数理解析研究所講究録
第1636巻 2009年 40-47

不完備情報の多段決定問題と評価について

千葉大学教育学部　中井　達 (Tōru Nakai)
Faculty of Education, Chiba University

1　はじめに

[8] において、評価と関連する状態に関する状態をもとに、支出を決定する逐次決定問題を扱った。また、状態は支出することによって変えることが出来た。ここでは、[8] などで扱った問題を費用最小化問題に応用することを考える。

いま、自動車や電化製品などに関して問題が生じたとき、問題点が大きくなければメーカーは個別的に対応することが可能であるが、問題が大きく部品の取り替える必要性が生ずるような場合には、リコールを行うことになる。そこで、製品の状態を $[0,\infty)$ によって表し、状態が 0 のとき問題がないと考えられる状態を表し、状態を表す値 s が大きくなれば製品の抱える問題が大きくなるとする。また、考える製品に対するクレームの大きさは、状態に応じて異なり、非負の確率変数 X_s で表されるとする。このとき、この確率変数の値を観測し、この値をもとに決定を行い、計画期間内で費用を最小化する問題を解析する。

2　逐次決定問題

状態空間を $[0,\infty)$ とし、状態が 0 のとき良い状態を表し、状態を表す値 s が大きくなれば状態が悪くなるとする。また、状態 $s\in[0,\infty)$ に対して、非負の確率変数 X_s を仮定する。これらの確率変数は $E[X_s]<\infty$ であり、s に関して確率的に増加とする。すなわち、s の値が大きくなればなるほどクレームの大きさも大きくなると考える。また、状態に関して不完備情報の場合には、これらの確率変数を観測することを情報過程とし、この値を観測することによって状態に関する情報を得る。さらに、$u(s)$ を最後の期の状態が s のときの終端費用とし、$c(x)$ は x を観測したときの費用とする。このとき、$u(s)$ は s に関して非減少で、非負の凹 (concave) 関数とし、$c(x)$ は x に関して非負の非減少関数とする。

はじめに、x を観測したとき x をクレームの大きさと考え、クレームの大きさ x にしたがって費用 $c(x)$ を支払って対応するか、それともリコールを行って問題を根本的に解決するかの 2 つの決定をとることが出来る場合を考える。したがって、状態はリコールのときにのみ変化し、その費用は状態によって変化し $C(s)$ とする。この $C(s)$ は、s に関して単調増加であり、非負で有界な凹 (concave) 関数とする。そのため、状態が確率的に推移しない場合から考える。

このとき、$u_n(s)$ を計画期間が n で状態が s のとき、最適に振る舞って得られる総期待費用とし、$u_n(s|x)$ を計画期間が n で状態が s のとき x を観測し、最適に振る舞って得られる総期待費用すれば、最適性の原理より、

$$u_n(s)=E_{X_s}[u_n(s|X_s)]$$

$$u_n(s|x) = \min\{C(s) + u_{n-1}(0), c(x) + u_{n-1}(s)\} \tag{1}$$

となる。ただし、$u_1(s|x) = \min\{C(s), c(x) + u(s)\}$ である。

$C(s) \leq u(s)$ とすれば、最適政策は何もしないことである。

補題 1 $u_n(s)$ は s に関する増加関数で、凹 *(concave)* 関数である。

つぎに、状態が確率的に推移する場合を考えてみる。状態は推移法則を $\boldsymbol{P} = (p_s(t))_{s,t\in[0,\infty)}$ とするマルコフ過程にしたがって推移し、確率変数 X_s とは独立とする。さらに、推移法則には、TP_2 を仮定する (定義 1)。

定義 1 推移法則 $\boldsymbol{P} = (p_s(t))_{s,t\in[0,\infty)}$ は、$s \leq t$ および $u \leq v$ となる任意の s, t, u と v に対して $(s, t, u, v \in [0, \infty))$、$\begin{vmatrix} p_s(u) & p_s(v) \\ p_t(u) & p_t(v) \end{vmatrix} \geq 0$ となる。

