Journal of the KSQC
Vol. 11, No. 2, 1983

Markov過程의 数理的 構造와 그 逐次決定過程
—On The Mathematical Structure of Markov Process and Markovian Sequential Decision Process—

金 裕 松*

ABSTRACT

As will be seen, this paper is tries that the research on the mathematical structure of Markov process and Markovian sequential decision process (the policy improvement iteration method,) moreover, that it analyze the logic and the characteristic of behavior of mathematical model of Markov process.

Therefore firstly, it classify, on research of mathematical structure of Markov process, the forward equation and backward equation of Chapman-kolmogorov equation and of kolmogorov differential equation, and then have survey on logic of equation systems or on the question of uniqueness and existence of solution of the equation.

Secondly, it classify, at the Markovian sequential decision process, the case of discrete time parameter and the continuous time parameter, and then it explore the logic system of characteristic of the behavior, the value determination operation and the policy improvement routine.

序 說

대저 이 논문에서는 Markov過程에 관하여 經營工學의 입장에서 첫째로, Markov過程의 論理와 作動特性에 따르는 數理的 構築과 Markov 型 逐次決定過程에 관한 理論的 모델을 구명하고자하였다. 무릇 Markov過程에 관하이는 A.N.Kolmogorov의 독창적인 方程式體系와 W.Feller의 확장된 數學的 定式化와 R.A.Howard의 엄밀한 Markov 모델의 연구가 이루어 졌다. 둘째로, Markov 型 逐次決定過程에 관하이는 R.Bellman과 S.E.Dreyfus 등에 의한 탁월한 數理的 定着이 이루어지고 있다.

이렇듯 Markov過程에 관한 理論的·基礎的 研究의 學際的 潮流(interdisciplinary stream)에 당면하여 經營科學(management sceience)의 입장에서 Markov過程에 관한 論理的 構築시스템을 분석하는 일은, Markov過程의 理論的·基礎的 研究에 있어서의 기본과제이며, 아울러 이것은 Markov過程과 그 逐次決定過程의 應用的·實證的 研究에 앞서는 必然的 當面課題라고 생각된다.

* 東國大學校 産業工學科 敎授

Ⅰ. Chapman-Kolmogorov 方程式

Markov過程의 運動方程式 -

Markov(A.A.Markov, 1856-1922)過程의 數理的 定式化를 한 것은 A.N.Kolmogorov[1]이며 그의 이론을 구명하고 확장함에 있어서 W. Feller[2], R.A.Howard[3] 등이 크게 기여하였다. 대저 Markov過程은 "시점 n에서 상태 i에 있었을때, 시점 $n+1$에서 상태 j로 이행하는 確率은, 시점 $n-1$ 이전에 어떠한 상태에 있었는가에 대하여는 無關係이다"라는 Markov性을 갖는 確率過程이다.

즉, 시점 $t_1, t_2, \cdots, t_{n-1}$에서 條件附確率 $X(tn)$은 시점 tn에 가장 가까운 시점 t_{n-1} 이전의 시점에 있어서의 $X(t)$에 의존하지 않는다는 관계를 표현하는 것이 Markov過程으로서 다음과 같이 표현한다. 따라서 Markov過程은 확율과정 $\{X(t); t \geq 0\}$은 어떤 시점 $t_1 < t_2 < \cdots < t_{n-1} < t_n$과 상태 $x_1, x_2, \cdots, x_{n-1}, x_n$에 대하여

$$P(X(t_n) \leq x_n \mid X(t_1) = x_1, \cdots, X(t_{n-1}) = x_{n-1}) = P(X(t_n) \leq x_n \mid X(t_{n-1}) = x_{n-1})$$

로 표현된다.

그리하여 지금 시점 m에서의 상태 i를 S_{mi}을 조건으로 하는, 시점 n에서의 상태 j에 관한 條件附確率을 $P\{S_{nj} \mid S_{mi}\}$와, 출발시점 n에서의 상태 j와 시점 r에서의 상태 k와의 결합된 확율 $P\{S_{nj} S_{rk} \mid S_{mi}\}$와의 관계를

$$(\text{I}.1.1)\quad P\{S_{nj} \mid S_{mi}\} = \sum_{k=1}^{n} P\{S_{nj} S_{rk} \mid S_{mi}\} \qquad (m \leq r \leq n)$$

로 표현한다.[4] 이것을 조건부확율에 의하여

$$(\text{I}.1.2)\quad P\{S_{nj} \mid S_{mi}\} = \sum_{k=1}^{N} P\{S_{rk} \mid S_{ni}\} P\{S_{nj} \mid S_{ik} S_{mi}\}$$

로 표현한다.

