ mensurável a Borel, tal que $q(\{f(x)\}|x)=1$, $\forall x \in X$; então, $q(.|x)$ é concentrada no ponto $f(x) \in Y$, $\forall x \in X$, isto é, $q(.|x)$ é degenerada, para cada $x \in X$, donde q é degenerada.

2º) Se q é degenerada, isto é, $q(.|x)$ é degenerada, para cada $x \in X$, então cada $q(.|x)$ é concentrada num ponto $y = f(x)$, donde $f: X \longrightarrow Y$ é tal que $q(\{f(x)\}|x) = 1$, $\forall x \in X$; resta mostrar que f é uma função (mensurável a Borel).

Ora, seja $B \in \mathfrak{B}(Y)$; $q(B|.)$ é uma função de Baire, donde:

$$f^{-1}(B) = \{x \in X;\ f(x) \in B\} = \{x \in X;\ y \in B \text{ e } q(y|x) = 1\} =$$
$$= \{x \in X;\ q(B|x) = 1\} =$$

pois $\{1\} \in \mathfrak{B}(Y)$.

Observação:

Uma tal $f: X \longrightarrow Y$ servirá para representar a $q(Y|X)$ degenerada a que está associada, no sentido seguinte: dada $u \in M(XY)$, se definirmos $fu(x) = u(x,f(x))$, $\forall x \in X$, então $fu(x)$ coincide com $qu(x)$; com efeito, como $q(.|x)$ é concentrada no ponto $f(x)$, tem-se:

$qu(x) = \int u(x,y)\, dq(y|x) = u(x,f(x)) = fu(x)$; note-se que $fu \in M(X)$.

Lema 6.7

Para quaisquer $q \in Q(Y|X)$, $u \in M(XY)$ e $\varepsilon > 0$, existe $f \in Q(Y|X)$, degenerada, tal que:

(1) $fu \geq qu$, isto é, $fu(x) \geq qu(x)$, $\forall x \in X$;

(2) $q(\{y;\ u(x_o, y) \geq u(x_o, f(x_o)) + \varepsilon\}|x_o) = 0$, $\forall x_o \in X$.

Dem.

Remetemos ao Apêndice II

§ 7 - CASO GERAL - PRELIMINARES

Generalizando o problema de programação dinâmica estocástica, com um fator de desconto β, $0 \leq \beta < 1$, consideraremos:

$\mathcal{S}$, A : conjunto de Borel de espaços métricos completos e separáveis, não vazios (respectivamente, o espaço dos "estados" e das "ações");

q : a lei de movimento do sistema, ou seja uma probabilidade condicional $q \in Q(\mathcal{S}|\mathcal{S}A)$;

r : o ganho, que é uma função real, limitada, mensurável a Borel, definida em $\mathcal{S}A\mathcal{S}$ isto é, $r \in M(\mathcal{S}A\mathcal{S})$.

Definição 7.1

Uma estratégia Π(*) é uma sequência $(\Pi_1, \Pi_2, \ldots..)$ onde cada $\Pi_n \in Q(A|H_n)$, sendo $H_n = \underbrace{\mathcal{S}A\mathcal{S}A\ldots\ldots\mathcal{S}}_{(2n-1)\text{ fatores}}$ o conjunto de todas as possíveis "histórias" do sistema, até o instante de ser tomada a n-ésima ação.

Obs.: Notaremos por $(\sigma_1, \alpha_1, \sigma_2, \ldots, \alpha_{n-1}, \sigma_n)$ cada elemento de H_n, onde σ_1 é o estado inicial, os σ_i $(i = 2, \ldots, n)$ são os estados subsequentes e α_i $(i = 1,2,\ldots,n)$ é a i-ésima ação.

Definição 7.2

Uma estratégia $\Pi = (\Pi_1, \Pi_2, \ldots.)$ diz-se uma estratégia de Markov, quando $\Pi_n \in Q(A|\mathcal{S})$, $n = 1,2,\ldots..$

Definição 7.3

Uma estratégia $\Pi = (\Pi_1, \Pi_2,\ldots..)$ diz-se uma estratégia não aleatórizada, de Markow, quando cada Π_n é um elemento "degenerado" de $Q(A|\mathcal{S})$, isto é, cada $\Pi_n \equiv f_n$ $\mathcal{S} \longrightarrow A$ (função mensurável a Borel)(*)

(*) Cada Π_n é "degenerado", no sentido de Def. 6.6; logo, $\Pi_n(f_n(s)|s) = 1$, $\forall\, s \in \mathcal{S}$

Definição 7.4

Uma estratégia Π é <u>não aleatorizado estacionário</u>, quando existe uma função mensurável a Borel $f: \mathcal{S} \to A$, tal que $\Pi = (f, f, \ldots) = (f^{(\infty)})$

Definição 7.5

Fixados uma estratégia $\Pi = (\Pi_1, \Pi_2, \ldots)$ e uma "lei de movimento" $q \in Q(\mathcal{S}|\mathcal{S}A)$, então:

a) Para cada $n \geq 1$ e $H^*_{n+1} = \underbrace{A\mathcal{S} \ldots \ldots A\mathcal{S}}_{2n \text{ fatores}}$, $\Pi_1 q \ldots \Pi_n q \in Q(H^*_{n+1}|\mathcal{S})$ é tal que(*): $(\Pi_1 q \ldots \Pi_1 q \ldots \Pi_n q)(A_1 S_2 \ldots A_n S_{n+1}|s) =$

$$= \int_{A_1 S_2 \ldots A_n S_{n+1}} d\Pi_1(\alpha_1|\sigma_1)\, dq(\sigma_2|\sigma_1\alpha_1) \ldots d\Pi_n(\alpha_n|\sigma_1, \alpha_1, \ldots, \sigma_n) .$$

$. \, dq(\sigma_{n+1}|\sigma_n\alpha_n)$, para $S_2, \ldots, S_{n+1} \in \mathcal{B}(\mathcal{S})$, $A_1, A_2, \ldots, A_n \in \mathcal{B}(A)$ (conjuntos de Borel) e o estado inicial $s = \sigma_1 \in \mathcal{S}$.

b) $e_\Pi = \Pi_1 q\, \Pi_2 q \ldots . \in Q(\Omega|\mathcal{S})$, $\Omega = A\mathcal{S}A\mathcal{S} \ldots$, é a extensão das $\Pi_1 q \ldots . \Pi_n q \in Q(H^*_{n+1}|\mathcal{S})$, $n \geq 1$.

Agora temos condições de definir o ganho total esperado, relativamente a estratégia Π escolhido e considerado um desconto β, $0 \leq \beta < 1$. Ora, dado um elemento de $\mathcal{S}\Omega$, êle é da forma (σ_1, w), com $w = (\alpha_1, \sigma_2, \alpha_2, \sigma_3, \alpha_3, \ldots)$; $(\sigma_1 w)$ diz-se uma "história" (ou "trajetória") do sistema, a que correspondem ganhos corrigidos, sucessivos: $\beta^{n-1}\, r(\sigma_n, \alpha_n, \sigma_{n+1})$, $n = 1, 2, \ldots .$, e portanto, um ganho total corrigido:

$$\boxed{u = \sum_{1}^{\infty} \beta^{n-1}\, r(\sigma_n, \alpha_n, \sigma_{n+1})} \;;$$

Então podemos observar que $u \in M(S\Omega)$.

Então podemos calcular o <u>ganho total esperado</u> (com o desconto ou correção β, $0 \leq \beta < 1$), para a escolha de uma certa estratégia Π é dada por:

(*) Observemos que $H_{n+1} = \mathcal{S}H^*_{n+1}$

$$\boxed{=V(\Pi) = e_\Pi \cdot u = \sum_{n=1}^{\infty} \beta^{n-1} \Pi_1 q \ldots \Pi_n qr}$$

que depende de cada estado inicial $s = \sigma_1$.

Com efeito, $e_\Pi \in Q(\Omega|\mathcal{S})$ e $u \in M(\mathcal{S}\Omega) \longrightarrow e_\Pi u \in M(\mathcal{S})$; ou seja para cada estado inicial s o valor de $V(\Pi)$ em s será denotado: $V(\Pi)(s)$.

Definição 7.6

(a) Dados $p \in P(\mathcal{S})$ e $\varepsilon > 0$, o plano Π^* diz-se (p.ε)-ótimo quando, para qualquer outro plano Π, tivermos:

$$p \{V(\Pi) > V(\Pi^*) + \varepsilon\} = 0$$

(b) Dado $\varepsilon > 0$, o plano Π^* diz-se ε-ótimo quando for (p,ε)-ótimo, $\forall$ $p \in P(\mathcal{S})$, ou equivalentemente, quando $V(\Pi)(s) \leq V(\Pi^*)(s) + \varepsilon$, para cada plano Π, $\forall$ $s \in \mathcal{S}$.

(c) O plano Π^* diz-se ótimo, quando for ε-ótimo, $\forall$ $\varepsilon > 0$, ou equivalentemente, $V(\Pi)(s) \leq V(\Pi^*)(s)$, para cada plano Π e $s \in \mathcal{S}$.

(d) Dado $p \in P(\mathcal{S})$, o plano Π^* diz-se p-ótimo, quando for (p,ε)-ótimo, $\forall$ $\varepsilon > 0$.

Observemos que a definição de plamo ótimo contém a da estratégia βótima, do "caso discreto".

Planos (p,ε)-ótimos sempre existem (vide § seguinte); porém, não se pode garantir a existência de planos p-ótimos ou ε-ótimos, conforme os exemplos abaixo.

Exemplo 1 (não existência de planos p-ótimos)

Sejam: $S = \{0\}$ (isto é, S tem um único elemento) e $A = \{1, 2, 3, \ldots\}$. Tomemos $r(0, a, 0) = (a-1)/a$.

O ganho (corrigido), para $(\sigma_1, \alpha_1, \sigma_2, \alpha_2, \sigma_3, \ldots) = (0, \alpha_1, 0, \alpha_2, \ldots)$, é:

$$\sum_{i=1}^{\infty} \beta^{i-1} \frac{\alpha_i - 1}{\alpha_i} < 1 + \beta + \beta^2 + \ldots = 1/(1-\beta).$$

Logo, não há nenhum Π com $V(\Pi) = 1/(1-\beta)$, embora $\sup_{\Pi} V(\Pi) = 1/(1-\beta)$, pois, com a estratégia Π_n que obriga à escolha da ação $\underline{n}$, em todos os instantes, tem-se:

$V(\Pi_n) = \sum_{i=1}^{\infty} \beta^{i-1} \frac{n-1}{n} = [1/(1-\beta)] [(n-1)/n]$ e $\sup V(\Pi_n) = 1/(1-\beta)$.

Exemplo 2 (não existência de planos ε-ótimos)

Seja $\mathcal{S} = A = [0,1]$. A transição é do tipo $(s,a) \in \mathcal{S}A \rightsquigarrow s$, isto é, o estado fica inalterado, de instante a instante, qualquer que seja a ação $\underline{a}$ escolhida, $q(s|s,a) = 1$; por outro lado, seja B um certo conjunto de Borel em $\mathcal{S}A$ e seja o ganho $\underline{r}$ dado por: $r(s,a,s) \equiv r(s,a) = 1$, se $(s,a) \in B$, e $r(s,a,s) \equiv r(s,a) = 0$, se $(s,a) \notin B$ (o ganho depende apenas de duas variáveis, pois não se permite mudança de estado).

Ora, para qualquer $\Pi = (\Pi_1, \Pi_2, \ldots)$ tem-se $\{s; \Pi_1\, qr > 0\}$ é um conjunto de Borel, na projeção D de B em $\mathcal{S}$.