このとき、$v_n(s)$ を計画期間が n で状態が s のとき、最適に振る舞って得られる総期待費用とし、$v_n(s|x)$ を計画期間が n で状態が s のとき x を観測し、最適に振る舞って得られる総期待費用すれば、最適性の原理より、

$$v_n(s) = E_{X_s}[v_n(s|X_s)]$$

$$v_n(s|x) = \min\left\{C(s) + v_{n-1}(0), c(x) + \int p_s(t)v_{n-1}(t)dt\right\} \tag{2}$$

となる。ただし、$n = 1$ のとき、$v_1(s|x) = \min\{C(s), c(x) + u(s)\}$ とする。

補題 2 $v_n(s)$ は s に関する増加関数である。

補題 3 推移法則が TP_2 のとき、$\int p_s(t)v_{n-1}(t)dt$ は s に関する増加関数である。

3 部分観測可能なマルコフ過程

状態空間が $[0, \infty)$ のマルコフ過程で、推移法則を $\boldsymbol{P} = (p_s(t))_{s,t\in[0,\infty)}$ とする。すなわち、$\boldsymbol{p}_s = (p_s(t))_{t\in[0,\infty)}$ は、任意の $s \in [0, \infty)$ に対して、$[0, \infty)$ 上の確率分布を表す。

このとき、このマルコフ過程の状態を直接知ることが出来ない不完備情報のマルコフ過程を考える。すなわち、このマルコフ過程の状態に関する情報は、状態空間 $[0, \infty)$ 上の確率分布 $\boldsymbol{\mu}$ で表し、$\mathcal{S}$ を状態に関する情報全体の集合とすれば、

$$\mathcal{S} = \left\{\boldsymbol{\mu} = (\mu(s))_{s\in[0,\infty)} \,\middle|\, \int_0^\infty \mu(s)ds = 1, \mu(s) \geq 0\,(s \in [0, \infty))\right\}$$

となる。

$\mathcal{S}$ に含まれる情報のあいだに、定義 1 と同じように、TP_2 を用いて半順序を定義する。この問題では、s が大きくなれば状態は悪くなるので、$\boldsymbol{\mu}$ が $\boldsymbol{\nu}$ よりこの半順序の意味で大きいとき、状態に関する情報は悪い情報を多く含むことになる。

いっぽう、$\boldsymbol{p}_s = (p_s(u))$ および $\boldsymbol{p}_{s'} = (p_{s'}(u))$ とおけば、$\boldsymbol{P}$ が定義 1 を満たすことから、任意の s, s' $(s \leq s',\ s, s' \in [0, \infty))$ に対して、$\boldsymbol{p}_{s'} \succeq \boldsymbol{p}_s$ となる。このとき、これらの仮定のもとで、Kijima and Ohnishi[2] などからつぎの性質が成り立つ。

補題 4 $\boldsymbol{\mu} \succeq \boldsymbol{\nu}$ ならば $(\boldsymbol{\mu}, \boldsymbol{\nu} \in \mathcal{S})$、$x$ に関する非減少な非負関数 $h(x)$ に対して、$\int_0^\infty h(x)dF_{\boldsymbol{\mu}}(x) \geq \int_0^\infty h(x)dF_{\boldsymbol{\nu}}(x)$ となる。

3.1 部分観測可能なマルコフ過程と情報

つぎに、それぞれの状態 s に対して、状態に依存する確率変数 X_s を情報プロセスとする。すなわち、それぞれの状態に関する情報を確率変数 X_s を通して得ることができる情報システムあるいは観測過程を考える。さらに、状態は直接には観測できず、状態に依存する確率変数 X_s を通じて情報が得られ ($s \in [0,\infty)$)、学習プロセスはベイズ学習にしたがって解析することから、仮定 1 を設ける。状態 s に対して、確率変数 X_s は絶対連続で、密度関数 $f_s(x)$ を持つとする ($s \in [0,\infty)$)。この仮定は、Nakai [6] にしたがって一般化でき、多段決定問題へ応用できる (Nakai [3, 4, 5] など)。

仮定 1 確率変数 $\{X_s\}_{s\in[0,\infty)}$ に対して、$s \le s'$ ならば、$X_{s'} \succeq X_s$ である ($s, s' \in [0,\infty)$)。すなわち、X_s は s に関して尤度比の意味で増加する。