그런데 Markov過程에 있어서는 시점 r에서 상태 k와 시점 m에서 상태 j에 있는 조건부확율 $P\{S_{nj} \mid S_{rk} S_{mi}\}$는 오직 현재 시점 r에서 상태 k에만 의존하므로

$$(\text{I}.1.3)\quad P\{S_{nj} \mid S_{rk} S_{mi}\} = P\{S_{nj} \mid S_{rk}\} \qquad (m \leq r \leq n)$$

로 된다. 따라서 (I.1.2)는

$$(\text{I}.1.4)\quad P\{S_{nj} \mid S_{mi}\} = \sum_{k=1}^{N} P\{S_{rk} \mid S_{mi}\} P\{S_{nj} \mid S_{rk}\} \qquad (m \leq r \leq m)$$

로 변형된다.

지금 시점 m에서 시점 n에의 推移確率[5]을 $\phi_{ij}(m, n)$으로 하면

$$(\text{I}.1.5)\quad \phi_{ij}(m, n) = P\{S_{nj} \mid S_{mi}\}$$

로 된다.[6] 이것은 초기시점 0을 포함하는 추이확율과 관계되므로

$$(\text{I}.1.6)\quad \phi_{ij}(n) = \phi_{ij}(0, n) = P\{S_{nj} \mid 0_i\}$$

로 된다. 따라서 또 (I.1.4)는

$$(\text{I}.1.7)\quad \phi_{ij}(m, n) = \sum_{k=1}^{N} \phi_{ik}(m, r) \phi_{kj}(r, n)$$

$(i = 1, 2, \cdots, N,\ j = 1, 2, \cdots, N,\ m \leq r \leq n)$으로 변형된다.[7] (I.1.7)은 어떤 시간구간에서의 離散型의 추이확율을 표현하는 Chapman-Kolmogorov (C-K) 方程式이다.[8]

여기에서 C-K前向方程式에 의하여 언급하고자 한다. 그런데 추이확율에 있어서 출발시점 t_{k-1}과 도착시점 t_{k+1}의 두 가지중 도착시점을 Δt 만큼 연장시킴으로서 시간에 관한 미분방정식을 구성하는 경우가 前向方程式[9](forward equation)이며, 반대로 출발시점을 Δt 만큼 연장시킴으로써 시간에 관한 미분방정식을 구성하는 경우가 後向方程式[10](backward equation)이다.

그리하여 지금 (I.1.8)에서 시점 m과 n 사이의 어떤 r의 값을 고를 수 있으므로 (I.1.8)에 의하여

$$(\text{I}.1.9)\quad \Phi(m, n) = \Phi(m, n-1) \Phi(n-1, n) \qquad (m \leq n-1)$$

을 얻는다. 또는 行列의 모양으로

$$(\text{I}.1.10)\quad \Phi(m, n) = \Phi(m, n-1) P(n-1) \qquad (m \leq n-1)$$

로 표현할 수 있다. (I.1.10)은 離散型 Markov過程에 있어서 시간구간 (m, n)에 관한 C-K前

---

1) Vgl. A.N. Kolmogorov, Grundbegriffe der Wahrscheilichkeits rechnung, Springer, 1933.

2) W. Feller, An Introduction to Prbability Theory. and Its Application, Vol. Ⅱ, Wiley, New York, 1891.

3) R.A. Howard, op. cit., P. 511.

4) R.A. Howard, op. cit., p. 512.

5) R.A. Howard, op. cit., pp. 511-512.

6) Cf. ditto. p. 513.

7) W. Feller, op. cit., Vol. I. p. 42.

向方程式이다.

다시금 微分作用素(difference operator)를 △로 표시하면

$$\triangle_n f(n) = f(n+1) - f(n)$$

을 얻으며, 또 이것을

$$(\text{I}.1.12)\quad \triangle_n \Phi(m, n-1) = \Phi(m, n) - \Phi(m, n-1)$$

로 표현할 수 있다. 여기에서 微分行列을 $D(n)$으로 표시하여

$$(\text{I}.1.14)\quad D(n) = P(n-1) - 1$$

로 할때 (I.1.13)은 다음과 같은

$$(\text{I}.1.15)\quad \triangle_n \Phi(m, n-1) = \Phi(m, n-1)\, D(n-1) \qquad (m \leq n-1)$$

의 모양으로 변형할 수 있다. (I. 1. 15)는 離散型 Markov過程에 있어서 미분방정식으로 표현된 C−K前向方程式[10]이다.

이제 다시금 C−K前向方程式에 대하여 논리를 전개하고자 한다. 즉, (I.1.15)의 r를 $r=m+1$로 할때

$$(\text{I}.1.16)\quad \Phi(m, n) = \Phi(m, m+1)\ \Phi(m+1, n) \qquad (m \leq n-1)$$

또는

$$(\text{I}.1.17)\quad \Phi(m, n) = P(m)\ \Phi(m+1, m) \qquad (m \leq n-1)$$

을 얻는다. (I.I.16) 또는 (I.1.17)은 시간구간 (m, n)에 있어서 표현한 것으로서 離散型 Markov過程에 있어서 시간구간(m, n)에 관한 C−K後向方程式이다.[11]

여기에서 또 미분작용소를 △m으로 표시한다면

$$(\text{I}.1.18)\quad \triangle_m f(m) = f(m+1) - f(m)$$

을 얻으며, 또 이것을

$$(\text{I}.1.19)\quad \triangle_m \Phi(m, n) = \Phi(m+1, n) - \Phi(m, n)$$

으로 표현할 수 있다.

