緩い対称性モデルにおける不確定情報の扱い
How Does Loosely Symmetric Model Treat Indefinite Information?

甲野 佑[†], 高橋 達二[‡]
Yu Kohno, Tatsuji Takahashi

[†] 東京電機大学大学院, [‡] 東京電機大学理工学部
School of Science and Engineering, Tokyo Denki University
yu.kohno.02@gmail.com

Abstract

We analyze the loosely symmetric model proposed by Shinohara. We show that the model reflects not only fundamental properties of human cognition, but also information theoretic and statistical treatment of uncertainty.

Keywords — Symmetry Bias; N-Armed Bandit Problems; Indefiniteness; Decision Making; Causal Induction

1. 概要

現実に生きる我々は常にその観測，記憶，思考，行動に時間的，量的な制限を受け続けている．計算機上でランダムなシミュレーションによるモンテカルロ計画法が上手く働くのは，それが定常的な環境を仮定しているからだと考えられる．しかし現実で働く人工知能を想定する場合，その仮定が十分満たされているとは限らない．現実の環境は複雑でむしろ非定常環境にあり，環境から受ける様々な制約の中で逐次的，連続的に行動を決定し続けなければならない．

そのような基礎的な課題として2本腕バンディット問題が存在する．この研究では人の非論理的バイアスを考慮した確率モデルであるLoosely Symmetric model(以下LS)DH[1]を2本腕バンディット問題上で扱った場合での分析において判明した，不確定な事象に対する扱い方を幾つかの側面から分析する．

LSは人間の因果帰納に対して最も相関が高いとされる服部のDH[4]と比較しても高い相関を得ているのみならず，多本腕バンディット問題においても高い成績を有する唯一のモデルであり[1]，因果帰納と意思決定に対する統合理論の仮説構築にも寄与している[5]．LSの認知的な妥当性は高橋らによって地の不変性等から示されている[3]．しかしながらその式そのものに関する意味的妥当性は未だ十分に述べられていない．本研究では上述した偶然の影響を排する事が出来ない不確定，非決定的な事象への扱いという観点からシミュレーション，理論的分析からLSの妥当性を論じる．

2. 2本腕バンディット問題

本研究では2本腕バンディット問題を例に不確定な情報の扱い方，値付けを論じている．ここでの不確定な情報とは事前情報がなく観測が不十分な情報を意味する．これは強化学習課題における初期において学習を促進するためにどのような方策，価値観数を用いるかの問題に対応する[2].

2本腕バンディット問題とは意思決定課題の一種で，動物が餌獲得のため複数(ここでは2種)の餌場から餌獲得が最も期待される餌場を選択する課題に例えられる．ここで問題となるのが知識の獲得とその利用のジレンマである．現在の観測結果から最も餌を獲得出来る確率の高い餌場を選ぶとして，それが真に最も高い確率で獲得出来る餌場とは限らない．観測された情報数が少なければ少ない程，実際はよく餌を得られるはずの餌場で“ 偶々 ”餌を得らなかった等，偶然の影響が強くなる．このような場合に人間は観測された客観的な確率をそのまま用いるだろうか．経験的には人間は主観的に確率を歪め，偶然を補正するような推定を行っているのだと思われる．

この課題を数学的モデルとして扱うため，2本腕バンディット問題を試行して得られた情報を以下の2×2分割表として扱う．ここで事象A，Bは選択可能な二つの手段のうち手段A，Bをそれぞれ試行した事象である．対して事象Eは結果として報酬が得られた事象である．先ほどの例では手段A,Bが餌場に，結果Eが餌の獲得に対応する．

表 1 共変情報の2×2分割表

	E	$\bar{E}$
A	a	b
B	c	d

a：手段Aで結果Eを得た回数
b：手段Aで結果Eを得なかった回数
c：手段Bで結果Eを得た回数
d：手段Bで結果Eを得なかった回数

表の頻度情報，a，b，c，dは試行総数$N = a+b+c+d$で除算する事により，結合確率分布(表2)として表す事が出来る．この事から手段A，Bを行っ

て結果Eが得られた確率が以下の式1，2で定義される．

表 2 結合確率分布

	E	$\bar{E}$
A	$P(A,E)$	$P(A,E)$
B	$P(B,E)$	$P(B,\bar{E})$

$$P(E|A) = \frac{P(E,A)}{P(A)} = \frac{a}{a+b} \quad (1)$$

$$P(E|B) = \frac{P(E,B)}{P(B)} = \frac{c}{c+d} \quad (2)$$

以降，式1，2で挙げられた条件付き確率を以下CPと呼び客観的な確率モデルとして扱う．

3. Loosely Symmetric model

本研究で主に扱うLSは篠原によって考案された信念の度合いを計算する確率モデルである[1]．LSは人の非論理的なバイアスを記述する対称性と相互排他性バイアスを緩く満たすモデルであり以下の式で表される．