仮定 1 において、$X_s \succeq X_{s'}$ のとき、$x < y$ ならば、$s \le s'$ となる任意の s と s' に対して ($s, s' \in [0,\infty)$)、$f_s(y)f_{s'}(x) \le f_s(x)f_{s'}(y)$ である。このことから、確率変数 X_s は s の値が大きくなるにしたがって、大きな値をとるようになり、状態 0 が一番良い状態であり、$\cdots$、s が大きくなるにしたがって悪くなり、それとともにクレームの大きさが大きくなる。また、推移法則に関する仮定から、現在の状態から、より悪い状態に推移する確率は、現在の状態が悪くなるにしたがって増加する。すなわち、それぞれの状態を表す s が大きくなれば、より悪い状態に推移する確率は大きくなるのである。

すなわち、この確率変数を観測することによって、状態に関して学習を行う。事前情報が μ のとき、まずはじめにこれらの確率変数 $\{X_s\}_{s\in[0,\infty)}$ を観測し、ベイズの定理を用いて学習を行う。その後、状態は推移し新しい状態になると考える。もちろん、この順序を変えても同じように解析できる。x を観測したとき、ベイズの定理にしたがって学習した事後情報を $\mu(x) = (\mu(x,s))_{s\in[0,\infty)}$ とし、その後で推移法則 $\boldsymbol{P}$ にしたがって状態が推移し、つぎの新しい状態に関する情報を $\overline{\mu(x)} = (\overline{\mu(x,s)})_{s\in[0,\infty)}$ と表す。

このとき、集合値関数 $\boldsymbol{h}(x,s)$ に対して、定義 2 によって単調性を定義する。

定義 2 任意の $s \in [0,\infty)$ と $x \in \Re_+$ に関する非負の集合値関数 $\boldsymbol{h}(x) = (h(x,s))_{s\in[0,\infty)}$ に対して、任意の s' と s ($s \le s'$ かつ $s, s' \in [0,\infty)$) について、$x < y$ ならば $\boldsymbol{h}(y) \succeq \boldsymbol{h}(x)$ ($\boldsymbol{h}(x) \succeq \boldsymbol{h}(y)$) とする。すなわち $h(x,s')\,h(y,s) \le h(x,s)\,h(y,s')$ ($h(x,s')\,h(y,s) \ge h(x,s)\,h(y,s')$) である。このとき、関数 $\boldsymbol{h}(x,s)$ を x に関する増加関数 (減少関数) という。

事前情報 μ と事後情報 $\overline{\mu(x)}$ のあいだには、マルコフ過程の推移法則に関する仮定と仮定 1 のもとで、つぎの基本的な性質が成り立つ (Nakai [6, 7] など)。

補題 5 $\mu \succ \nu$ ならば、任意の y に対して、$\mu(x) \succ \nu(x)$ および $\overline{\mu(x)} \succ \overline{\nu(x)}$ である。任意の μ に対して、$\mu(x)$ と $\overline{\mu(x)}$ は x に関する増加関数である。

補題 5 から、事前情報 μ における順序関係は、$\mu(x)$ と事後情報 $\overline{\mu(x)}$ に対して保たれる。さらに、同じ事前情報 μ であれば、観測した値 x が大きくなれば、事後情報 $\overline{\mu(x)}$ もまたよくなる。

4 不完備情報の多段決定問題

部分観測可能なマルコフ過程で考える。すなわち、状態を直接知ることは出来ず、状態に関する情報を知っている場合である。いま、μ を状態に関する事前情報とし、TP_2 による半順序を仮定する。このとき、$v_n(\mu)$ を計画期間が n で状態に関する事前情報が μ のとき、最適に振る舞って得られる総期待費用とし、$v_n(\mu|x)$ を計画期間が n で状態に関する事前情報が μ のとき x を観測し、最適に振る舞って得られる総期待費用すれば、最適性の原理より、

$$v_n(\mu) = E_X[v_n(\mu|X)] = \int E_{X_s}[v_n(s|X_s)]\mu(s)ds$$

$$v_n(\mu|x) = \min\left\{\int C(s)\mu(s)ds + v_{n-1}(0), c(x) + \int\left\{\int p_s(t)v_{n-1}(t)dt\right\}\mu(s)ds\right\} \tag{3}$$