그러므로 (I.1.19)는 (I.1.17)에 의하여

$$(\text{I}.1.20)\quad \triangle_m \Phi(m, n) = \Phi(m+1, n) - P(m)\Phi(m+1, n) = [\,I - P(m)\,]\,\Phi(m+1, n) = D(m)\,\Phi(m+1, n) \qquad (m \leq n-1)$$

으로 표현할 수 있다. (I.1.20)은 離散型 Markov過程에 있어서 미분방정식으로 표현한 C−K後向方程式[12]이다.

그리하여 Chapman-Kolmogorov의 前向 또는 後向方程式의 解는 같은 값으로 되며 (I.1.10)에 의하여 다음의 방정식에서 구할 수 있다. 즉,

$$\begin{aligned}
&\Phi(m, m+1) = \Phi(m, m)\,P(m) = P(m) \\
&\Phi(m, m+2) = \Phi(m, m+1)\,P(m+1) = P(m)\,P(m+1) \\
&\Phi(m, m+3) = \Phi(m, m+2)\,P(m+2) = P(m)(m+1)\,P(m+2)
\end{aligned} \qquad (\text{I}.1.21)$$

이거나 또는 그 일반형

$$(\text{I}.1.22)\quad \Phi(m, n) = P(m)P(m+1)\cdots P(n-1) \qquad (m \leq n-1)$$

에서의 n의 값에서 구할 수 있다.

이것은 (I.1.21) 또는 (I.1.22)에서 (I.1.15)의 基本方程式을 만족한다.

무릇 C-K方程式體系에 있어서는 一意的 解(unique solution), 즉 推移行列에서의 Frobenius根과 微分方程式體系에서의 Brouwer 不動點으로서 不動벡터(fixed vector)가 존재한다.[13]

Ⅱ. Markov 型 逐次決定過程

1. 離散型 逐次決定過程

Markor 型 逐次決定過程[14](Markovian sequential decision process: MSDP), 즉 政策改良反覆法(policy improvement-iteration method : PIIM)은 R.A.Howard에 의하여 창안되었는데, 여기에서는 離散型 逐次決定過程에 관하여 생각하여 보기로 한다.[15]

첫째로, 利得反覆方程式(value iteration equation)으로서, 우선 상태 i에서 戰略k를 선택할 때의 期待利得 q_i^k, 상태 i에서 상태 j로 이행하는 추이확율을 p_{ij}^k, 또 상태 i에서 상태 j로 이행할때 전략 k에서의 기대이득 r_{ij}^k로 하면 q_i^k는

$$(\text{Ⅱ}.1.1)\quad q_i^k = \sum_{j=1}^{N} p_{ij}^k\, r_{ij}^k$$

를 얻는다.

10) 11) R.A.Howard, op. cit., p.514.

12) R.A.Howard, op. cit., pp.514-515.

13) R.Bellman, Dynamic Programming, Prinston Univ. Press, 1957, pp.321-325, p.330.

14) R.A.Howard, Dynamic Programmming and Markov Processes, The M.I.T.Press, 1960, p.30.

15) ditto. p.29.

그리고 상태 i에서 출발하여 전략 k를 선택할 때 기간 $n+1$ 사이에서의 多段階 決定過程에서 얻는 總期待利得(total expected rewards)은

(Ⅱ.1.2) $V_i(n+1)=\max_k \sum_{j=1}^{N} P_{ij}^k [r_{ij}^k+V_j(n)]$

$(n=0,1,2,\cdots,n)$

의 利得反覆方程式에 의하여 구한다. 그리고 시점 $n+1$ 사이에 있어서 총기대이득으로서 (Ⅱ. 1.2)의 우변

(Ⅱ. 1. 3) $\sum_{j=1}^{N} P_{ij}^k [r_{ij}^k+V_j(n)]$

을 최대로 하는 것이 最適政策[16](optimal policy)이다. (Ⅱ.1.2)에서 전략 k를 선택할때의 直接期待利得(expected immediate rewards)은

(Ⅱ.1.4) $V_i(n+1)=\max_k [q_i^k+\sum_{j=1}^{N} P_{ij}^k V_j(n)]$

의 再歸關係式(recursive equation)으로 표현된다.