$$LS(E|A) = \frac{a+\frac{b}{b+d}d}{a+\frac{b}{b+d}d+b+\frac{a}{a+c}c} = \frac{P(A,E)+S_p}{P(A,E)+S_p+P(A,\bar{E})+S_n} \quad (3)$$

$$\text{Positive bias}: S_p = P(\bar{E})P(A|\bar{E})P(B|\bar{E}) \quad (4)$$

$$\text{Negative bias}: S_n = P(E)P(A|E)P(B|E) \quad (5)$$

人の思考には“ $p \rightarrow q$ ”という命題が真であるとき，“ $q \rightarrow p$ ”もまた真であると思い込む非論理的なバイアスが存在する．これが先に述べた“ 対称性バイアス ”であるまた，同じように“ $\bar{p} \rightarrow \bar{q}$ ”を真であると思い込む性質を“ 相互排他性バイアス ”という．これらは論理学において誤りだが，現実にはそのように考える事がより適切な場合もある[5]．対称性，相互排他性バイアスを完全に満たすモデルとして以下のRSが考案されている．

$$RS(E|A) = \frac{a+d}{a+d+b+c} = P(A,E)+P(B,\bar{E}) \quad (6)$$

しかし，人が常にそのようなバイアスを完全に働かせているとは考え辛く，むしろ緩く柔軟に満たすものだと考えられる．LSは非対称なCPと完全対称性なRSモデルの中間，緩く対称性を満たす数学的モデルである．また，LSは条件付き確率に対するバイアスの値(式4,5)を定数で調節するのではなく，地の不変性から導く事の出来る関数を用いる事で，対称性，相互排他性バイアスを柔軟に変化させている[3]．この式は人の因果帰納実験に対して高い相関を得ており[1]，複雑な多本腕バンディット問題においても高い成績を有している[7]．またモンテカルロ計画法や，単純な追跡問題においても優れた結果を残している[8]．

4. 偶然と必然

上述のようにLSの有用性は諸問題に対する成績から正当化されるが，その理論的な分析は未だ不十分であった．特にバイアス項(式4,5)が何故このような式になるのか強い理由が示されていない．この章では不確定の解釈として用いられる偶然という概念から，バイアス項と平均情報量，分散との関係性を論じる．

4.1 バイアス項と平均情報量

LSのポジティブなバイアス項である式4は“ 試行全体において報酬を得なかった確率 ”である$P(\bar{E})$と，確率の偏りから値の大きさを制御する項$P(A|\bar{E})P(B|\bar{E})$からなる．即ちLSは大まかに言って，現在得られている報酬確率が低ければ“ 偶然報酬が得られてない ”と考え評価を上げ，ネガティブなバイアス項(式5)であれば逆に“ 偶然報酬を得過ぎている ”として評価を下げる性質がある．

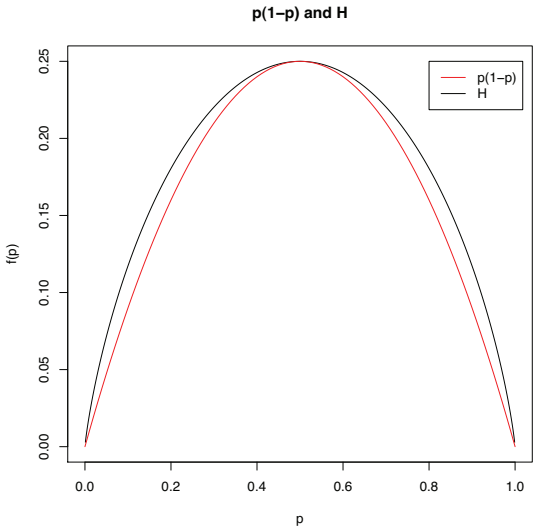

図 1 平均情報量への近似

この効果を偶然の補正と捉えるならば，試行を重ねれば減少するべきである．LSの制御項$(P(A|\bar{E})P(B|\bar{E}))$は“ 当たった時，どの手段を行っていたか ”という報酬から選択への確率$p = P(A|\bar{E})$で記述され，pに対する平均情報量の4分の1倍に近似される（表1）．平均情報量の特性から制御項は逆確率が0.5のとき最大となり，最も偏ったとき