となる。ただし、$v_1(\mu|x) = \min\left\{\int C(s)\mu(s)ds, c(x) + \int u(s)\mu(s)ds\right\}$ である。である。このとき、つぎの性質が成り立つ。

補題 6 $v_n(\mu)$ は μ に関する増加関数である。

補題 7 $\int \mu(s)ds \int p_s(t)v_{n-1}(t)dt$ は μ に関する増加関数である。

また、部分観測可能なマルコフ過程での問題を考えることができる。すなわち、クレームの大きさによって状態に関する情報を得て決定を行う場合である。このとき、$\bar{v}_n(\mu)$ を計画期間が n で状態に関する事前情報が μ のとき、最適に振る舞って得られる総期待費用とし、$\bar{v}_n(\mu|x)$ を計画期間が n で状態に関する事前情報が μ のとき x を観測し、最適に振る舞って得られる総期待費用すれば、最適性の原理より、

$$\begin{aligned}\bar{v}_n(\mu) &= E_X[\bar{v}_n(\mu|X)] = \int E_{X_s}[\bar{v}_n(s|X_s)]\mu(s)ds \\ \bar{v}_n(\mu|x) &= \min\left\{\int C(s)\mu(x)(s)ds + \bar{v}_{n-1}(0),\right. \\ &\qquad \left. c(x) + \int\left\{\int p_s(t)\bar{v}_{n-1}(t)dt\right\}\mu(x)(s)ds\right\} \end{aligned} \tag{4}$$

となる。ただし、$\bar{v}_1(\mu|x) = \min\left\{\int C(s)\mu(s)ds, c(x) + \int u(s)\mu(x)(s)ds\right\}$ である。このとき、つぎの性質が成り立つ。

補題 8 $\bar{v}_n(\mu)$ は μ の増加関数であり、$\bar{v}_n(\mu|x)$ は μ と x の増加関数である。

4.1 決定が推移に影響する多段決定問題

クレームの大きさを見てリコールするかしないかを決定するのではなく、[8] で考えたように、リコールを全て取り替えることから、一部を取り替えるといったように、クレームの大きさによって対応を変化させることが出来る場合を考える。すなわち状態 s を、観測値 x によって、変化させることができると考える。部分的にでもクレームに対応することで、状態が良くなり、その変化の度合いは決定に依存する場合である。

このとき、$w_n(s)$ を計画期間が n で状態が s のとき、最適に振る舞って得られる総期待費用とし、$w_n(s|x)$ を計画期間が n で状態が s のとき x を観測し、最適に振る舞って得られる総期待費用すれば、最適性の原理より、

$$w_n(s) = E_{X_s}[w_n(s|X_s)]$$

$$w_n(s|x) = c(x) + \min_{0\leq\alpha\leq 1}\{C(\alpha) + w_{n-1}(\alpha s), w_{n-1}(s)\} \tag{5}$$

とする。ただし、$w_1(s|x) = c(x) + \min_{0\leq\alpha\leq 1}\{C(\alpha) + u(\alpha s)\}$ とする。このとき、観測値 (クレームの大きさ) によって、クレームにどの様に対応できるかは、現在の状態 s の大きさに関わりなく、等倍率で状態を良くできるとし、そのための費用は絶対量ではなく倍率で定まると考える。すなわち、状態が s のとき、この状態は s の α 倍にすることができ $(0 \leq \alpha \leq 1)$、状態を α 倍だけよくするための費用を $C(\alpha)$ とする。この $C(\alpha)$ は状態を α 倍だけよくするための費用だから、α に関して減少関数である。