둘째로, 數値決定演算[1](value determination operation)으로서, 지금 完全에르고드的(ergodic) 過程, 즉 極限狀態確率 π_i는 出發狀態와 無關係라는 것을 가정하는 것이므로 이득은

(Ⅱ.1.5) $g=\sum_{i=1}^{N} \pi_i g_i$

의 利得方程式으로 표현된다. Markov過程에서는 어떤 시점에서의 상태 i에서 출발하여 n회 移行하였을때의 총기대이득 $V_i(n)$은

(Ⅱ.1.6) $V_i(n)=q_i+\sum_{j=1}^{N} P_{ij} V_j(n-1)$

$(i=1,2,\cdots,N,\quad n=1,2,3,\cdots)$

의 再歸關係式으로 표현된 完全에르고드的 過程에 대하여 $V_i(n)$은

(Ⅱ.1.7) $V_i(n)=ng+V_i$

의 漸近式으로 표현된다.

여기에서 (Ⅱ.1.7)은 벡터의 모양으로

$V(n)=q+PV(n-1) \qquad (n=1,2,3,\cdots)$

로 표현되므로 (Ⅱ.1.7)은 (Ⅱ.1.6)에 의하여

$$nq+V_i=q_i+\sum_{j=1}^{N} P_{ij}[(n-1)g+V_j] \qquad (i=1,2,\cdots,N)$$

(Ⅱ.1.8) $$=q_i+(n-1)g\sum_{j=1}^{N} P_{ij}+\sum_{j=1}^{N} P_{ij}V_j$$

로 변형할 수 있다.

그리고 $\sum_{j=1}^{N} P_{ij}$ 이므로 위식 들은

$$g+V_i=q_i+\sum_{j=1}^{N} P_{ij}V_j \qquad (i=1,2,\cdots,N)$$

로 변형할 수 있다. (Ⅱ.1.9)는 어떤 상태 i에서의 극한상태확율 π_i를 곱하고 이것을 i에 관하여 더하면

(Ⅱ.1.10) $g\sum_{i=1}^{N}\pi_i+\sum_{i=1}^{N}\pi_i V_i=\sum_{i=1}^{N}\pi_i P_i+\sum_{j=1}^{N}\sum_{i=1}^{N}\pi_i P_{ij} V_j$

로 된다. 여기에서 $\pi=\pi P$ $\sum_{i=1}^{N}\pi_i=1$ 이므로 (Ⅱ.1.10)은

(Ⅱ.1.11) $g=\sum_{i=1}^{N}\pi_i q_i$

로 표현된다.

세째로, Markov型 逐次決定過程에 있어서의 政策改良루틴[17](policy-improvement routine : PUR)이다. 지금 $n+1$ 단계에 있어서 상태 i에 있어서의 최적정책은 (Ⅱ.1.4)에 의하여

(Ⅱ.1.12) $q_i^k+\sum_{j=1}^{N} P_{ij}^k V_j(n)$

으로 주어진다. (Ⅱ.1.12)는 총기대이득을 최대로 하는 多段階決定過程을 축차적 반복을 함으로써 얻어진 것이다. 여기에서

$V_i(n)=ng+V_i \qquad (i=1,2,\cdots,N)$

을 (Ⅱ.1.12)에 대입함으로써

(Ⅱ.1.13) $q_i^k+\sum_{j=1}^{N} P_{ij}^k (ng+V_j)$

를 최대화하도록 다단계결정과정을 반복한다. 따라서 $\sum_{j=1}^{N} P_{ij}^k=1$ 이므로 상태 i에서의 최적정책은 각 단계에 있어서의 전략에 관하여

(Ⅱ.1.14) $q_i^k+\sum_{j=1}^{N} P_{ij}^k V_j$

를 최대화하도록 한다.[18]

끝으로 Markov型 逐次決定過程에 있어서 割引率을 도입하는 경우의 數理性을 여기에서 구명하여 보기로 한다. 앞에서의 所得反覆方程式과 再歸關係式에 관하여는 割引率을 도입하는 경우를 고려하여야 한다. 따라서 지금 어떤 시점의 상태 i에서 n단계로 이행하였을때의 총기대이득 $V_i(n)$의 現在價値는 割引率 β, 즉 現在時點에서 將來時點을 예상하는 경우의 이득을 割引하여 評價[19]하여야 한다. 따라서 利得反覆方程式은

16) P.Bellman, op. cit., Ch.Ⅱ, p.83.

17) R.A.Howard, op. cit., pp.37-38, ; E.Bellman, op. cit., pp.321-329, p.330 ; R.Bollman and S.E. Dreyfus, op. cit., p.30.

19) R.A.Howard, op. cit., pp.74-91.

$$(\text{Ⅱ}.1.15)\quad V_i(n) = \sum_{j=1}^{N} P_{ij}\,[r_{ij} + \beta V_j(n-1)]$$
$$(i = 1, 2, \cdots, N,\quad n = 1, 2, 3, \cdots)$$

로 된다.

또 이 경우 再歸關係式은 직접기대이득

$$q_i = \sum_{j=1}^{N} P_{ij}\, r_{ij}$$

에 의하여

$$V_i(n) = q_i + \beta \sum_{j=1}^{N} P_{ij} V_j(n-1)$$
$$(i = 1, 2, \cdots, N,\quad n = 1, 2, 3, \cdots)$$

로 표현된다.