Então, escolhamos B tal que D não seja um conjunto de Borel, em $\mathcal{S}$; assim, existe $s_0 \in D$, tal que $\Pi_1 r = 0$, pois o conjunto dos $s \in \mathcal{S}$ tais que $\Pi_1\, qr \equiv \Pi_1\, r > 0$ é um subconjunto de Borel de D; ora, D não sendo um conjunto de Borel, então possui pelo menos um ponto $s_0 \in D$, tal que $s_0 \notin \{s; \Pi_1\, qr > 0\}$ ou seja $\Pi_1\, r = 0$; logo $V(\Pi) = \sum_{1}^{\infty} \beta^{n-1} \Pi_1\, q \ldots \Pi_n qr$, aplicando a s_0, nos dá: $V(\Pi)(s_0) = \sum_{2}^{\infty} \beta^{n-1} (\Pi_1 q \ldots \Pi_n qr)(s_0) \leq \beta + \beta^2 + \ldots \ldots = \beta/(1-\beta)$.

Como existe um Π^* tal que $V(\Pi^*)(s_0) = 1/(1-\beta)$, pois escolhemos a_0 tal que $(s_0, a_0) \in B$; isto é possível, pois, $s_0 \in D = \text{proj}.B$. Tomaremos Π^* tal que a ação escolhida seja sempre a_0; nêste caso, o ganho $\underline{r}$ em cada etapa é sempre 1, para s_0 inicial, donde $V(\Pi^*)(s_0) = 1 + \beta + \beta^2 + \ldots \ldots = 1/(1-\beta)$, vemos que:

$V(\Pi^*)(s_o) - V(\Pi)(s_o) = \dfrac{1}{1-\beta} - \dfrac{1}{1-\beta} = 1$, dondo

$V(\Pi^*)(s_o) - V(\Pi)(s_o) > \varepsilon$, $\forall\ \varepsilon < 1$, e portanto Π não é ε-ótima.

§ 8. CASO GERAL - EXISTÊNCIA DE ESTRATÉGIAS (p,ε) - ÓTIMOS E DE ESTRATÉGIAS NÃO ALEATORIZADAS (p,ε) - DOMINANTES

Teorema 8.1

Para cada $p \in P(\mathfrak{S})$ e cada $\varepsilon > 0$, existe uma estratégia (p,ε)-ótima.

Dem. (*)

1º) Sejam Π uma estratégia arbitrária; se $p \in P(X)$ é a distribuição de probabilidade para o estado inicial, então $p.V(\Pi)$ é o ganho esperado. Lembremos que de acôrdo com a notação adotada, tem-se $p.V(\Pi) = \int V(\Pi)(s)\, dp(s)$. Particularmente, se fôr fixado um estado inicial σ_1, tem-se uma distribuição de probabilidade concentrada em σ_1, isto é $p(\sigma_1) = 1$, donde $p.V(\Pi) = V(\Pi)(\sigma_1)$.

Tomemos $v = \sup_{\Pi} p.V(\Pi)$, e portanto, podemos escolher uma sequência de estratégias $\Pi^{(1)}, \Pi^{(2)}, \ldots$, tais que $p.V(\Pi^{(n)}) \longrightarrow v$; seja $\mu(s) = \sup_n V(\Pi^{(n)})_{(s)}$.

Em seguida, para cada n, consideremos o conjunto S_n de todos os $s \in \mathfrak{S}$, para os quais n é o menor dos k que satisfazem $V(\Pi^{(k)})(s) \geq \mu(s) - \varepsilon$, pela própria definição esta sequência é de conjuntos disjuntos e para todo $s \in \mathfrak{S} \exists n ; s \in S_n$.

Logo, $\mathfrak{S} = S_1 + S_2 + \ldots + S_n + \ldots + \ldots$.

Então consideremos a estratégia Π^*, que coincide com $\Pi^{(n)}$ em cada $s \in S_n$, isto é, se $\Pi^* = (\Pi_1^*, \Pi_2^*, \ldots)$, tem-se:

$\Pi_m^*(\cdot|\sigma_1, a_1, \ldots, \sigma_m) = \Pi_m^{(n)}(\cdot|\sigma_1, a_1, \ldots, \sigma_m)$, para todo $\sigma_1 \in S_n$ e $m = 1, 2, \ldots$; em consequência, teremos:

$V(\Pi^*)(s) = V(\Pi^{(n)})(s)$, $\forall\, s \in S_n$, $\forall\, n$, e como nesta circunstância $V(\Pi^{(n)})(s) \geq \mu(s) - \varepsilon$, segue-se:

$V(\Pi^*)(s) \geq \mu(s) - \varepsilon$, $\forall\, s \in S_n$, $\forall\, n$, e uma vez que $\mathfrak{S} = S_1 + S_2 + \ldots\ldots$, obtemos, finalmente, que $V(\Pi^*)(s) \geq \mu(s) - \varepsilon\ \forall\, s \in S$, isto é:

(*) Apesar de simples, infelizmente esta demonstração não é construtiva.

$$\boxed{V(\Pi^{*}) \geq \mu - \varepsilon} \qquad (1)$$

2º) Agora mostremos que para cada Π se tem $p\{V(\Pi) \leq \mu\} = 1$, isto é, $p\{s \in \;;\; V(\Pi)(s) \leq \mu(s)\} = 1$, ficando então provado que Π^{*} é (p,ε)-ótimo, pois (1) vale para todo $s \in \mathcal{S}$.

Ora, tomemos qualque Π e qualquer real $\gamma > 0$; em vez de considerarmos a sequência $\Pi^{(1)}, \Pi^{(2)}, \ldots, \Pi^{(n)}, \ldots$, consideremos a sequência $\Pi, \Pi^{(1)}, \Pi^{(2)}, \ldots, \Pi^{(n)}, \ldots$; da mesma forma como se construiu Π^{*} partindo da primeira sequência, construiremos Π^{**} partindo da segunda sequência, e sendo $\mu' = \sup\,(V(\Pi), V(\Pi^{(1)}), \ldots)$, obtemos $\mu' = \max\,(V(\Pi), \mu)$; analogamente, teremos $V(\Pi^{**})(s) \geq \mu'(s) - \gamma$, $\forall\, s \in \mathcal{S}$, isto é, $V(\Pi^{**}) \geq \mu' - \gamma$, donde $V(\Pi^{**}) \geq \max(\mu, V(\Pi)) - \gamma$.

Mas, $pV(\Pi^{**}) \leq v \leq p\mu$ (*), enquanto $pV(\Pi^{**}) \geq p\,\max(\mu, V(\Pi)) - \gamma$, donde se obtem:

$p\,\max\,(\mu, V(\Pi)) \leq p\mu + \gamma$ como $\gamma > 0$ é um número real arbitrário, segue-se $p\,\max(\mu, V(\Pi)) \leq p\mu$, e como $p\mu \leq p\,\max(\mu, V(\Pi))$ segue-se:

$$\mu \overset{q.s}{=} \max(\mu, V(\Pi)),$$

donde $V(\Pi) \overset{q.s}{\leq} \mu$, ou seja:

$$p\{V(\Pi) \leq \mu\} = 1.$$

<u>Teorema 8.2</u>

Para cada $p \in P(\mathcal{S})$ e cada $\varepsilon > 0$ e cada plano Π, existe um plano Π^{*}, **não aleatorizado**, de Markov, que (p,ε)-domina o plano Π, ou seja:

(*) A primeira desigualdade é óbvia; a segunda, verifica-se como se segue:

$$v = \sup_{\Pi} p\,V(\Pi) = \sup_{n} pV(\Pi^{(n)}) = \sup_{n} \int V(\Pi^{(n)})\,dp \leq \int \sup_{n} V(\Pi^{(n)})\,dp = \int \mu\,dp = p\mu.$$

$$p\{V(\pi^*) \geq V(\pi) - \varepsilon\} = 1$$

<u>Dem.</u>

1º) Suponhamos que π e π', são planos que coincidem nos N primeiros instantes; se $u \in M(\mathcal{S})$, notaremos $||u|| = \sup_s |u(s)|$; segue-se:

$$V(\pi) - V(\pi') = e_\pi u - e_{\pi'} u = \sum_{n=1}^{\infty} \beta^{n-1} \pi_1 q \ldots \pi_n qr -$$

$$- \sum_{n=1}^{\infty} \beta^{n-1} \pi_1' q \ldots \pi_n' qr = \sum_{n=N+1}^{\infty} \beta^{n-1} \pi_1 q \ldots \pi_n qr -$$

$$- \sum_{n=N+1}^{\infty} \beta^{n-1} \pi_1' q \ldots \pi_n qr, \text{ donde:}$$

$$|| V(\pi) - V(\pi') || = || \beta^N \sum_{n=1}^{\infty} \beta^{n-1} [\pi_1 q \ldots \pi_{n+N} qr -$$

$$- \pi_1' q \ldots \pi_{n+N}' qr] || \leq \beta^N \sum_{n=1}^{\infty} \beta^{n-1} || \pi_1 q \ldots \pi_{n+N} qr -$$

$$- \pi_1' q \ldots \pi_{n+N}' qr || \leq \beta^N \sum_{n=1}^{\infty} \beta^{n-1} (||\pi_1 q \ldots \pi_{n+N} qr || +$$

$$+ ||\pi_1' q \ldots \pi_{n+N}' qr ||) = \beta^N \sum_{n=1}^{\infty} \beta^{n-1} (\sup_s |\pi_1 q \ldots \pi_{n+N} qr | +$$

$$+ \sup_s |\pi_1' q \ldots \pi_{n+N}' qr |);$$

por outro lado, é imediato, que $\sup_s |\pi_1 q \ldots \pi_{n+N} q| \leq || r ||$ e analogamente, para as π_j' em lugar das π_j.

Em consequência:

$$|| V(\pi) - V(\pi') || \leq \beta^N \sum_{n=1}^{\infty} \beta^{n-1} (||r|| + ||r||),$$

isto é,

$$|| V(\pi) \quad V(\pi') || \leq 2\beta^N \frac{||r||}{1 - \beta},$$

2º) Ora, se tomarmos $\pi' = (\pi_1, \ldots \pi_N, f_{n+1}, f_{N+2}, \ldots)$, donde as f_i, $i \geq N+1$ são não aleatorizadas, e com N suficientemente grande de sorte que $2\beta^N ||r|| / (1-\beta) < \varepsilon/2$, teremos:

$$|| V(\pi) - V(\pi') || < \varepsilon/2.$$

(observemos que Π' se comporta como não estocástica, de Markov, para $n > N$).

Em seguida, pode-se mostrar que existe f_N tal que, para $\Pi'' = (\Pi_1, \ldots \Pi_{N-1}, f_N, f_{N+1}, \ldots)$ se tenha $p\{V(\Pi'') \geq V(\Pi') - \gamma\} = 1$ onde $\gamma = \varepsilon/2N$ (será deixado para se provar no final).

Ora, aceito êste fato, podemos determinar $\Pi''' = (\Pi_1, \ldots \Pi_{N-2}, f_{N-1}, f_N, f_{N+1}, \ldots)$ tal que $P\{V(\Pi''') \geq V(\Pi'') - \gamma$, e assim sucessivamente, até encontrar uma $\Pi^* = (f_1, f_2, \ldots)$ que é não estocástica, de Markov, e teremos:

$$P\{V(\Pi^*) \geq V(\Pi') - N.\varepsilon/2N\} = p\{V(\Pi^*) \geq V(\Pi') - \varepsilon/2\} = 1.$$

3º) Como $||V(\Pi) - V(\Pi')|| < \varepsilon/2$ implica(*) $|V(\Pi)(s) - V(\Pi')(s)| < \varepsilon/2$, $\forall$ s, que por sua vez implica $V(\Pi')(s) > V(\Pi)(s) - \varepsilon/2$, $\forall$ s; isto, combinado com $p\{V(\Pi^*) \geq V(\Pi') - \varepsilon/2\} = 1$, fornece:

$$p\{V(\Pi^*) \geq V(\Pi) - \varepsilon\} = 1,$$

isto é, o plano Π^* é (p,ε) - dominante com relação ao plano Π, e além disso, Π^* é não estocástica, de Markov.