最小となる．LSのバイアス項から得られる偶然への補正はその偏りによって試行回数を表現し，効果の減少を制御していると言える．

4.2 標本平均とその分散

ベルヌーイ試行で得られた標本の平均は中心極限定理により正規分布に近似される．その分散は試行回数nが増える程に小さくなって行く．言い換えれば標本平均とはその試行によりある結果が生じる確率pであり，その分散は誤差である．この考えは前述した我々の直感的な偶然の扱い，そしてLSモデルのバイアス項の挙動と一致している．実際に“ 報酬を獲得した回数 ”$a+c$をnとし，報酬を得た時，それが“ 手段Aを試行していた確率 ”$P(A|\bar{E})=a/(a+c)$をpとした時，二項分布の分散$np(1-p)$とLSのネガティブなバイアスと一致する(式7).

$$np(1-p)=(a+c)\frac{a}{a+c}(1-\frac{a}{a+c})=\frac{ac}{a+c} \qquad (7)$$

このように分散の概念からもLSがバイアス項により誤差を補正して真の確率を推定していると考えられる．

4.3 認知的事象歪曲

高橋らはLSのバイアス項の値が全ての手段に対して一定である事を示した．それは視覚における図と地の関係に例えられ認知的に正当化される．即ちLSは注目対象を図として正確に捉えるが，それ以外の非注目対象を地として図がなんであるかに関わらず同様に曖昧に捉える性質を持つ．これを地の不変性としている[3].

LSは上述した平均情報量等から結合確率分布(表2)を歪めて表3を生成し，対称性，相互排他性バイアスから評価値を導いていると考えられる．これは環境から曖昧な領域として，視覚等における地(Ground)の確率変数Gを新たに生成しているに等しい（図2）.

表 3 歪められた結合確率分布

	E	$\bar{E}$
A	$P(A,E)$	$P(A,\bar{E})$
$Ground$	S_n	S_p

これは全事象Uそのものを変化させる事で，注目する事象に対して主観的な確率を導いていると解釈出来る(図3)．人は地の不変性から，全ての事象を常に正確には扱っていないと考えられる．LSはこのような性質から対称性，相互排他性バイアスを緩く調節していると考えられる．

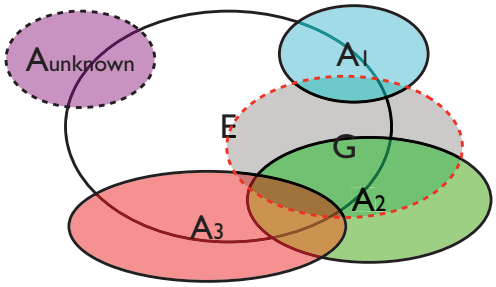

図 2 確率変数Gの生成

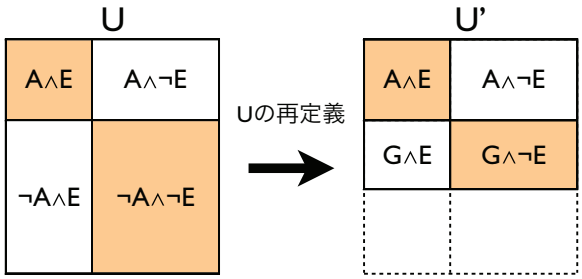

図 3 全事象Uの歪曲

5. LS policyと環境適応

意思決定課題である2本腕バンディット問題においては意識的に変更出来る変数とそうでない値が存在する．手段A（あるいはB）を行い報酬を得られた割合（$P(A|\bar{E})$, $P(B|\bar{E})$）はいくら試行を行っても任意に確率を変える事は出来ない．意思決定者が直接変化させられるのは“ 手段A（あるいはB）を行った割合 ”である$P(A)$のみである．