다시금 총기대이득의 벡터를 $V(n)$으로 하면

$$(\text{Ⅱ}.1.17)\quad V(n+1) = q + \beta P V(n)$$

으로 표현할 수 있다. 여기에서 (Ⅱ.1.17)을 Z變換을 함으로써 行列方程式으로서

$$Z^{-1}[V(z) - V(0)] = \frac{1}{1-z}\, q + \beta P V(z)$$

를 얻는다. 위식을 변형함으로써

$$V(z) - V(0) = \frac{z}{1-z}\, q + \beta Z P V(z)$$

$$(I - \beta Z P)\, V(z) = \frac{z}{1-z}\, q + V(0)$$

으로 되므로, 이에 의하여

$$(\text{Ⅱ}.1.18)\quad V(z) = \frac{z}{1-z}(I - \beta Z P)^{-1} q + (I - \beta Z P)^{-1} V(0)$$

을 얻는다. 여기에서 (Ⅱ.1.18)은 $V(n)$의 Z變換을 할 수 있으므로 (Ⅱ.1.17)은 무시하고 (Ⅱ.1.18)을 적용하는 것이 편리하다.

그런데 여기에서 축차적 개량을 위하여 다단계 반복을 시도함에 있어서 割引率 β를 도입하면

$$(\text{Ⅱ}.1.19)\quad V_i(n+1) = \max_k [\, q_i^k + \beta \sum_{j=1}^{N} P_{ij}^k V_j(n)]$$

을 얻는다. (Ⅱ.1.19)는 상태 i에 있어서 최적정책을 선택할때 $n+1$ 단계에서의 現在價値를 의미한다. 그리고 각 단계의 최적정책은

$$q_i^k + \beta \sum_{j=1}^{N} P_{ij}^k V_j(n)$$

을 최대화 하도록 축차적 반복을 시도하는데 있다.

둘째로, 數値決定演算[20]에 관하여 지금 n단계에 있어서의 기대이득의 現在價値 벡터를 Z變換을 하면

$$(\text{Ⅱ}.1.20)\quad V(z) = \frac{z}{1-z}\,(I - \beta Z P)^{-1} q + (I - \beta Z P)^{-1} V(0)$$

으로 된다. (Ⅱ.1.20)에서 $(I - \beta Z P)^{-1}$은

$$(\text{Ⅱ}.1.21)\quad (I - \beta Z P)^{-1} = \frac{1}{1-\beta Z}\, S + \mathcal{T}(\beta z)$$

로 변형할 수 있다. S는 극한상태확률로 구성되는 行列이고 $\mathcal{T}(\beta z)$는 n이 커질때 $\mathcal{T}(z)$보다 먼저 0으로 접근하는 성질을 지니고 있다. 그러므로 (Ⅱ.1.21)은

$$(\text{Ⅱ}.1.22)\quad V(z) = \frac{z}{1-z}\left[\frac{1}{1-\beta z} S - \mathcal{T}(\beta z)\right] q + \left[\frac{1}{1-\beta z}\, S + \mathcal{T}(\beta z)\right] V(0)$$

으로 된다.

여기에서 (Ⅱ.1.16)의 $V_i(n)$에 현재가치 $V_i = \lim_{n\to\infty} V_i(n)$을 대입하면

$$(\text{Ⅱ}.1.23)\quad V_i = q_i + \beta \sum_{j=1}^{N} P_{ij} V_j \qquad (i = 1, 2, \cdots, N)$$

를 얻는다. 그러므로 상태 i에서 상태 j에의 推移確率 P_{ij}와 직접기대이득 q_i가 주어지면 (Ⅱ.1.23)에 의하여 總期待利得의 現在價値를 구할 수 있다.

세째로, 割引率을 도입하는 경우의 政策改良루틴[21]에 관하여, 지금 전략 k에 있어서의 再歸關係式 (Ⅱ.1.23)의 우변

$$q_i^k + \beta \sum_{j=1}^{N} P_{ij}^k V_j(n)$$

을 최대로 하도록 단단계 결정과정을 축차적으로 반복을 시도하는 과정이 政策改良루틴이다. 따라서 위식에서 총기대이득 $V_j(n)$에 현재가치 V_j를 대입하여 상태 i에서의

$$q_i^k + \beta \sum_{j=1}^{N} P_{ij}^k V_j$$

를 최대로 하는 절차를 취한다.

2. 連續型 逐次決定過程

Markov過程에 있어서의 連續型 逐次決定過程[22]에 관하여는 첫째로, 利得反覆方程式으로서 총기대이득을 $V_i(t)$, 시간변동 $t + dt$에서의 총기대이득 $V_i(t + dt)$는

20) R.A.Howard, op. cit., pp. 81-83 ; cf.E.Bellman, op. cit., p.303, · cf.R.Bellman and S.E.Dreyfus, op.cit., pp. 297-320.

21) R.A.Howard, op. cit., pp.83-84.