4º) Só resta provar o resultado indicado no item (2º), ou seja: dada $\Pi' = (\Pi_1, \ldots, \Pi_N, f_{N+1}, \ldots)$, existe f_N, tal que $p\{V(\Pi'') \geq V(\Pi') - \gamma\} = 1$, onde $\gamma = \varepsilon/2N$ e $\Pi'' = (\Pi_1, \ldots, \Pi_{N-1}, f_N, f_{N+1}, \ldots)$.

Observe-se que:

$\Pi_1 q \ldots \Pi_{N-1} q\, r(s_k, a_k, s_{k+1}) = \Pi_1 q \ldots \Pi_k\, qr(s_k, a_k, s_{k+1})$, para $k \leq N-1$ (com efeito, Π_j q não depende de s_k, a_k, s_{k+1}, quando $j > k$.

Logo, se $J = \Pi_1 q \ldots \Pi_{N-1} q\, (u + \beta^{N-1} \Pi_N\, q\, v)$ onde $u\,(s_1, a_1, \ldots, s_N) = \sum_{k=1}^{N-1} \beta^{k-1} r(s_k, a_k, s_{k+1})$ e $v = V(s_N, a_N, s_{N+1}) = r(s_N, a_N, s_{N+1}) + \sum_{k=1}^{\infty} \beta^k f_{N+1} q \ldots f_{N+k}\, qr$(**), segue-se:

(*) pois $||V(\Pi) - V(\Pi')|| = \sup_s |V(\Pi)(s) - V(\Pi')(s)|$.

(**) Notaremos u no lugar de $u(s_1, a_1, \ldots, s_{N+1})$, v no lugar de $v(s_N, a_N, s_{N+1})$ e r no lugar de qualquer $r(s_n, a_n, s_{n+1})$.

$$J = \Pi_1 q \ldots \Pi_{N-1} q \,[(\sum_{k=1}^{N-1} \beta^{k-1} r) +$$

$$+ \beta^{N-1} \Pi_N q (r + \sum_{k=1}^{\infty} \beta^K f_{N+1} q \ldots\ldots f_{N+k} qr)] =$$

$$= \sum_{k=1}^{N-1} \beta^{k-1} \Pi_1 q \ldots\ldots \Pi_k qr + \beta^{N-1} \Pi_1 q \ldots \Pi_{N-1} q \Pi_N qr +$$

$$+ \sum_{k=N+1}^{\infty} \beta^k \Pi_1 q \ldots \Pi_{N-1} q f_{N+1} q \ldots f_{N+k} r = \sum_{k=1}^{\infty} \Pi'_1 q \ldots \Pi'_k qr =$$

$= V(\Pi')$, isto é,

$$J = \Pi_1 q \ldots \Pi_{N-1} q (u + \beta^{N-1} \Pi_N qu) = V(\Pi'), \qquad (1)$$

onde u e v são definidos como anteriormente.

Em seguida, encontraremos f_N, tal que:

$$P\{\Pi_1 q \ldots \Pi_{N-1} qf_N w = \Pi_1 q \ldots \Pi_{N-1} q \Pi_N - \gamma\} = 1 \quad (2)$$

onde $w = \beta^{N-1} qv \in M(SA)$.

Ora, consideremos a probabilidade $m = p\Pi_1 q \ldots \Pi_N$, em $\mathcal{S}A \ldots \mathcal{S}A$ (2N fatores), e representemos as coordenadas variáveis por $\sigma_1, \alpha_1, \ldots, \sigma_N, \alpha_N$; para qualquer f_N, tem-se que $x = \Pi_1 q \ldots \Pi_{N-1} qf_N w(\sigma_1)$ é uma versão de $E(w(\sigma_N, f_N(\sigma_N))|\sigma_1)$, enquanto $y = \Pi_1 q_1 \ldots \Pi_{N-1} q \Pi_N w(\sigma_1)$ é uma versão de $E(w(\sigma_N, \alpha_N)|\sigma_1)$. Então, se escolhermos f_N tal que $w(\sigma_N, f(\sigma_N)) \geq w(\sigma_N, \alpha_n) - \gamma$, com probabilidade 1, segue-se $x > y - \gamma$, com probabilidade 1, que é equivalente a (2).

Que um tal f_N existe, é decorrência do Lema 6.8, com $X = \mathcal{S}$; $Y = A$; q uma versão da distribuição condicional de α_N, dada σ_N; $u = w$; $\varepsilon = \gamma$.

Isto completa a prova.

<u>Corolário 8.3</u>

Para todo $p \in P(\mathcal{S})$ e $\varepsilon > 0$, existe um plano de Markov Π^* que é (p,ε)-ótimo.

Dem.

Conforme o Teorema 8.1, existe um plano Π que é $(p,\varepsilon/2)$-ótimo; pelo Teorema 8.2, existe Π^* que $(p,\varepsilon/2)$-domina o plano Π. Então, Π^* é (p,ε)-ótimo.

C.Q.D.

O seguinte exemplo mostra que o teorema 8.2 não pode ser melhorado, no sentido de substituirmos a (p,ε)-dominação pela ε-dominação.

Exemplo (Uma estratégia Π que não pode ser ε-dominado por nenhuma estratégia de Markov)

1º) Tomamos $X = Y = [0, 1]$ e B um subconjunto de Borel do quadrado unitário XY tal que a projeção D, de B sôbre X, não é de Borel; $\mathcal{S} = B \cup X$.

Consideremos a seguinte lei de movimento, para $x \in X$ e $(x,y) \in B$:

$$(x,y) \to x$$

$$x \to x;$$

logo se trata de uma lei de movimento q, degenerada e independe da ação $\underline{a}$, isto é,

$q(\{x\} \mid (x,y)) = 1$, $\forall\ (x,y) \in B$, e $q(\{x\} \mid x) = 1$, $\forall\ x \in X$.
O conjunto de ações é $A = [0, 1]$, sendo r dada por:

$$(1) \quad \begin{cases} r(x, a, x) = 1, & \text{se } (x, a) \in B \\ r = 0, & \text{caso contrário.} \end{cases}$$

2º) Agora, tomamos qualquer estratégia $\Pi^* = (\Pi_1^*, \Pi_2^*, \ldots)$, tal que $\Pi_n^*(\cdot|s_1, a_1, \ldots, s_n)$ é degenerado em y, quando $s_1 = (x,y)$; mostraremos que, nêste caso, $V(\Pi^*) = \beta/(1-\beta)$ em B, isto é,

$$V(\Pi^*)(s) = \beta/(1-\beta), \ \forall\ s = (x,y) \in B.$$

Tem-se: $V(\Pi^*) = \sum_{n=1}^{\infty} \beta^{n-1} \Pi_1^* q \ldots \Pi_n^* qr =$

$= \Pi_1^* qr + \sum_{n=2}^{\infty} \beta^{n-1} \Pi_1^* q \ldots \Pi_n^* qr$; $(\Pi_1^* qr)(s) = 0$

porque $s = (x, y) \in B$ e de acôrdo com (1), $r((x,y),y,x) = 0$.

A seguir mostraremos que $(\Pi_1^* q \ldots \Pi_n^* qr)(s) = 1$ se $s = (x,y) \in B$; assim, para qualquer "trajetória" (s, a, s', a', s''), começando com $s = (x,y) \in B$, teremos obrigatòriamente $s' = s'' = x$, devido à lei de movimento considerada; por outro lado, como $\Pi_1^* (.|(x,y))$ é uma probabilidade concentrada em y, a primeira ação será obrigatòriamente $a = y$, e da mesma forma como $\Pi_2^*(.|(x,y),y,x)$ é concentrada em y, a segunda ação será $a' = y$; em consequência:

$(\Pi_2^* q \Pi_1^* qr)(x,y) = \int\int\int [\int r(x,y,x)\, dq]\, d\Pi_1^*\, dq\, d\Pi_2^* = 1$, uma vez que $r(x, y, x) = 1$, de acôrdo com (1).

Anàlogamente, obtemos $\Pi_1^* q \ldots \Pi_n^* qr = 1$, para todo estado inicial $s = (x, y) \in B$, e portanto:

$V(\Pi^*)(s) = \beta + \beta^2 + \ldots + \beta^n + \ldots = \beta(1 + \beta + \ldots + \beta^{n-1} + \ldots) =$

$= \beta/(1-\beta)$, $\forall\, s = (x, y) \in B$.

3º) Finalmente, seja $\Pi = (\Pi_1, \Pi_2, \ldots.)$ tal que $\Pi_2 \in Q(A|\mathcal{S})$, isto é, Π_2 não depende do estado inicial(*). Por outro lado, consideremos o conjunto $x \in X$, tais que $\Pi_2\, qr > 0$, isto é, $\{x \in X;\ (\Pi_2\, qr)(x) < 0\}$; observemos que, para $s = x$, temos: $(qr)(x,a) = \int r(x, a, s')\, dq\, (s'|x)$, e como $q\, (\{x\}|x) = 1$, obtemos $(qr)(x,a) = r(x, a, x)$. Ora, $(\Pi_2\, qr)(x) > 0$ implica

$(\Pi_2\, qr)(x) = \Pi_2\, (qr(x, a)) = \Pi_2\, (r(x, a, x)) =$

$= \int r\,(x, a, x)\, d\,\Pi_2 > 0$, donde devemos ter $r\,(x, a, x) = 1$, isto só ocorrendo quando $(x,a) \in B$, donde $x \in \mathrm{proj}_1.B = D$.

(*) Geralmente, $\Pi_2 \in Q(A|H_2) = Q(A|\mathcal{S}A\mathcal{S})$; se $\Pi_2 \in Q(A|\mathcal{S})$, então Π_2 não não depende do estado inicial, nem da primeira ação empreendida.

Logo temos: $\{x \in X;\ (\Pi_2\ qr)(x) > 0\} \subset D$, sendo aquêle um conjunto de Borel(*); como D não é de Borel, esiste $x_o \in D$, tal que $(\Pi_2\ qr)(x_o) \leq 0$(**); mas r só pode assumir os valores $\underline{0}$ a $\underline{1}$, donde a integral $(\Pi_2\ qr)(x_o) \geq 0$; as duas desigualdades fornecem $(\Pi_2\ qr)(x_o) = 0$. Em seguida, para qualquer y_o tal que $(x_o, y_o) \in B$ (repetições de técnicas já empregadas), obtemos:

$$V(\Pi)(x_o, y_o) \leq \beta^2/(1-\beta).$$

Esta desigualdade, junto com $V(\Pi^*) = \beta/(1-\beta)$, em B, fornece: $V(\Pi)(x_o, y_o) - V(\Pi^*)(x_o, y_o) \leq [\beta^2/(1-\beta)] - [\beta/(1-\beta)] = -\beta$, donde $V(\Pi)(x_o, y_o) \leq V(\Pi^*)(x_o, y_o) - \beta$, isto é, $V(\Pi^*)(x_o, y_o) > V(\Pi)(x_o, y_o) + \varepsilon$, $\forall\ \varepsilon < \beta$, portanto, nenhum dos planos Π^* pode ser ε-dominado por um plano Π tal que $\Pi_2 \in Q(A|\mathcal{B})$, e, em particular por nenhum plano de Markov.

(*) $\Pi_2\ qr$ é uma função de Baire e portanto $\{\Pi_2\ qr > 0\}$ é um conjunto de Borel.

(**) Do contrário, teríamos D igual ao segundo subconjunto, absurdo!