$u(s)$ が s に関する増加関数だから、$w_1(s|x)$ も s に関する増加関数である。さらに帰納法により、$w_{n-1}(s)$ が s の増加関数で、$w_{n-1}(\alpha s)$ は α の増加関数だから、$w_n(s|x)$ も s に関する増加関数である。したがって、$w_n(s)$ も s に関する増加関数となる。また、$w_{n-1}(\alpha s)$ も、α の増加関数である。さらに、$\alpha = 1$ のときは、$v_{n-1}(\alpha s) = v_{n-1}(s)$ であり、$\alpha = 0$ のときは、$v_{n-1}(\alpha s) = v_{n-1}(0)$ である。

注 1 $C(\alpha)$ と $u(s)$ がともに、凹 (*concave*) 関数とすれば、$C(\alpha) + u(\alpha s)$ は α に関する凹 (*concave*) 関数である。したがって、$\alpha = 1$ または、$\alpha = 0$ で最小値をとるので、リコールするかしないかの 2 決定問題と同じとなる。

つぎに、状態がマルコフ過程にしたがって推移する場合を考える。いま、$\bar{w}_n(s)$ を計画期間が n で状態が s のとき、最適に振る舞って得られる総期待費用とし、$\bar{w}_n(s|x)$ を計画期間が n で状態が s のとき x を観測し、最適に振る舞って得られる総期待費用すれば、最適性の原理より、

$$\bar{w}_n(s) = E_{X_s}[\bar{w}_n(s|X_s)]$$

$$\bar{w}_n(s|x) = c(x) + \min_{0\leq\alpha\leq 1}\{C(\alpha) + \int p_{\alpha s}(t)\bar{w}_{n-1}(t)dt, \int p_s(t)\bar{w}_{n-1}(t)dt\} \tag{6}$$

とする。ただし、$\bar{w}_1(s|x) = c(x) + \min_{0\leq\alpha\leq 1}\{C(\alpha) + u(\alpha s)\}$ とする。このとき、つぎの性質が成り立つ。

補題 9 $\bar{w}_n(s)$ は s の増加関数であり、$\bar{w}_n(s|x)$ は s と x の増加関数である。

4.2 Gradually Condition

[8] において、状態空間を $[-\infty, \infty]$ のとき、不完備情報のマルコフ過程での最適決定問題を考えるための条件を考えた。[8] で考えた支出モデルでは、決定がつぎの期の状態に影響することからもこれらの条件が必要であった。このなかで、状態が s のとき決定 x をとれば、状態は $s(x) = s + d(x)$ となると仮定した。このとき、$d(x)$ は、$d(0) = 0$ で、x に関する増加関数である。このとき、$\mu_x(t)$ を事前情報が μ のと

き、決定 x をとったときの、状態空間上の事後分布とする。ここで、$s(0)=s$ だから、$\overline{\mu}=\int_0^\infty \mu(s)p_s(t)ds=\mu_0$ である。

さらに、状態の推移、学習、決定と事後情報との関係を見るため、つぎの性質と仮定を置いた。いま、状態全体の集合 S に含まれる確率分布 μ が「$s<t, s'<t'$ と $s-s'=t-t'=c<0$ を満たす任意の $s<s', t\leq t'$ に対して、$\frac{\mu(s)}{\mu(s')}\geq\frac{\mu(t)}{\mu(t')}$」の性質を満たすとき、この μ は gradually condition を満足するということにする。

簡単な計算により、状態空間上の正規分布 $\mu(s)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(s-a)^2}{2\sigma^2}}$ はこの性質を満足する。

補題 10 状態全体の集合 S に含まれる確率分布 μ が *gradually condition* を満足するとき、$x>y$ ならば、$\mu_x\succeq\mu_y$ である。ただし、$\mu_x=(\mu_{s(x)})$ とする。

ここで、$\overline{\mu}(t)=\int_0^\infty \mu(s)p_s(t)ds$ とおく。つぎの性質を導くため、推移法則に関してつぎの仮定をおく。

仮定 2 任意の $s<s', t\leq t'$ および $u<v$ となる s,s',t,t',u,v に対して $p_u(s)p_v(t')-p_u(t)p_v(s')\geq p_v(s)p_u(t')-p_v(t)p_u(s')$ とする。