22) R.A.Howard, op. cit., pp.99-104, p.99

(Ⅲ.2.1) $V_i(t+dt)$

$$= (1-\sum_{j=1} a_{ij}\,dt)\,[r_{ij}\,dt + V_i(t)] + \sum_{j\neq i} a_{ij}\,dt\,[r_{ij}+V_j(t)]$$

로 된다. 여기에서

$$a_{ii} = -\sum_{i\neq j} a_{ji}$$

에 의하여 (Ⅱ.2.1)을

$$V_i(t+dt) = (1+a_{ii}\,dt)\,[r_{ii}\,dt + V_i(t)] + \sum_{j\neq i} a_{ij}\,dt\,[r_{ij}+V_j(t)]$$

로 변형한다. 다시금 위식의 양변에서 $V_i(t)$를 제외하고 dt로 나누면

(Ⅲ.2.2) $$\frac{V_i(t+dt)-V_i(t)}{dt} = r_{ii} + \sum_{j\neq i} a_{ij}\,r_{ij} + \sum_{j=1}^{N} a_{ij}V_j(t)$$

로 된다.

여기에서 $dt \to 0$을 반복함으로써

$$\frac{d}{dt}V_i(t) = r_{ii} + \sum_{j\neq i} a_{ij}\,r_{ij} + \sum_{j=1}^{N} a_{ij}V_j(t) \qquad (i=1,2,\cdots,N)$$

을 얻는다. 다시금 總期待利得 q_i를

(Ⅲ.2.3) $$q_i = r_{ii} + \sum_{j\neq i} a_{ij}\,r_{ij}$$

로 표시하고 (Ⅲ.2.3)에 의하여 (Ⅱ.2.1)은 線型微分方程式

(Ⅲ.2.4) $$\frac{d}{dt}V_i(t) = q_i + \sum_{j=1}^{N} a_{ij}V_j(t) \qquad (i=1,2,\cdots,N)$$

로 쓸 수 있다.

(Ⅲ.2.4)는 어떤 시점의 상태 i에서 출발하여 시간 t 사이에 있어서의 총기대이득을 q_i, q_{ij}에 관련시켜 표현한 것이다. 여기에서 총기대이득 $V_i(t)$를 원소로하는 列벡터 $V(t)$, q_i를 원소로 하는 利得벡터를 q로 할때 다음과 같이 行列의 모양으로

(Ⅲ.2.5) $$\frac{d}{dt}V(t) = q + AV(t)$$

로 쓸 수 있다. (Ⅲ.2.5)의 解는 一意的 解(unique solution)로서 Frobenius 根 또는 Brouwer 不動點으로서 주어진다.[23]

여기에서 (Ⅲ.2.5)는 線型聯立微分方程式으로서 Laplace 變換을 하는 것이 편리하므로 이것을 Laplace 變換을 하면

$$sV(s) - V(0) = \frac{1}{s}q + AV(s)$$

즉,

$$(sI-A)V(s) = \frac{1}{s}q + V(0)$$

으로 된다. 그리고 또

(Ⅲ.2.6) $$V(s) = \frac{1}{s}(sI-A)^{-1}q + (sI-A)^{-1}V(0)$$

을 얻는다. 그리하여 (Ⅱ.2.6)은 $V(t)$의 Laplace 變換과 $(sI-A)^{-1}$, 利得벡터 q, 최종이득벡터 $V(0)$의 상호관계를 표시하는 것으로서 (Ⅲ.2.6)을 다시금 逆變換을 함으로써 利得벡터 $V(t)$를 구할 수 있다.

둘째로, 連續型 逐次決定過程에 있어서의 數値決定演算[24]에 관하여, 지금 어떤 정책이 주어 졌을때 시간 t에서의 총기대이득은

$$\frac{d}{dt}V_i(t) = q_i \sum_{j=1}^{N} a_{ij}V_j(t) \qquad (i=1,2,\cdots,N)$$

에 의하여 결정된다. 여기에서 t가 커졌을 때 $V_i(t)$에 관한 漸近式

(Ⅲ.2.8) $$V_i(t) = tg_i + V_i$$

에 의하여 (Ⅲ.2.7)을 再歸關係式

$$g_i = q_i + \sum_{j\neq i}^{N} a_{ij}(tg_j + V_j)$$

즉,

(Ⅲ.2.9) $$g_i = q_i + t\sum_{j=1}^{N} a_{ij}g_j + \sum_{j\neq i}^{N} a_{ij}V_j \qquad (i=1,2,\cdots,N)$$

로 변형할 수 있다.

그리하여 (Ⅲ.2.9)는 모든 커지는 t에 관하여 성립되므로

(Ⅲ.2.10) $$\sum_{j=1}^{N} a_{ij}g_j = 0 \qquad (i=1,2,\cdots,N)$$

과

(Ⅲ.2.11) $$g_i = q_i + \sum_{j=1}^{N} a_{ij}V_j \quad (i=1,2,\cdots,N)$$

로 된다. (Ⅲ.2.11)에 의하여 각 상태의 이득을 再歸過程의 이득으로 표시할 수 있다. 또 (Ⅲ.2.11)에 의하여 나머지 再歸過程의 이득까지 구할 수 있다.