§ 9. ESTRATÉGIAS ESTACIONÁRIAS E OPERADORES

Definição 9.1 - Operador T

Associado a cada função de Baire $f: \mathcal{S} \to A$, consideremos o operador $T: M(\mathcal{S}) \to M(\mathcal{S})$, tal que $Tu = fq\,(r + \beta u)$, $\forall\, u \in M(\mathcal{S})$, onde o u da direita, considerado como função em $\mathcal{S}A\mathcal{S}$, depende só da última coordenada; assim, f podendo ser interpretada como uma distribuição condicional degenerada, em A, para casa $s \in \mathcal{S}$, segue-se que a notação precedente indica:

$$(Tu)(s) = (fq\,(r + \beta u))(s) =$$
$$= \int \left[r(s, f(s), s') + \beta u(s')\right] dq(s' | s, f(s)), \ \forall\, s \in \mathcal{S}.$$

Interpretação: Tu é o ganho esperado, em função do estado inicila, supondo que utilizamos $f^{(\infty)}$ e que o processo finda no início do 2º instante, com um ganho final $u(s')$, sendo s' o estado final; analogamente, $T^n u$ é o ganho esperado, também em função do estado inicial, ainda com a suposição de utilizarmos $f^{(\infty)}$ e de que o processo finda no início do (n+1)-ésimo instante, com ganho final $u(s')$, se s' é o estado final.

O seguinte Teorema resume as propriedades do operador T:

Teorema 9.2

(a) T é monótono, isto é $u \leq v$ implica $Tu \leq Tv$

(b) Para toda constante c, temos $T(u + c) = Tu + \beta c$.

(c) Para toda estratégia de Markov, $\Pi = (f_1, f_2, \ldots)$, temos $TV(\Pi) = V(f,\Pi)$, onde (f,Π) é a estratégia de Markov $(f, f_1, f_2, \ldots\ldots)$.

Dem. de (a)

$(Tu)(s) = (fqr)(s) + \beta(fu)(s) = \int r(s,f(s),s')\, dq(s'|s,f(s)) +$
$+ \beta \int u(s')\, dq(s'|s,f(s))$, $\forall\, s$; logo, $u \leq v$, isto é, $u(s') \leq v(s')$, $\forall\, s'$, implica $\int u(s')\, dq(s'|s\ ,f(s)) \leq \int v(s')\, dq(s'|s,f(s))$, $\forall\, s$, donde $(Tu)(s) \leq (Tv)(s)$, $\forall\, s$, isto é, $Tu \leq Tv$.

Dem. de (b)

$T(u + c) = fq\ (r + \beta(u + c)) = fq(r + \beta u + \beta c) = fq\ (r + \beta u) +$
$+ fq\ \beta c = Tu + \beta c$.

Demo. de (c)

$V(\Pi) = \sum_{n=1}^{\infty} \beta^{n-1}\, \Pi_1\, q\ \ldots.\ \Pi_n\, qr = \sum_{n=1}^{\infty} \beta^{n-1}\, f_1\, q\ \ldots.\ f_n\, qr$; segue-se: $TV(\Pi) = T(V(\Pi)) = fgr + \ \beta fgV(\Pi) = fgr + \beta\, fq(\sum_{\Pi=1}^{\infty} \beta^{n-1} f_1 q \ldots f_n qr) =$
$= fqr + \sum \beta^n\, f\, q\, f_1\, q\ \ldots\ f_n\, qr =$
$= \sum_{n=1}^{\infty} \beta^{n-1} g_1\, q\ \ldots\ g_n\, qr = V(\bar{\Pi}) = V(f,\Pi)$, onde $\bar{\Pi} = (f,\Pi) =$
$= (g_1, q_2, \ldots., g_n, \ldots.)$ e $g_1 = f$, $g_2 = f_1$, $\ldots.$, $g_n = f_{n-1}$, $\ldots.$

C.Q.D.

Definição 9.3

(i) Dado uma estrátégia $\Pi = (f_1, f_2, \ldots)$, de Markov dizemos que $f: S \twoheadrightarrow A$ é uma função Π-gerada quando existir uma partição $\mathcal{S} = s_1 + s_2 + \ldots$, o- s_i todos conjuntos de Borel, tais que, a restrição de f a S_n é igual a f_n, $n = 1, 2, \ldots$, (isto é, $f \mid S_n = f_n$).

(ii) $\Pi' = (g_1, g_2, \ldots.)$ diz-se uma estratégia Π- gerada, quando cada g_n é Π- gerada.

Definição 9.4 - Operador U

Associado a cada estratégia $\Pi = (f_1, f_2, \ldots)$, de Markov consideraremos o operador $U: M(\mathcal{S}) \longrightarrow M(\mathcal{S})$, tal que:

$$U(u) = \sup_{u} T_n u,$$

onde T_n é o operador (vide Def. 9.1) associado a f_n, para $n = 1, 2, \ldots..$

Interpretação: $U^n u$ é o ganho esperado ótimo, relativamente a todas as estratégias Π', de Markov, que são Π-geradas, se iniciamos usando Π' e terminamos no início do $(n + 1)$-ésimo dia com um ganho final $u(s')$, onde s' é o estado final, nesse dia(*).

O seguinte Teorema resume as propriedades do operador U.

Teorema 9.5

(a) U é monótono

(b) Para cada constante c, temos $U(u + c) = Uu + \beta c$.

(c) Para todo operador T associado a uma f que é Π-gerada, temos $Tu \leq Uu$.

(d) Se Π é uma estratégia de Markov, então para cada $u \in M(\mathcal{S})$ e cada $\varepsilon > 0$, existe f que é Π-gerada e é tal que o operador a ela associado satisfaz $Tu \geq Uu - \varepsilon$.

Dem. de (a)

$$u \leq v \xrightarrow{\text{Teor. } 9.2(a)} T_n u \leq T_n v,\ n = 1, 2, \ldots.. \Longrightarrow T_n u \leq \sup_n T_n v = Uv \Longrightarrow Uu = \sup_n T_n u \leq Uv.$$

Dem. de (b)

$$U(u + c) = \sup_n T_n (u + c) = \sup_n (T_n u + \beta c) = \sup_n T_n u + \beta c = Uu + \beta c.$$

Demo. de (c)

Uma vez que $\Pi = (f_1, f_2, \ldots\ldots)$ e f é Π-gerada, temos $f|S_n = f_n$, $n = 1, 2, \ldots.$ e $\mathcal{S} = S_1 + S_2 + \ldots\ldots$

(*) Adiante, apresentaremos uma cabal justificativa para tal interpretação.

Logo, dado $s \in \mathcal{S}$, existe n tal que $s \in S_n$, donde $f_n(s) = f(s)$, e portanto:

$$(Tu)(s) = (fqr)(s) + \beta(fqu)(s) = (f_n qr)(s) + \beta(f_n qu)(s) =$$
$$= (T_n u)(s) \leq \sup_n \left[(T_n u)(s)\right] = (Uu)(s);$$

assim, $(Tu)(s) \leq (Uu)(s)$, $\forall s \in \mathcal{S}$, isto é $Tu \leq Uu$.

<u>Dem. de (d)</u>

Sejam: $S_n = \{s;\ T_i u < Uu - \varepsilon$, se $i < n$, e $T_n u \geq Uu - \varepsilon\}$, $n = 1, 2, \ldots$; assim, os S_n são disjuntos dois a dois, e cada $s \in \mathcal{S}$ pertence a algum S_n; donde $\mathcal{S} = S_1 + S_2 + \ldots$; então, podemos definir f tal que $f(s) = f_n(s)$, para $s \in S_n$, $n = 1, 2, \ldots$.

Em consequência, para cada v, tem-se $Tv = T_n v$, em S_n, onde T é o operador associado a f; em particular, $Tu = T_n u \geq Uu - \varepsilon$, em S_n, donde $Tu \geq Uu - \varepsilon$, em todo $\mathcal{S}$.

Agora, estamos em condições de apresentar uma justificativa à interpretação de $U^n u$.

Observemos, inicialmente, que para cada estratégia $\Pi' = (g_1, g_2, \ldots)$ de Markov, o ganho total devido a Π', supondo-se o processo com término no $(n+1)$-ésimo instante e ganho final u, é dado através de:

$$V_n(\Pi', u) = T_1' T_2' \ldots T_n' u,$$

onde $V_n(\Pi', u)$ representa o ganho acima referido e T_i' é o operador associado a g_i. Se Π' é -gerada, então $T_i' v \leq Vv$, para cada i, por (9.5) (c), como decorrência, substituindo-se em $V_n(\Pi', u) = T_1' T_2' \ldots T_n' u$, obtemos $V_n(\Pi', u) \leq V^n u$. Então para encontrar Π', tal que $V_n(\Pi', u) \geq V^n u - \varepsilon$, escolhamos números positivos ε_i e funções g_i que sejam Π-geradas, tais que:

$$T_i' U^{n-i} u \geq UU^{n-i} u - \varepsilon_i = U^{n-i+1} u - \varepsilon_i$$

(as g_i existem como decorrência do Teorema 9.5(d)).

Por indução sôbre i, começando por $i = n$, obtemos:

$$T_i' \ldots\ldots T_n' u \geq U^{n-i+1} - d_i,$$

onde $d_i = \varepsilon_i + \beta\varepsilon_{i+1} + \ldots + \beta^{n-i} \varepsilon_n$. Para $i = 1$, obtemos $V_n(\Pi', u) \geq U^n u - d_1$, e os ε_i podem ser escolhidos tais que $d_1 \leq \varepsilon$.

Teorema 9.6

Se U é qualquer operador com propriedade (a) e (b) do Teorema 9.5, então U é uma contração com módulo β, isto é, $||Uu - Vv|| < \beta\, ||u - v||$, de sorte que, pelo Teorema do ponto fixo de Banach, U possui um único ponto fixo u^*, e $||U^n u - u^*|| \leq \beta^n ||u - u^*||$, para cada n.

Dem.

Lembremos que $||w|| = \sup|w(s)|$, onde $w \in M(S)$ (trata-se de uma norma). Então, para cada estado s, tem-se: $v(s) = u(s) + v(s) - u(s) \leq u(s) + |v(s) - u(s)| = u(s) + |u(s) - v(s)| \leq u(s) + \sup_{s'} |u(s') - v(s')| = u(s) + ||u-v|| (s)$, isto é $v \leq u + ||u-v||$.

Segue-se $Uv \leq U(u + ||u-v||) = Uu + \beta||u-v||$, conforme (a) e (b) do Teorema 9.5, isto é:

$$Uv \leq Uu + \beta||u-v|| \qquad (1).$$

Trocando os papeis de v e u, obtemos:

$$Uu \leq Uv + \beta||u-v|| \qquad (2)$$

Ora, (1) e (2) implicam que $Uv - Uu \leq \beta||u-v||$ e $Uu - Uv \leq \beta||u-v||$, donde $|Uu - Uv| \leq \beta||u-v||$, isto é, $|Uu - Uv|(s) \leq \beta||u-v||$, $\forall\, s \in \mathcal{S}$, donde $||Uu - Vv|| = \sup_s |Uu - Uv|(s) \leq \beta||u-v||$, isto é,

$$||Uu - Uv|| \leq \beta||u-v||.$$

Em consequência, pelo "Teorema do ponto fixo de Banch", o ope

dor U possue um ponto fixo u^* (isto é, tal que $Uu^* = u^*$) e
$||V^n u - u^*|| \leq \beta^n \, ||u - u^*||$, $\forall$ n.

C.Q.D.

Os principais resultados sôbre estratégias ótimos estão contidos no Teorema seguinte (são indicados os resultados correspondentes, obtidos por Dubins & Savage [13]).

Teorema 9.7

(a) Para cada estratégia $\Pi = (f_1, f_2, \ldots.)$ de Markov, representando por T_n o operador associado a f_n e $U = \sup T_n$ o operador associado a Π, o ponto fixo u^* de U é o ganho ótimo entre todos as estratégia de Markov Π' que Π-gerados, isto é,

(i) $V(\Pi') \leq u^*$, $\forall \, \Pi'$ que é Π-gerada;

(ii) $\exists$ f que é Π-gerada e tal que $V(f^{(\infty)}) \geq u^* - \varepsilon$.

Qualquer f, como $Tu^* \geq u^* - \varepsilon(1-\beta)$, satisfaz essa desigualdade.

(b) Para cada $p \in P(\mathcal{D})$ e $\varepsilon > 0$ existe uma estratégia estacionária $(p.\varepsilon)$-ótima.