表 4 2本腕バンディット問題の変数分類表

意思決定変数	環境変数	目的変数
$P(A)$	$P(A|E)$,$P(A|\bar{E})$	$P(E)$
$P(B)$	$P(B|E)$,$P(B|\bar{E})$	$P(\bar{E})$

ここで手段Aの真の報酬確率をP_A，手段Bの真の報酬確率P_Bとする．$P_A>P_B$，即ち手段Aが最も効率的な手段である場合，以下の式が成り立つ.

$$\lim_{P(A)\to 1.0} P(E)=P(E|A)\approx P_A \qquad (8)$$

即ち，バンディット問題とは選択の割合を$P(A)$もしくは$P(B)$に偏らせる事で報酬獲得割合$P(E)$を最大化する問題だと考えられる．この事からLSをはじめとする2本腕バンディット問題のモデルは意思決定変数$P(A)$の関数と捉える事が出来る．

5.1 LSの基準点

2本腕バンディット問題は真の報酬獲得確率P_A, P_Bから環境を定義出来る．特に報酬確率の高低

で図4のように環境を大別する事が出来る[7].

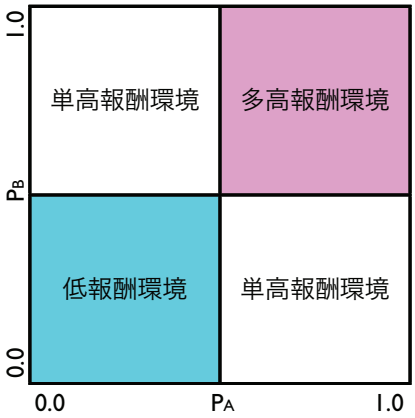

図 4 報酬獲得確率からなる問題環境の区分

以下はこの分類に合わせ，単高報集環境($P_A = .8, P_B = .2$)，多高報集環境($P_A = .8, P_B = .7$)，低高報集環境($P_A = .3, P_B = .2$)の環境に分けてLSの挙動の変化を論じる．

図5はある時点で観測された環境において，意思決定変数$P(A)$の値からどのように評価値が変化するか，LSを$P(A)$の関数として捉えた図である．

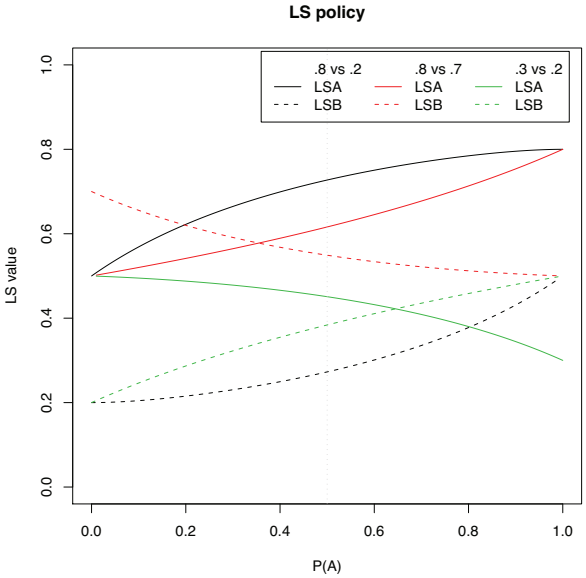

図 5 意思決定変数$P(A)$に対するLSの値

この関数においては意思決定変数$P(A)$に対して評価値が逆転する交点の有無が客観的な確率に逆らい探索を行うかを決定する．環境観測変数を固定した時，単高報酬環境(どちらか一方の手段のみが報酬獲得確率が0.5以上)という易しい問題環境では交点が存在せず探索を行わない．対して多高報酬環境，低報酬環境のような難しい問題環境では交点が現れ探索するようになる．

また，$P(A) = 0$，即ち手段Aを全く行っていない場合には，その手段に対する評価はどの環境においても$LS(E|A) = 0.5$に収束している．因果帰納におけるLSの0.5は人が無相関と感じる規準が一致する[6]．逆に$P(A) = 1.0$に近づけば近づく程値が頻度的に観測された客観的な条件付き確率$P(E|A)$に近似される．即ち，LSにはあまり観測されていない未知の手段を無相関規準である0.5に値付けし，十分観測された手段に対しては客観的な値に収束する9.