例 1 正規分布による推移法則 $p_v(s)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(s-v)^2}{2\sigma^2}}$ は、仮定 *2* の条件を満足する。

情報プロセスからの観測値 y が得られたときの事後情報 $\overline{\mu(y)}=(\overline{\mu(y,s)})$

補題 11 状態全体の集合 S に含まれる確率分布 μ が *gradually condition* を満足するならば、任意の y に対して $\overline{\mu(y)}$ もまた *gradually condition* を満足する。

ここで、観測できない状態に関する情報が μ で、決定が x のときの、状態空間上の確率分布を $\overline{\mu(x)}$ とおく。このとき、これまでの性質をまとめれば、つぎの補題が得られる。

補題 12 状態全体の集合 S に含まれる確率分布 μ が *gradually condition* を満足するならば、$\overline{\mu(x)}$ もまた *gradually condition* を満足する。

補題 13 状態全体の集合 S に含まれる確率分布 μ と ν が *gradually condition* を満足するとき、$\mu\succeq\nu$ ならば、任意の $x(\geq 0)$ に対して $\overline{\mu(x)}\succeq\overline{\nu(x)}$ である。

簡単な計算から、任意の x に対して推移法則 $(p_{s(x)})(t))_{0\leq s\leq 1}$ が TP_2 であるから、これまでに議論してきた仮定の下で、つぎの性質が成り立つ。

補題 14 状態全体の集合 S に含まれる確率分布 μ が *gradually condition* を満足するとき、$x>y$ ならば $\overline{\mu(x)}\succeq\overline{\mu(y)}$ である。

例 2（**対数正規分布**）確率変数 Y を正規分布 $Y\sim N(\mu,\sigma^2)$ とするとき、$X:=e^Y$ で定義される確率変数を対数正規分布といい、$y>0$ のとき、事象 $\{X\leq x\}$ と事象

$\{Y \leq \log x\}$ は等しいので、$Pr(X \leq x) = Pr(Y \leq \log x)$ より、X の分布関数 $F_X(x)$ は

$$F_X(x) = \int_{-\infty}^{\log x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

であり、密度関数 $f_X(x)$ は $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}}$ である。いま、正規分布 $N(\mu, \sigma^2)$ の密度関数を $\phi(x)$ とすれば、$f_X(x) = \phi(\log x)$ だから、$f_X(\alpha x) = \phi(\log \alpha x) = \phi(\log \alpha + \log x)$ となる。

[8] で扱った、評価を考慮した支出モデルでは、状態が s のとき、決定 x をとれば、状態を $s(x) = s + d(x)$ となる場合に、不完備情報のマルコフ過程での多段決定問題の性質を、gradually condition の性質を仮定して考えた。ここでは、状態が s のとき、決定 α をとれば、状態を αs と仮定した。すなわち、$s(\alpha) = \alpha s$ と考えればよい。よって、状態全体の集合 $\mathcal{S}$ に含まれる確率分布 μ が

$s < t, s' < t'$ と $\frac{s}{s'} = \frac{t}{t'} = \alpha < 0$ を満たす任意の $0 < s < s'$, $0 < t < t'$ に対して、$\frac{\mu(s)}{\mu(s')} \geq \frac{\mu(t)}{\mu(t')}$ となる

の性質を満たすときを考える。上記の性質から、集合 $\mathcal{S}$ に含まれる確率分布 μ として対数正規分布を考えれば、この条件を満足するので、この場合を考える。

4.3 不完備情報の多段決定モデル

状態がマルコフ過程にしたがって推移し、その状態を直接知ることができず、クレームの大きさによって状態に関する情報を得る場合の逐次決定問題を考えることにしよう。クレームの大きさを知ることを、状態に関する情報を得る情報プロセスと考える。したがって、このモデルは、3 節の部分観測可能なマルコフ過程での逐次決定問題として定式化できる。