(c) Se para cada $\varepsilon \geq 0$ existir $\Pi^* = (\Pi_1, \Pi_2, \ldots.)$ que é ε-ótima, então existe uma estratégia estacionária que é $\varepsilon/(1-\beta)$-ótima (é o Teorema 3.9.6, de [13]).

(d) Para cada $a \in A$ representemos por T_a o operador associado a $f \equiv a$. Para todo u se $T_a u \leq u$, para todo a, então $V(\Pi) \leq u$, para todo Π, isto é u é um limite superior para os ganhos esperados correspondentes a qualquer estratégia Π. (Teoremas 2.12.1 e 3.3.1, de [13]).

(e) Se para cada $\varepsilon > 0$, existir uma estratégia ε-ótima, então os ganho esperados u^* definido em a é uma função de Baire que satisfaz a equação:

$$u^* = \sup_a T_a u^* \quad \text{(equação de otimibilidade).}$$

(Teorema 3.3.1, de [13]).

(f) Uma estratégia Π é ótima se e só se seu ganho esperado satisfaz à equação $u^* = \sup_a T_a u^*$.

Dem. de (a)

1º) Seja $\Pi' = (g_1, g_2, \ldots\ldots)$ uma estratégia de Markov, Π-gerada; tem-se: $V(\Pi') = T_1', \ldots\ldots T_n' u_n$, onde $u_n = V(g_{n+1}, g_{n+2}, \ldots\ldots)$ e T_i' é o operador associado a g_i, isto é, $T_i' u = g_i q(r + \beta u)$, para cada $u \in M(\mathcal{S})$. De fato, isso pode ser verificado por indução, como se segue.

Vale para $n = 1$, isto é, $V(\Pi') = T_1' u_1$, pois: $T_1' u_1 =$
$= T_1' (V(g_2, g_3, \ldots\ldots)) = g_1 q (r + \beta V(g_2, g_3, \ldots\ldots) = g_1 q (r +$
$+ \beta \sum_{n=1}^{\infty} \beta^{n-1} g_2 q \ldots g_{n+1} qr) = g_1 q(r + \sum_{n=1}^{\infty} \beta^n g_2 q \ldots g_{n+1} qr) =$
$= g_1 qr + \sum_{n=1}^{\infty} \beta^n g_1 q \ldots g_{n+1} qr = g_1 qr + \sum_{n=2}^{\infty} \beta^{n-1} g_1 q \ldots g_n qr =$
$= \sum_{n=1}^{\infty} \beta^{n-1} g_1 q \ldots\ldots g_n qr = V(\Pi')$.

Valendo para n, vale para n+1, isto é $V(\Pi') = T_1' \ldots\ldots T_n' u_n \longrightarrow$
$\longrightarrow T_1' \ldots\ldots T_{n+1}'$, pois: $T'_{n+1} u_{n+1} = T'_{n+1}(V(g_{n+2}, g_{n+3}, \ldots\ldots)) =$
$= g_{n+1} qr + \beta g_{n+1} q V(g_{n+2}, g_{n+3}, \ldots) = V(g_{n+1}, g_{n+2}, \ldots\ldots) =$
u_n, e portanto $T_1', \ldots T'_{n+1} u_{n+1} = (T_1' \ldots T_n')(T'_{n+1} u_{n+1}) =$
$T_1', \ldots\ldots T_n' u_n = V(\Pi')$.

2º) Como cada T_i' é uma contração de módulo β, obtem-se:

$$||T_1' \ldots\ldots T_n' u_n - T_1' \ldots\ldots T_n' u^*|| \leq \beta^n ||u_n - u^*|| \leq \beta^n (\frac{||r||}{1 - \beta} + ||u^*||) \quad (1)$$

A primeira desigualdade, acima, pode ser verificada por indução (não apenas para u_n e u^*m mas em geral, dados u e u^* quaisquer), como se segue:

(i) Vale para n = 1: $||T_1' u - T_1' u'|| \leq \beta ||u - u'||$, conforme Teorema 9.6;

(ii) <u>Valendo para n, vale para n + 1:</u>

$$||T_1' \ldots\ldots T_{n+1}' u - T_1' \ldots\ldots T_{n+1}' u'|| \leq ||(T_1' \ldots . T_n')(T_{n+1}' u)$$
$$-(T_1' \ldots . T_n')(T_{n+1}' u')|| \leq \beta^n ||T_{n+1}' u - T_{n+1}' u'|| \leq \beta^n (\beta ||u - u'||)$$
$$= \beta^{n+1} ||u - u'||.$$

A segunda desigualdade em (1) é verificada da seguinte maneira: $||u_n|| = |u_n| = \sup_s |V_n(g_{n+1}, g_{n+2}, \ldots)| =$

$$= \sup_s \left| \sum_{m=1}^{\infty} \beta^{m-1} g_{n+1} q \ldots g_{n+m} q \right| \leq \sup_s \sum_{m=1}^{\infty} \beta^{m-1} g_{n+1} q \ldots g_{n+,} q |r| \leq$$

$$\leq \sum_{m=1}^{\infty} \beta^{m-1} g_{n+1} q \ldots g_{n+m} q \; \sup_s |r| = \sum_{m=1}^{\infty} \beta^{m-1} g_{m+1} q \ldots g_{n+m} q ||r|| =$$

$$= \sum_{m=1}^{\infty} \beta^{m-1} ||r|| = ||r|| / (1-\beta), \text{ donde:}$$

$$\beta^n ||u_n - u^*|| \leq \beta^n (||u_n|| + ||u^*|| \leq \beta^n (\frac{||r||}{1 - \beta} + ||u^*||).$$

Agora, observando que $n(\frac{||r||}{1 - \beta} + ||u^*||) \longrightarrow 0$, quando $n \longrightarrow \infty$ (pois, $0 < \beta < 1$), pode-se concluir de (1) que:

$||T_1' \ldots . T_n' u_n - T_1' \ldots T_n' u^*|| \longrightarrow 0$, quando $n \longrightarrow \infty$, ou seja:

$V_n(\Pi', u^*) = T_1' \ldots . T_n' u^* \longrightarrow T_1' \ldots T'_n u_n = V(\Pi')$ quando $n \longrightarrow \infty$. (2)

Porém:

$$V_n(\Pi', u^*) = T_1' \ldots .. T_n' u^* \leq U_n u^* \tag{3}$$

(vide observação que se segue ao Teorema 9.5), e portanto, (2) e (3) implicam:

$$V(\Pi') \leq u^* \tag{4}$$

Ora, pelo Teorema 9.5 (d), existe uma f que é Π-gerada, tal que $T u^* \geq u^* - \varepsilon' = u^* - \varepsilon'$, onde $\varepsilon' = \varepsilon(1 - \beta)$ e T é associada a f, então, por indução, obtemos:

$$T^n u^* \geq u^* - \varepsilon'(1 + \beta + \ldots + \beta^{n-1}), \quad n \geq 1 \tag{5}$$

Com efeito:

i) Vale para n = 1 (trata-se da própria desigualdade (4)):

ii) Valendo para n, vale para n+1:

$$T^n u^* \geq u^* - \varepsilon'(1 + \beta + \ldots\ldots + \beta^{n-1}) \longrightarrow T^{n+1}u^* = T(T^n u^*) \geq$$
$$\geq T(u^* - \varepsilon'(1 + \beta + \ldots\ldots + \beta^{n-1})) = Tu^* - \varepsilon'(1 + \beta + \ldots \beta^{n-1})$$
$$\geq u^* - \varepsilon' - \varepsilon'(\beta + \beta^2 + \ldots\ldots + \beta^N) = u^* - \varepsilon'(1+\beta+\beta^2+\ldots+\beta^n).$$

4º) por outro lado, para T associada a f, temos:

$$T^n u^* \longrightarrow V(f^{(\infty)}), \quad \text{quando } n \longrightarrow \infty \tag{6}$$

Então, de (5) e (6) se segue: $V(f^{(\infty)}) \geq u^* -$

$$- \varepsilon'(1 + \beta + \ldots\ldots + \beta^{n-1} + \ldots\ldots) = u^* - \frac{\varepsilon'}{1-\beta} = u^* - \frac{\varepsilon(1-\beta)}{1-\beta} = u^* - \varepsilon$$

isto é, $V(f^{(\infty)}) \geq u^* - \varepsilon$,

C.Q.D.

Dem. de (b)

Pelo Corolário 8.3, existe uma estratégia de Markov $(p,\varepsilon/2)$-ótima, que designaremos por $\Pi = (f_1, f_2, \ldots\ldots)$, donde:

$$V(\Pi)(s) \geq V(\Pi')(s) - \varepsilon, \ \forall\ \Pi' \text{ e quase todo } s \in \mathcal{S} \quad (1).$$

Por (a), existe uma estratégia de Markov estacionária $f^{(\infty)} \equiv (f, f, \ldots\ldots)$ tal que $V(f^{(\infty)}) \geq u^* - (\varepsilon/2)$, onde u^* é um ponto fixo do operador U associado a Π; por outro lado, conforme (a), se tem: $V(\Pi) \leq u^*$, donde se segue:

$$V(f^{(\infty)}) \geq V(\Pi) - \varepsilon/2 \tag{2}$$

de (1) e (2), vem: $V(f^{(\infty)})(s) \geq [V(\Pi')(s) - \varepsilon/2] - \varepsilon/2$, $\forall\ \Pi'$ e para quase todo $s \in S$, ou seja $p\{V(f^{(\infty)}) \geq V(\Pi) - \varepsilon\} = 1$, portanto $f^{(\infty)}$ é uma estratégia de Markov que é (p,ε)-ótima.

C.Q.D.

Dem. de (c)

1º) Para qualquer estratégia $\Pi^* = (\Pi_1, \Pi_2, \ldots\ldots)$ temos:

$$V(\Pi^*) = \Pi_1 \, q(r + \beta w) \tag{1}$$

onde $w \in M(\mathcal{S} A \mathcal{S})$ e $w(s,a,s') = V(\Pi_{s,a})(s')$, $\Pi_{s,a}$ denotando a estratégia construida a partir de Π^*, da maneira como se segue:

$\Pi_{s,a} = (\Pi_1', \Pi_2', \ldots)$ e $\Pi_n'(.|s_1, a_1, \ldots, s_n) = \Pi_{n+1}(.|s,a,s_1,\ldots s_n)$.