세째로, 政策改良루틴[25]에 관하여, 지금 최적정책이 주어졌을때 시간 t의 변동에 따르는 정책의 총기대이득을 $V_i(t)$로 할때 (Ⅲ.2.7)에 의하여 상태 i에서의 전략 k에 관하여

23) R.Bellman, op. cit., pp.321-325, p.330

24) R.Bellman and S.E.Dreyfus, op.cit., pp.297-320, p.330, · cf. Bellman, op. cit., pp.317-336.

25) R.Bellman and S.E.Dreyfus, op.cit., p.304, pp.297-320, · R.Bellman, op.cit., pp.317-336.

(Ⅲ.2.12) $q_i^k + \sum_{j=1}^{k} a_{ij}^k V_j(t)$

를 최대로 함으로써 $V_i(t)$를 최대화할 수 있다. 그리고 시점 t가 커질때에는 $V_j(t) = tg_j + V_j$ 에 의하여 상태 i에서의 再歸關係式

$$q_i^k + \sum_{j=1}^{N} a_{ij}^k (tg_j + V_j)$$

즉,

(Ⅲ.2.13) $q_i^k + \sum_{j=1}^{N} a_{ij}^k V_j + t \sum_{j=1}^{N} a_{ij}^k g_j$

얻는다. 그리하여 政策改良루틴에 있어서 (Ⅲ.3)을 최대로 하는 축차적 반복을 시도한다.

끝으로 連續型 逐次決定過程에 있어서 割引率을 도입하는 경우[26], 지금 $V_i(t)$가 시간 t에서의 총 기대이득이라고 할때 割引率을 α로 하면 (Ⅱ.2.1)에 의하여

(Ⅲ.2.14) $V_i(t+dt) = (1-\alpha dt)\{(1-\sum_j a_{ij}\, dt + V_i(t)\}$

를 얻는다. 여기에서 $a_{ii} = -\sum_{i,j} a_{ij}$ 이므로

(Ⅲ.2.15) $V_i(t+dt) = (1-\alpha dt)\{1 + a_{ii}\, dt\,[r_{ii}\, dt + V_i(t)] + \sum_{j \neq i} a_{ij}\, dt\,[r_{ij} + V_j(t)]\}$

또는

(Ⅲ.2.16) $V_i(t+dt) = (1-\alpha\, dt)\{(r_{ii} + \sum_{j \neq i} a_{ij}\, r_{ij})\, dt + V_i(t) + \sum_{j \neq i}^{N} a_{ij}\, dt\, V_j(t)\}$

를 얻는다.

다시금 (Ⅱ.2.16)을 dt보다 고차항을 무시함으로써

(Ⅲ.2.17) $V_i(t+dt) = (r_{ii} + \sum_{j \neq i} a_{ij}\, r_{ij})dt + V_i(t) + \sum_{j \neq i}^{N} a_{ij}\, dt\, V_j(t) - \alpha\, dt\, V_i(t)$

로 변형할 수 있나. 여기에서 (Ⅱ.2.2)로부터 收益率(earning rate)을 도입하여 정돈하면

(Ⅲ.2.18) $V_i(t+dt) - V_i(t) + \alpha\, dt\, V_i(t) - q_i\, dt + \sum_{j=1}^{N} a_{ij}\, dt\, V_j(t)$

를 얻는다.

다시금 (Ⅲ.2.18)을 dt로 나누어 $dt \to 0$일때

(Ⅲ.2.19) $\frac{d}{dt} V_i(t) + \alpha V_i(t) = q_i + \sum_{j=1}^{N} a_{ij} V_j(t)$ $(i = 1, 2, \cdots, N)$

로 된다. (Ⅲ.2.19)의 線型聯立微分方程式은 다음과 같이 行列의 모양

(Ⅲ.2.20) $\frac{d}{dt} V(t) + \alpha v(t) = q + AV(t)$

로 바꾸어 쓸 수 있다.

그런데 (Ⅲ.2.20)은 선형연립미분방정식이므로 Laplace 變換을 하는 것이 편리하다. 그러므로 이것을 Laplace 變換을 하면

$$sV(s) - V(0) + \alpha V(s) = \frac{1}{s} q + AV(s)$$

즉,

$$[(s+\alpha) I - A]\, V(s) = \frac{1}{s} q + V(0)$$

으로 된다. 그리고 끝으로

(Ⅲ.2.21) $V(s) = \frac{1}{s}[(s+\alpha) I - A]^{-1} q + [(s+\alpha) I - A]^{-1} V(0)$

을 얻는다. 그런데 여기에서

(Ⅲ.2.22) $(sI - A)^{-1} = \frac{1}{s} S + \mathcal{J}(s)$

로 되는데, S는 극한상태확률의 行列을 표시하며 또 $\mathcal{J}(s)$는 하나의 원소이므로

(Ⅲ.2.23) $[(s+\alpha) I - A]^{-1} = \frac{1}{s+\alpha} S + \mathcal{J}(s-\alpha)$

로 된다. 그러므로 (Ⅱ.2.23)을 (Ⅱ.2.22)에 대입함으로써

(Ⅲ.2.24) $V(s) = \frac{1}{s}[\frac{1}{s+\alpha} S + \mathcal{J}(s+\alpha)]\, q + [\frac{1}{s+\alpha} S + \mathcal{J}(s+\alpha)]\, V(0)$

을 얻는다.