このような部分観測可能なマルコフ過程での逐次決定問題において、観測できない状態に関する情報は、状態空間上の確率分布として表され、前節で考えた性質を持つものとする。このとき、クレームの大きさを観測値とし、この値をもとにベイズの定理にしたがって学習を行う。また、3 節の部分観測可能なマルコフ過程においては、それぞれの状態 s $(s \in [0, \infty))$ に対して、クレームの大きさを表す確率変数 Y_s を観測過程と考え、この値を観測することが情報プロセスとなる。いま、状態に関する事前情報を μ とし、計画期間が n のとき、最適政策にしたがって得られる総期待費用を $\overline{w}_n(\mu)$ とする。このとき、最適性の原理より、つぎの再帰方程式が得られる。

$$\begin{aligned} \overline{w}_n(\mu) &= \int_0^\infty \overline{w}_n(\mu|x) d\mu(x) \\ \overline{w}_n(\mu|x) &= \max_{0 \leq \alpha \leq 1} \left\{ -c(x) + \overline{w}_{n-1}(\overline{\mu(x)_\alpha}) \right\} \end{aligned} \tag{7}$$

ここで、$\overline{w}_0(\mu) = \int_0^1 u(t) d\mu(t)$ とする。(7) 式において、$\mu(x)$ は、情報プロセスからクレームの大きさとして得られた値 x をもとに情報を改良したあとの、状態に関する

事後情報とする。 すなわち、事前情報が μ のとき、まず始めに情報プロセスから観測値 x を観測し、状態に関する情報をベイズの定理にしたがって $\mu(x)$ 改良する。そのあと、決定 α をとったあとに推移する。さらに、決定後の状態が s であれば、推移法則 $(p_{s(x)}(t))_{0\leq s\leq 1}$ にしたがって状態が推移し、1 期間移動する。こうして、この確率過程は新しい状態となり、この新しい状態に関する情報は $\overline{\mu(x)_\alpha}$ となる。これは、学習したあと 1 期間経過後の状態空間上の確率分布である。そのあとで、最適政策にしたがって得られる残り計画期間での総期待利得は $\overline{w}_{n-1}(\overline{\mu(x)_\alpha})$ となる。よって、n に関する帰納法を用いれば、つぎの性質が得られる。.

性質 1 状態全体の集合 S に含まれる確率分布 μ と ν が前節の条件を満足するとき、$\mu \geq \nu$ ならば、$\overline{w}_n(\mu) \geq \overline{w}_n(\nu)$ である。

参考文献

[1] F. De Vylder, Duality Theorem for Bounds in Integrals with Applications to Stop Loss Premiums, *Scandinavian Actuarial Journal*, 129–147, (1983).

[2] M. Kijima and M. Ohnishi, Stochastic Orders and Their Applications in Financial Optimization, *Mathematical Methods of Operations Research*, **50**, 351–372, (1999).

[3] T. Nakai, A Sequential Stochastic Assignment Problem in a Partially Observable Markov process, *Mathematics of Operations Research*, **11**, 230–240, (1986).

[4] T. Nakai, An Optimal Selection Problem on a Partially Observable Markov process, In *Stochastic Modelling in Innovative Manufacturing*, Lecture Notes in Economics and Mathematical Systems **445**, (Eds. A. H. Christer, S. Osaki and L. C. Thomas), pp. 140–154, Springer-Verlag, Berlin, (1996).

[5] T. Nakai, An Optimal Assignment Problem for Multiple Objects per Period – Case of a Partially Observable Markov process, *Bulletin of Informatics and Cybernetics*, **31**, 23–34, (1999).

[6] T. Nakai, A Generalization of Multivariate Total Positivity of Order Two with an Application to Bayesian Learning Procedure, *Journal of Information & Optimization Sciences*, **23**, 163–176, (2002).

[7] T. Nakai, A Sequential Expenditure Problem for Public Sector Based on the Outcome, *Recent Advances in Stochastic Operations Research* (Eds. T. Dohi, S. Osaki and K. Sawaki), World Scientific Publishing, 277–295, 2007.

[8] T. Nakai, A Sequential Decision Problem based on the Rate Depending on a Markov Process, *Prodeedings of International Workshop on Recent Advances in Stochastic Operations Research II*, (Eds. T. Dohi, S. Osaki and K. Sawaki), Systems Reliability Engineering Laboratory, Hiroshima University, 171–178, 2007.