Com efeito, $\Pi_1 q(r + \beta w) = \Pi_1 qr + \beta\Pi_1 qw$ e $V(\Pi^*) =$

$$= \sum_{n=1}^{\infty} \beta^{n-1}\Pi_1 q \ldots\ldots \Pi_n qr = \Pi_1 qr + \sum_{n=2}^{\infty} \beta^{n-1} \Pi_1 q \ldots\ldots \Pi_n qr =$$

$$= \Pi_1 qr + \beta\Pi_1 q \sum_{n=1}^{\infty} \beta^{n-1} \Pi_2 q \ldots \Pi_{n+1} qr = \Pi_1 qr + \beta\Pi_1 q V(\Pi_{s,a}) =$$

$$= \Pi_1 qr + \beta\Pi_1 qw = \Pi_1 q(r, \beta w).$$

2º) Ora, se Π^* for ε-ótima, teremos então $V(\Pi_{s,a})(s') \leq V(\Pi^*)(s') + \varepsilon$, $\forall\, s' \in \mathcal{S}$, ou seja:

$$w(s,a,s') \leq V(\Pi^*)(s') + \varepsilon, \ \forall\, s' \in \mathcal{S} \tag{2}.$$

De (1) e (2), obtemos $V(\Pi^*) \leq \Pi_1 q(r + \beta(V(\Pi^*) + \varepsilon))$, e chamando $h = q(r + \beta V(\Pi^*) + \beta\varepsilon)$, tem-se $V(\Pi^*) \leq \Pi_1 h$; então pelo Lema 6.8, existe f tal que $fh \geq \Pi_1 h$, e portanto:

$$\begin{aligned} V(\Pi^*) \leq fh &= fq(r + \beta V(\Pi^*) + \beta\varepsilon) \\ &= fq(r + \beta V(\Pi^*) + \beta\varepsilon; \end{aligned} \tag{3}$$

lembrando que $TV(\Pi^*) = V(f,\Pi^*) = f\, qr + \beta V(\Pi^*)$, segue-se:

$$V(\Pi^*) \leq T(V(\Pi^*)) + \beta\varepsilon \tag{4},$$

e por indução, usando o mesmo procedimento pelo qual se obteve (5) de 9.7 (a), obtemos:

$$T^n(V(\Pi^*)) \geq V(\Pi^*) - \varepsilon(\beta + \ldots\ldots + \beta^n) \tag{5}$$

Ora, $T^n V(\Pi^*) \longrightarrow V(f^{(\infty)})$, quando $n \longrightarrow \infty$, e portanto, obtemos de (5):

$$V(f^{(\infty)}) \geq V(\Pi^*) - \varepsilon(\beta + \ldots + \beta^n + \ldots) =$$
$$= V(\Pi^*) - \frac{\beta\varepsilon}{1-\beta}\text{, isto é, } V(f^{(\infty)}) \geq V(\Pi^*) - \frac{\beta\varepsilon}{1-\beta}\text{, ou ainda:}$$

$$V(f^{(\infty)})(s) \geq V(\Pi^*)(s) - \frac{\beta\varepsilon}{1-\beta}, \forall s \in \mathcal{S}. \qquad (6).$$

Por outro lado, como Π^* é ε-ótima, isto é, $V(\Pi)(s) \leq V(\Pi^*)(s) + \varepsilon$, $\forall \Pi$, $\forall s \in \mathcal{S}$, segue-se de (6) que:

$V(f^{(\infty)})(s) \geq V(\Pi)(s) - \varepsilon - [\beta\varepsilon/(1-\beta)] = V(\Pi)(s) - [\varepsilon/1-\beta)]$, $\forall s \in \mathcal{S}$, $\forall \Pi$, ou seja, $f^{(\infty)}$ é $\varepsilon/(1-\beta)$-ótima.

C.Q.D.

Dem. de (d)

Para cada $s_0 \in \mathcal{S}$, e para cada $\varepsilon > 0$, existe $f^{(\infty)} = (f, f, \ldots)$, tal que:

$$V(\Pi)(s_0) \leq V(f^{(\infty)})(s_0) + \varepsilon, \forall \Pi \qquad (1);$$

para isto, basta escolher $f^{(\infty)}$ que seja (p,ε)-ótima, onde p é a probabilidade concentrada em s_0.

Ora, $T_a u \leq u$, $\forall a \in A$, implica $Tu \leq u$, para todo u, e em particular, para T associado a $f^{(*)}$, de onde se obtem $T^n u \leq T^{n-1}u$, $\forall n$, e portanto, $T^n u \downarrow V(f^{(\infty)})$ (isto é, $T^n u$ decresce para $V(f^{(\infty)})$, em particular:

$$(T^n u)(s_0) \downarrow V(f^{(\infty)})(s_0) \quad \therefore \quad u(s_0) \geq V(f^{(\infty)}(s_0) \qquad (2).$$

e substituindo-se em (1), vem:

$V(\Pi)(s_0) \leq u(s_0) + \varepsilon$, $\forall \Pi$, e fazendo $\varepsilon \longrightarrow 0$, obtem-se $V(\ )(s_0) \leq u(s_0)$, $\forall \Pi$.

(*) Sendo T associado a f, tem-se $tu = fq(r + \beta u)$, donde $(tu)(s) = (fq(r + \beta u))(s) = \int [r(s,f(s),s') + \beta u(s')]\, dq(s'|s,f(s))$; por outro lado, $(T_a\, u)(s) = \int r(s,a,s')\, dq(s'|s,a)$, e portanto $Tu(s) = (T_{f(s)}\, u)(s)$, $\forall s \in \mathcal{S}$. Desta forma,
$T_a\, u \leq u$, $\forall a \in A \longrightarrow T_{f(s)}\, u \leq u$, $\forall s \in \mathcal{S} \longrightarrow (T\, u)(s) = (T_{f(s)}\, u) \leq u(s)$, $\forall s \in \mathcal{S}$ isto é, $T\, u \leq u$.

Como s_o foi escolhido arbitrariamente, tem-se:

$$V(\Pi) \leq u, \forall \Pi.$$

C.Q.D.

Dem. de 6(e)

1º) Por (c), a hipótese implica que existe, para cada $n = 1, 2, \ldots.,$ uma estratégia $f_n^{(\infty)} = (f_n, f_n, \ldots..)$ que é $(1/n)$-ótima.

Tomemos $\Pi = (f_1, f_2, \ldots.., f_n, \ldots..)$; seja U o operador associado a Π (logo, $Uu = \sup T_n\, u$, onde T_n é associado a cada f_n); seja u^* o ponto fixo de U; então, por (a), segue-se que $V(\Pi') \leq u^*$, para cada estratégia Π', que seja Π-gerada.

Em particular $\Pi = (f_1, f_2, \ldots.)$ é Π-gerada, donde cada f_n é Π-gerada, e então, cada $f_u^{(\infty)} = (f_n, f_n, \ldots.)$ é π-gerada, e portanto, $V(f_n^{(\infty)}) \leq u^*$, $n = 1, 2, \ldots..$, isto é,

$$V(f_n^{(\infty)})(s) \leq u^*(s), \forall s \in \mathcal{S}, n = 1, 2, \ldots.. \quad (1);$$

como cada $V(f_n^{(\infty)})$ é $1/n$-ótima, temos ainda:

$$V(\Pi')(s) \leq V(f_n^{(\infty)})(s) + (1/n), \forall s \in \mathcal{S}, \forall \Pi, n = 1, 2, .. \quad (2)$$

onde Π' é agora qualquer estratégia Π'.

Ora, (1) e (2) implicam que $V(\Pi')(s) - (1/n) \leq u^*(s)$, $\forall s \in \mathcal{S}$, $n = 1, 2, \ldots..$, e fazendo $n \longrightarrow \infty$, segue-se: $V(\Pi')(s) \leq u^*(s)$ $\forall s \in \mathcal{S}$, ou seja, $V(\Pi') \leq u^*$, (assim, u^* é um ganho ótimo, em relação a todas as possíveis estratégias Π').

2º) Tem-se:

$$\sup_S T_a u^* \geq Uu^* = u^* \quad (3);$$

Lembrando-se que u^* é o ponto fixo de U, verificamos a desigualdade.

Com efeito: $(Uu^*)(s) = \sup_n [(T_n\, u^*)(s)]$ e $(T_n\, u^*)(s) =$

$= (T_{f_n}(s)\, u^*)(s)$, $\forall\, s \in \mathcal{S}$, donde $(Uu^*)(s) = \sup_{n}[(T_{f_n(s)}u^*)(s)] \leq$ $\leq \sup_{a}[(T_a u^*)(s)]$, $\forall\, s \in \mathcal{S}$ (esta última desigualdade decorre de, no lado direito, aparecerem todos os T_a, enquanto no lado esquerdo concorrem apenas os T_a tais que $a = f_n(s)$).

3º) Tem-se $u^* \leq V(f^{(\infty)}) + (1/n)$, $n = 1, 2, \ldots$, com efeito, se fôsse $u^*(s) > V(f_n^{(\infty)})(s) + (1/n)$, para algum índice n e algum $s \in$, ao tomarmos $\varepsilon = u^*(s) - V(f_n^{(\infty)})(s) - (1/n)$, obteriamos uma certa $f^{(\infty)}$ que seria $f^{(\infty)}$-gerada e tal que $V(f^{(\infty)})(s) \geq u^*(s) - \varepsilon = V(f_n^{(\infty)})$ $(s) + (1/n)$, conforme 9.7 (a); e portanto, uma contradição em relação ao fato de ser $f_n^{(\infty)}$ uma estratégia $(1/n)$-ótima.

Assim, $u^* \leq V(f^{(\infty)}) + (1/n)$, $n = 1, 2, \ldots$, implica $T_a u^* \leq$ $\leq T_a\,(Vf_n^{(\infty)} + (1/n)) = T_a(V(f_n^{(\infty)}) + (\beta/n) = V(a, f^{(\infty)}) + (\beta/n) \leq$ $\leq u^* + (\beta/n)$, $n = 1, 2, \ldots$, onde a última desiguâldade decorre do fato de ser u^* um ganho esperado ótimo, a igualdade que lhe antecede decorre de (c) e do fato de $f \equiv a$ ser associada a T_a, e finalmente, a outra igualdade é conforme (b); ou seja, se tem $T_a u^* \leq$ $\leq u^* + (\beta/n)$, $n = 1, 2, \ldots$, e fazendo $n \longrightarrow \infty$, fica $T_a u^* \leq u^*$, donde:

$$\sup_{s} T_a u^* \leq u^* \qquad (4).$$

Ora, (3) e (4) implicam: $\sup_{a} T_a\, u^* = u^=$.

<u>Dem. de (f)</u>

<u>1a.) parte</u> ($\longrightarrow$)

Se Π é ótimo, a hipótese de (e) é satisfeita, donde $u^* = V(\Pi)$ é um ganho ótimo e satisfaz, portanto, à equação de otimibilidade.

2a.) parte ($\longleftarrow$)

Suponhamos qu $V(\Pi)$ satisfaz à equação, ou seja que $V(\Pi) =$ $= \sup_{a} T_a(V(\Pi))$, donde $T_a(V(\Pi)) \leq V(\Pi)$, $\forall\, a \in A$, donde por (d), com

$u = V(\Pi)$, se obtem: $V(\Pi') \leq V(\Pi)$, $\forall\ \Pi'$, isto é, Π é ótimo..

C.Q.D.

Obsevações:

1. (d) é estremamente útil para se se provar a otimibilidade; com efeito, se sabemos que u é o ganho esperado de uma estratégia Π, tal que $T_a u \leq u$, $\forall\ a \in A$, então (d) garante que Π é ótimo.

2. Não sabemos se o ganho ótimo sempre satisfaz à equação de otimibilidade, ou mesmo se sob as hipótese de (e), a solução (limitada) é única.

§ 10. RESULTADOS PARTICULARES

Se A é enumerável, com elementos a_1, a_2, ..., cada estratégia de Markov é Π^*-gerada, com $\Pi^* = (g_1, g_2, \ldots)$ e $g_n \equiv a_n$, ou seja, cada $\Pi = (f_1, f_2, \ldots)$ é tal que f_k é Π^*-gerada, $k = 1, 2, \ldots$, de fato, fixado o índice k, seja, $S_n = \{s; f_k(s) = a_n\}$ onde $S = S_1 + S_2 + \ldots + S_n + \ldots$ e $f_k(s) = g_n(s)$ $\forall$ $s \in S_n$, isto é, $f_k | S_n = g_n$.

Reciprocamente, para cada estratégia $\Pi = (f_1, f_2, \ldots)$, o estudo de estratégias Π-geradas pode reduzir-se ao caso em que A é enumerável; com efeito, se interpretarmos "ação n" no estado s, como sendo a seleção de $f_n(s)$.

Contudo, é preferível conservar o conjunto de ações A original e introduzir o conceito de enumerabilidade essencial.

Definição 10.2

Duas ações a e b dizem-se equivalentes no estado s quando $r(s,a,.) = r(s,b,)$ e $q(.|s,a) = q(.|s,b)$, ou seja, $r(s,a,s') = r(s,b,s')$ e $q(s'|s,a) = q(s'|s,b)$, $\forall$ $s' \in s$, ou ainda:

$$T_a u(s) = T_b u(s) \quad , \forall \ u \in M(s).$$

Definição 10.3

(i) Se $\Pi = (f_1, f_2, \ldots)$ é de Markov, A diz-se essencialmente enumerável por Π quando para cada (s,a) existir um n tal que $f_n(s)$ é equivalente a a, no estado s.