그런데 현재까지 V_i의 벡터 V를

$$V = \lim_{t \to \infty} V(t)$$

로 하면

$$V = [\frac{1}{\alpha} S + \mathcal{J}(\alpha)]\, q$$

또는 (Ⅲ.2.22)에 의하여

(Ⅲ.2.25) $V = (\alpha I - A)^{-1} q$

로 된다. 여기에서 V는 장시간에 걸친 割引利得이며, (Ⅲ.2.25)에 의하여 現在價値와 割引率 α 또는 收益率 q의 관계를 알 수 있다.

結　　論

이상에서 經營科學 분야의 입장에서 Markov過程과 그 逐次決定過程의 數理的 모델로서 Chapman-Kolmogorov 方程式시스템과 Kolmogorov 微分方程式시스템을 解析的으로 분석하고 解의 一意

性과 存在問題를 다루었다. 무릇 理論經濟學에서는 均衡解의 一意性과 存在問題의 論證을 위하여 Walras 모델과 Arrow-Debreu 모델 그리고 Brouwer 不動點定理[27]에 의한 분석이 성공적으로 이루어지고 있다.

마찬가지로 經營工學의 관점에서는 Markov 過程의 定礎를 이루는 推移確率行列과 Chapman-Kolmogorov 方程式體系에 있어서 Frobenius 根[28](最大非負固有値)과 Brouwer 不動點定理에 의한 解의 一意性과 存在問題에 대한 究明이 이루어지고 있다. 무릇 Markov型 確率過程論的 現象을 "論理와 直觀의 融合"으로 嚴密한 數理的 構築을 위한 座標軸을 建設하는 일은 喫緊한 과제라고 생각된다.

參 考 文 獻

(1) A.N. kolmogorov, Grundbegriffe der Wahrscheilichkeitsrechnung, Ergebnisse der Mathematik und ihrer Grenzgebriete, Springer, 1933.

(2)_____, (English transration by No Morrison, Foundations of the Theory of Probability, 2nd English ed. Chelsea Publishing Company, New York, 1957.

(3) R.A. Howard, Dynamic Probabilistic Systems, Vol. I. Markov Models, John Wiley & Sons, Inc. New York, 1971.

(4)_____, Dynamic Programming and Markov Processes, The M.I.T. Press, 1960.

(5) W. Feller, An Introduction to Probability Theory and its Application, Vol. I. 2nd ed. John Wiley & Sons, Inc. New York, 1957.

(6)_____, ditto., Vol. II. John Wiley & Sons. Inc. New York, 1966.

(7) R. Bellman, Dynamic Programming

(8)_____, Princeton Univ. Press, Princeton, New Jersey, 1957. Dynamic Programming and Modern Control Theory, Academic Press New York, 1965.

(9) .R. Bellman, S.E. Dreyfus, Applied Dynamic Programming,

(10) S.E. Dreyfus, Dynamic Programming and Calculus of Variatiem, Academic Press, New York, 1965.

(11) J.G. Kemeny, J.I. Snell, Finite Markov Chains, Van Nostrand, 1960.

(12) K.I. Chung, Markov Chains with Stationary Transition Probabilities, Springer, 1960.

(13) M. Frechet, Méthode desfontions arbitraires, Theorie des Evénément en Chaines dans les cas d'un Nombre fini d'etats Passibles. Traite du Calcul des Probabilites et ses Application, Tome I, Secord Libre, 1938.

(14) C.F. Derman, Finite State Markovian Decision Processes, Academic Press, 1970.

(15) H. Wielandt, "Unzerlegbare, nicht negative Matrizen, Mathematische Zeitschrift, 52. 1950.

(16) Arrow, S. Karlin and H. Sarf, Studies in Applied Probabilities and Management Science Stanford, Calif, 1962.

(17) S.M. Ross, Applied Probability Models with Optimization Applications, Holden-Day, San Francisco, 1970.

(18) S. Karlin, A First Course in Stochastic Processes, Academic Press, New York, 1966.

(19) H. Mine, S. Osaki, Markovian Decision Processes, Elsevier Publishing Company, 1970.

(20) ORSOJ ed., Selected Papers for Operations Research Society of Japan Anual Paper Award, (By S. Osaki, " System Reliability Analysis by Markov Renewal Processes"), Sep., 1982, pp. 80-141.