(ii) A diz-se essencialmente finito por Π quando existe uma partição de S em conjuntos de Borel S_1, S_2, ... tais que, para cada pra (s,a) com $s \in S_n$, então pelo menos uma das ações $f_1(s), \ldots f_n(s_n)$ é equivalente a $\underline{a}$ em $\underline{s}$.

Teorema 10.4

(a) Se A é essencialmente enumerável por $\Pi = (f_1, f_2, \ldots.)$ então o ponto fixo u^* do operador U associado a Π, é o ganho ótimo. U é idêntico ao operador $\sup_a T_a$, de sorte que u^* é a única solução (limitada) da equação de otimibilidade. Para cada $\varepsilon > 0$, há uma única estratégia estacionária que é ε-ótima.

(b) Se A é essencialmente finito por $\Pi = (f_1, f_2, \ldots)$, então existe uma estratégia estacionária ótima.

Dem. de (a)

Para u e s, tem-se $T_n u(s) = T_a u(s)$, onde $a = f_n(s)$; então:
$Uu(s) = \sup_n T_n u(s) = \sup_n T_{f_n(s)} u(s) \leq \sup_a T_a u(s)$, $\forall\, s \in S$, isto é,

$$Uu \leq \sup_a T_a u \qquad (1).$$

Porém, para cada $a \in A$, $T_a u(s) = T_n u(s)$, para algum $\underline{n}$ donde $T_a u(s) \leq U\, u\,(s)$, e portanto, $\sup_a T_a u(s) \leq Uu(s)$, isto é,

$$\sup_a T_a u \leq Uu \qquad (2);$$

então, (1) e (2) implicam $U\, u = \sup_a T_a\, u$.

Ora, o Teorema 9.7 (d) garante que $V(\Pi) \leq u^*$, $\forall\, \Pi$, e pelo Teorema 9.7 (a), existe uma estratégia estacionária $f^{(\infty)} = (f, f, \ldots.)$, com $V(f^{(\infty)}) \geq u^* - \varepsilon$; então, $V(f^{(\infty)}) \geq V(\pi) - \varepsilon$, $\forall\, \Pi$, ou seja, $f^{(\infty)}$ é ε-ótima.

Dem. de (b)

Se A fôr essencialmente finito por Π, definimos:

$B_n = \{s;$ n é o menor $\underline{i}$ tal que $T_i u^*(s) = \sup T_n u^*(s)\}$; observemos que a sequência $\{T_n u^*(s)\}$ contém apenas um número finito de valores distintos.

Definimos $f = f_n$ em B_n, tal que $Tu^* = Uu^* = u^*$ onde T é associada com f. Então μ^*, como ponto fixo de T, é o ganho de $f^{(\infty)}$, e $f^{(\infty)}$ é ótima.

A seguir são apresentadas extensões de rotinas para melhoramento de estratégias ("improvement routines"), já conhecidas no caso de $\mathcal{S}$ e A finitos.

<u>Teorema 10.5</u>

(a) ("Howard improvement") Se $V(g,\Pi) \geq V(\Pi)$, então $V(g^{(\infty)}) \geq$ $\geq V(g,\Pi) \geq V(\Pi)$.

(b) ("Eaton-Zadeh improvement") [14]. Dadas $f, g: \mathcal{S} \longrightarrow A$, definimos $h(s) = f(s)$, quando $V(f^{(\infty)})(s) \geq V(g^{(\infty)})(s)$ e $h(s) =$ $= g(s)$, quando $V(g^{(\infty)})(s) > V(f^{(\infty)})(s)$; então, $V(h^{(\infty)}) \geq$ mas $\{V(f^{(\infty)}), V(g^{(\infty)})\}$.

<u>Dem. de (a)</u>

Se T é associado a g, temos $T(V(\Pi)) = V(g(\Pi) \geq V(\Pi)$, donde, $T^n(V(\Pi)) \geq T^{n-1}(V(\Pi))$, $\forall$ n, donde $T^n(V(\Pi)) \uparrow V(g^{(\infty)})$ e $V(g^{(\infty)}) \geq V(g,\Pi)$.

<u>Dem. de (b)</u>.

h está bem definida, pois $\mathcal{S} = S' + S''$, onde $S' = \{s; V(f^{(\infty)})(s) \geq V(g^{(\infty)})(s)\}$ e $S'' = \{s; V(g^{(\infty)})(s) > V(f^{(\infty)})(s)\}$.

Consideremos T_1, T_2 e T associados a f, g, e h, respectivamente; temos:

$$(T\,u)(s) = (T_1 u)(s),\ \forall\ s \in S'$$

$$(T\,u)(s) = (T_2 u)(s),\ \forall\ s \in S''$$

Com $u = mas\{V(f^{(\infty)}), V(g^{(\infty)})\}$, obtemos $(Tu)(s) =)T_1u)(s) = (T_1\, mas\{V(f^{(\infty)}), V(g^{(\infty)})\})(s) \geq T_1V(f^{(\infty)})(s) = V(f^{(\infty)}) = \max\{V(f^{(\infty)})(s), V(g^{(\infty)})(s)\} = u(s)$, isto é $(Tu)(s) \geq u(s)$, se $s \in S'$; analogamente, obtemos $(Tu)(s) = (T_2u)(s) \geq (T_2V(g^{(\infty)})(s) = (V(g^{(\infty)})(s) = u(s)$, isto é, $(Tu)(s) \geq u(s)$, para $s \in S''$.

Então, uma vez que $\mathcal{S} = S' + S''$, obtemos $(T\,u)(s) \geq u(s)$, $\forall\, s \in \mathcal{S}$ isto é, $T\,u \geq u$, donde $V(h^{(\infty)}) \geq u$, ou seja, $V(h^{(\infty)}) \geq \max\{V(f^{(\infty)}), V(g^{(\infty)})\}$.

C.Q.D.

Faremos, aqui, a verificação da <u>Proposição 1.6</u>: Se Q é uma matriz de Markov, então existe outra matriz de Markov Q^*, tal que $Q^* = \lim_{n \to \infty} (Q + \ldots + Q^n) / n$.

Algumas definições e resultados que serão necessários para esta verificação, são indicados a seguir, com a referência da página do Livro de Gantmacher (4), onde poderão ser encontrados.

Uma matriz quadrada A diz-se uma <u>matriz redutível</u> se existe uma mesma permutação dos índices i, j das linhas e colunas, permitindo que ela passe a se escrever sob a forma:

$$\tilde{A} = \begin{bmatrix} B & 0 \\ & \\ C & D \end{bmatrix}$$

onde os blocos B e D são matrizes quadradas e 0 é um bloco nulo; c contrário, diz-se <u>irredutível</u> (pág. 50).

Uma matriz quadrada A diz-se sob <u>forma normal</u> quando se esc

$$A = \begin{bmatrix} A_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_g & 0 & \cdots & 0 \\ A_{g+1,1} & A_{g+1,2} & \cdots & A_{g+1,g} & A_{g+1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ A_{p,1} & A_{p,2} & \cdots & A_{p,g} & A_{p,g+1} & \cdots & A_p \end{bmatrix}$$

onde os blocos A_i (i = 1, a, ..., p) da "diagonal" são matrizes das irredutíveis, e em cada linha $f \geq g+1$, pelo menos uma das matr $A_{f,1}, \ldots, A_{f,f-1}$ é não nula. Toda matriz pode ser posta sob a forma normal, com uma conveniente permutação de linhas e colunas (pp.74-77).

Se uma matriz $A \geq 0$ (isto é, com termos todos não negativos) é irredutível e seus autovalores que têm módulo máximo r são $\lambda_1, \ldots, \lambda_h$

(no caso de uma matriz de Markov, o módulo máximo é 1, conforme Prop. 1.5, § 1), então: A diz-se primitiva quando h = 1; imprimitiva, quando h > 1 (e o "índice de imprimitividade" da matriz é o inteiro h > 1) (pág. 80).

Resultados importantes são os seguintes: 1º) Se $A \geq 0$ é uma matriz imprimitiva, com índice de imprimitividade igual a h, então A^h tem h autovalores iguais ao módulo máximo r; assim, se A é uma matriz de Markov, então possui h autovalores iguais a 1 (os restantes com módulo inferior a 1)(pág. 82). 2º) Se uma matriz de Markov A é escrita sob a forma normal, então os autovalores de módulo máximo 1, ocorrem nos blocos isolados $A_1, \ldots, A_g$ (isto é, blocos sôbre a "diagonal", que são precedidos somente por blocos nulos) (pág. 84-85).

Uma matriz de Markov A diz-se regular quando todos os autovalores, de módulo máximo, são iguais a 1. Outro resultado importante é o seguinte: 3º) Se A é uma matriz de Markov, regular, então existe uma matriz de Markov $A^{\infty} = \lim_{q \to \infty} A^q$ (pp 88-89).

Agora, estamos em condições de provar a Proposição 1.6.

Dem.

1º) Supponhamos que a matriz de Markov A esteja escrita sob a forma normal, onde $A_1, \ldots, A_g$ são os "blocos isolados", com índices de imprimitividade $h_1, \ldots, h_g$ (se alguma dessas matrizes fôr primitiva, tomaremos 1 como índice de imprimitividade). Seja h o mínimo múltiplo comum de $h_1, \ldots, h_g$. Em consequência, cada $(A_i)^h = (A_i^{h_i})^{\alpha_i}$ tem h_i autovalores iguais a 1; então, a matriz Q de $(A_1)^h, \ldots, (A_g)^h$ comparecem como blocos isolados sôbre a diagonal", possui $h_1 + \ldots + h_g$ autovalores iguais a 1, sendo todos os restantes autovalores de módulo inferior a 1, e assim, Q^h é matriz regular, e existe $(Q^h)^{\infty} = \lim_{q \to \infty} Q^{hq}$; conforme (1º), (2º) e

2º) Notaremos: $P_n = \sum_{k=0}^{n} Q^h / (n+1)$; $n \geq 0$; assim, desejamos provar existe $\lim_{n \to \infty} P_n$.

Observe-se que a sequência $(P_n)_{n\geq 0}$ pode ser decomposta em h sequências, por partição dos índices n = 0, 1, 2, ..., a saber:

$$(P_q^{(r)})_{q\geq 0} = (P_{qh+r})_{q\geq 0}\ ,\ r = 0,\ 1,\ \ldots,\ h-1;$$

verificaremos que tôdas estas subsequências possuem um mesmo limite, e portanto ficará provado que a sequência $(P_n)_{n\geq 0}$ também possui êste limite(*).

3º) Inicialmente, mostraremos que a subsequência $(P_q^{(h-1)})_{q\geq o}$; possui um limite; tem-se:

$$P_q^{(h-1)} = P_{qh+(h-1)} = \sum_{k=0}^{qh+(h-1)} Q^k/)qh+h) =$$

$$= \frac{(I + Q + \ldots + Q^{h-1})(I + Q^h + \ldots + Q^{qh}}{qh + h} =$$

$$= \frac{I + Q + \ldots + Q^{h-1}}{h} \cdot \frac{I + Q^h \ldots + Q^{gh}}{q + 1} \qquad (1)$$

Assim, a existência de $\lim_{q\to\infty} P_q^{(h-1)}$ dependerá unicamente, da existência de $\lim_{q\to\infty} (I+Q^h+\ldots+Q^{qh}) / (q+1)$, ora, tal limite existe e é igual a $(Q^h)^\infty$, como será verificado adiante, e em consequência: $\lim_{q\to\infty} P_q^{(h-1)} =$

$$= (1/h)(I+Q+\ldots+Q^{h-1})(Q^h)^\infty \qquad (2)$$

Agora, prova-se que, de fato, toda outra sequência $(P_q^{(r)})_{q\geq 0}$, r = = 0,1,...,h-2, possui o mesmo limite em (2); com efeito:

$$P_q^{(r)} = \sum_{k=0}^{qh+r} Q^h / (qh+r+1) = \left[\sum_{k=0}^{qh+h+1} Q^h / (qh+h)\right] \left[(qh+h) / (qh+r+s)\right] -$$

$$- \left[(Q^{qh+r+1} + \ldots + Q^{qh+h+1}) / (h-r-1)\right] \times \left[(h-r-1) / (qh+r+1)\right] \quad (3)$$

como existem:

(*) (Raciocinemos em termos das (i,j)-componentes das matrizes; com efeito, fixado (i,j), sejam $P_{ij}^{(qh+r)}$ a (i,j)-componente de $P_q^{(n)} = P_{qh+r}$, e suponhamos que existam os limites $\lim_{q\to\infty} P_{ij}^{(qh+r)}$, r=0,1,...,h-1, todos iguais P_{ij}; isto é, dado $\varepsilon>o$, existem índices n_r, tais que $|P_{ij}^{(qh+r)}-P_{ij}|<\varepsilon$, $\forall\ q\geq n_r$, r=0,1,...,h-1. Segue-se $|P_{ij}^{(qh+r)}-P_{ij}|<\varepsilon$, $\forall\ q\geq \max\{n_0,n_1,\ldots n_{h-1}$ $\forall$ r=0,1,...,h-1, donde $|P_{ij}^{(n)}-P_{ij}|<\varepsilon$, $\forall\ n\geq h \max\{n_0,n_1,\ldots,n_{h-1}\}+\max\{0,1,\ldots,h-1\}$, em consequência, $\forall\ n\geq h[\max\{n_0,n_1,\ldots,n_{h-1}\}+1]$ ou seja, $\exists \lim_{n\to\infty} P_{ij}^{(n)} = P_{ij}$, onde $P_{ij}^{(n)}$ é a (i,j)-componente da matriz P_n.

$$\lim_{q\to\infty} (qh+h) / (qh + r + 1) = 1$$

$$\lim_{q\to\infty} (h - r - 1) / (qh + r + 1) = 0 \quad \text{e}$$

$$\lim_{q\to\infty} (Q^{qh+r+1} + \ldots + Q^{qh+h+1}) / (h - r - 1) =$$

$$= \lim_{q\to\infty} Q^{qh}(Q^{r+1} + \ldots + Q^{h-1}) / (h - r - 1) =$$

$= (Q^h)^\infty(Q^{r+1} + \ldots + Q^{h-1}) / (h - r - 1)$, segue-se que existe o limite da expressão à direita, em (3), a saber:

$\lim_{q\to\infty} P_q^{(h-1)} = \lim_{q\to\infty} \sum_{k=0}^{qh+h-1} Q^k/(qh+h)$, e portanto:

$$\lim_{q\to\infty} P_q^{(r)} = \lim_{q\to\infty} P_q^{(h-1)} \quad , r = 0, 1, \ldots, h-2 \tag{4}$$

De (2) e (4), concluimos que:

$$\exists \lim_{n\to\infty} P_n = (1/h)(I + Q + \ldots + Q^{h-1})(Q^h)^\infty = Q^*$$

4º) É fácil ver que Q^* é uma matriz estocástica; com efeito, as P_n o são (conforme Proposição 1.3), donde se q^*_{ij}, é a(i,j) - componente de Q^* e $P^{(n)}_{i,j}$ a (i,j)-componente de P_n (i,j = 1,2,...,S), segue-se:

$$\sum_{j=1}^{S} q^*_{ij} = \sum_{j=1}^{S} \lim_{n\to\infty} P^{(n)}_{i,j} = \lim_{n\to\infty} \sum_{j=1}^{S} P^{(n)}_{i,j} = \lim_{n\to\infty} 1 = 1,$$

para cada i = 1, 2, ..., S.

5º) Para completar a demonstração, falta verificar que $\lim_{q\to\infty} (I+Q+\ldots+Q^{qh}) / (q+1) = (Q^h)^\infty$; mas isto é fácil, pois existinto $\lim_{q\to\infty} (Q^h)^q = (Q^h)^\infty$ (uma vea que Q^h é regular), segue-se (é um resultado clássico) que a sequência de médias aritméticas converge para o mesmo limite(*).

(*) Remetemos a Knopp (7) (pp.72).

A P Ê N D I C E II

Fazemos, nêste apêndice, a verificação do:

Lema 6.8 - Dados $q \in Q(Y \mid X)$, $u \in M(XY)$ e $\varepsilon > 0$, existe $f \in Q(Y \mid X)$, degenerada, tal que:

(1) $$f\,u \geq q\,u\,;$$

(2) $$q(\{y\ ;\ u(x,y) \geq u(x_o, f(x_o)) + \varepsilon\} \mid x_o) = 0,\ \forall\, x_o \in X.$$

Antes de iniciar essa verificação, recordemos que $qu(x) = \int u(x,y)\, dq(y \mid x)$; no caso de $f \in Q(Y \mid X)$ degenerada, pode se identificar a uma função de Baire $f: X \longrightarrow Y$, de sorte que: $f\,u\,(x) = u(x, f(x))$.

Dem. de (1)

Consideremos o conjunto $D = \{(x,y) \in XY\ ;\ u(x,y) \geq qu(x)\}$; para cada $x \in X$, seja $D_x = \{y\ ;\ (x,y) \in D\}$ afirmamos que sempre se tem $q(D_x \mid x) > 0$.

Com efeito, suponhamos que se tivesse $q(D_x \mid x) = 0$, isto é. $q(D_x^c \mid x) = 1$, para algum x; observemos que, para $y \in D_x^c$, se tem $u(x,y) < qu(x)$. Em consequência, obteríamos:

$$qu(x) = \int u(x, y)\, dq \mid x) = \int_{D_x^c} u(x, y)\, dq(y \mid x) <$$

$$< \int_{D_x^c} qu(x)\, dq(y \mid x) = \int qu(x)\, dq(y \mid x) = qu(s)$$

isto é, $qu(x) < qu(x)$, contradição:

Então, por um resultado de Blackwell e Ryll-Nardzewski[(a)], existe uma função de Baire $f: X \longrightarrow Y$, cujo gráfico é um subconjunto de D, e portanto,

$$u(x, f(x) \geq qu\,(x),\ \forall\, c \in X,$$

que é precisamente a condição:

(1) $$f\,u \geq q\,u\,.$$

Demo. de (2)

Procede-se como na demonstração de (1), só que em vez de D, usa-se $D_1 = \{(x,y) \in XY\ ;\ u(x,y) > S(x) - \varepsilon/2\}$ onde $S(x)$ é o "supremo condicional essencial de $u(x,y)$", isto é,:

$$S(x) = \sup\{r \in Q;\ q(\{y\ ;\ u(x,y) > r\}|x) > 0\}$$

(Q = conjunto dos números racionais).

Temos: $D_{1,x} = \{y\ ;\ (x,y) \in D_1\} = \{y\ ;\ u(x,y) > S(x) - \varepsilon/2\}$, e pela definição de $S(x)$, $q(D_{1,x}\ |\ x) > 0$, $\forall\, x \in X$. Assim podemos escolher f cujo gráfico esteja em D_1, e obtemos:

$$u(x, f(x)) > S(x) - \varepsilon/2, \quad \forall\, x \in X \tag{A}$$

por outro lado, ainda utilizando a definição de $S(x)$, se tem:

$$q(\{y;\ u(x,y) > S(x) + \varepsilon/2\}|x\ ) = 0\ , \qquad \forall\, x \in X,$$

$$q(\{y;\ u(x,y) \leq S(x) + \varepsilon/2\}|x\ ) = 1\ , \qquad \forall\, x \in X,$$

ou seja ainda, fixado x:

$S(x) \geq u(x,y) - \varepsilon/2$, para quase todo $y \in Y$, relativamente a $q(.|x)$ (B);

(A) e (B) então fornecem:

$u(x, f(x)) > u(x,y) - \varepsilon$, para quase todo $y \in Y$, relativamente a $q(.|x)$,

ou seja,

$q(\{y\ ;\ u(x,f(x)) > u(x,y) - \varepsilon\}|x) = 1$, isto é, vale:

(2) $\quad q(\{y\ ;\ u(x,y) \geq u(x,f(x)) + \varepsilon\}|x) = 0,\ \forall\ x \in X.$

Partindo de $D \cap D_1$ podemos construir a função f satisfazendo as mesmas condições.

(*) O resultado utilizado é o Teorema 2 contido no artigo de Blackwell, D & Ryll-Nardzewski, C.; "Non-existence of everywhere proper condicional distributions", Ann. Math. Statist., 34 (1963), p.p. 223-225, cujo enunciado (devidamente adaptado à nossa notação)é o seguinte:

"Dados:

X e Y, conjuntos de Borel de um espaço métrico completo e separável; $\mathcal{A} = \mathcal{B}(X)$ e $\mathcal{B} = \mathcal{B}(Y)$, as σ-álgebras correspondentes;

$\psi: X \times \quad \longrightarrow R$, tal que: a) $\psi(x,.) \in P(X),\ \forall\ x \in X$;

b) $\psi(.,A)$ é uma função de Baire, para cada $A \in \mathcal{A}$;

$D \in \mathcal{A} \times \mathcal{B}$, tal que: $\psi(x,D_x) > 0,\ \forall\ x \in X$, onde D_x é a x-secção de D, isto é, $D_x = \{y;\ (x,y) \in D\}$; então existe uma função de Baire $f: X \longrightarrow Y$, cujo gráfico é um subconjunto de D, isto é, tal que $(x,\ f(x)) \in D,\ \forall\ x \in X$."

As linhas da prova do Lema 6.8, seguidas por nós, são as que se encontram indicadas no artigo de D. Blackwell, "Memoryless strategies in finite-stage dynamic programming", Ann. Math. Stat., 35 (1964), p.p. 863-865.

B I B L I O G R A F I A

1. Blackwell, D. Discrete Dynamic Programming, Ann. Math. Statis., 33, 719-726, (1962).

2. Blackwell, D, Discountes Dynamic Programming, Ann. Math. Statis., 35, 863-865 (1964).

3. Belmann, R. Introduction to Matrix Analysis, Mc Graw-Hill Book Co., 1960.

4. Gantmacher, F.R. The Theory of matrices, vol. 2, Chelsea Publ. Co. N.Y., 1959.

5. Kemeny & Snell, Finite Markov Chains, P.Van Nostrand Company Inc. 1960.

6. Howard R. Dynamic Programming and Markov Processes, Technology Press and Wiley, N.Y., 1960.

7. Knopp, K. Theory and application of infinite series, Harfner Publ. Co. New York.

8. Bartle, R.G., The Elements of Integration, John Wiley & Sons, 1966.

9. Neveu, J., Bases Mathématiques du Calcul des Probabilités, Masson et Cie., Ed., Paris, 1964.

10. Hennequin, P.I. & Tortrat, A., Théorie des Probabilités et quelques Applications, Masson & Cie., Éd., Paris, 1965.

11. Blackwell, D. Memoriless Strategies in finite stage dynamic programming, Ann. Math, Statist., 35, 863-865 (1964).

12. Blackwell, D. and Ryll-Nardzewski, C. Non-existence of everywhere proper conditional distribution, Ann Math. Statis.,34 223-225, (1963).

ÍNDICE

§ 0 - Introdução. 1

§ 1 - Matrizes de Markov. 4

§ 2 - Caso discreto - Preliminares. 13

§ 3 - Caso discreto - Estratégias ótimas no caso $\beta < 1$. 18

§ 4 - Aplicações 23

§ 5 - Caso discreto - Estratégias ótimas para $\beta = 1$. 28

§ 6 - Conjunto de Borel e funções mensuráveis a Borel. 52

§ 7 - Caso geral - Preliminares. 57

§ 8 - Caso geral - Existência de estratégias (p,ε)-ótimas e de estratégias não aleatorizadas (p,ε)-dominantes 62

§ 9 - Estratégias estacionárias e operadores . .. 70

§10 - Resultados particulares. 84

Apêndice I 88

Apêndice II. 92

Bibliografia 95

$\varepsilon \phi \varepsilon$