$$\lim_{P(A)\to 1.0} LS(E|A) = P(E|A) \qquad (9)$$
$$\lim_{P(A)\to 0.0} LS(E|A) = 0.5$$

図5，LSの極限(式9)から，$P(E|A)$の値が0.5を境に$LS(E|A)$と$LS(E|B)$のLSの評価が逆転し，低報酬獲得環境において探索をより行う事にが示された．この事からLSは“ 少なくともいずれか一つの手段は0.5以上の報酬確率を持つであろう ”というポリシーを持ち，報酬確率が0.5以上の手段を求めて探索を行うと解釈出来る．

5.2 選択バイアスの状態遷移

LSには低報酬環境ではよく探索し，多高報得環境(どちらの手段も報酬獲得確率が0.5以上)では報酬を最大化させる傾向がある．これは環境に適応するよう緩やかな状態遷移を行っていると言える(図6)．このようにLSは一つの価値観数でありながら，環境の状態に応じて方策を複数種類有している．

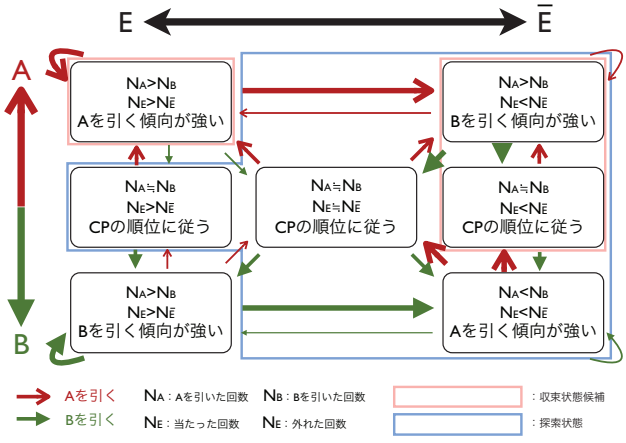

図 6 LSの方策状態遷移

2本腕バンディット問題で優秀とされる$UCB1$と比べても多数の状態を表現しており，また$UCB1$は試行回数が増える程，方策が客観的な評価に収束する．LSは試行数が増えても上述したような柔軟な状態変化を行う能力を失わないため，複雑

に変化する環境に対しても柔軟な判断を行う事が可能である．

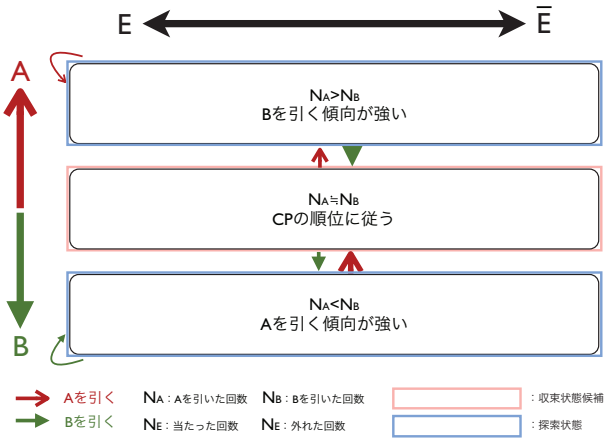

図 7 $UCB1$の方策状態遷移

6. 総合考察

以上によりLSは環境の複雑さを地の不変性，平均情報量，分散からなるバイアスの補正項という形で値に取り入れ，異なる環境毎にその挙動を変化させて不確定な情報を扱っていると考えられる．LSにおいて手段の大まかな評価は手段と報酬間の因果関係で表され，0.5という規準は試行による損失と利益に対する相関の境界を示しているのだと解釈出来る．

また，LSはある種の主観確率を記述すると考えられる．主観確率とは認知を構成する上で最も基礎的なカテゴリーに位置する．論理で扱えない矛盾を省くだけでなく，確率においても初期において近似的な解を効率的に与え，環境変化に対しても高い適応性を示す．曖昧さをどのような扱うべきかという議論は今までも行われてきた．ファジィ集合等で扱われるファジィ理論はその代表例ともいえる．それは人間の離散的だが定性的な判断を定量化し，論理に持ち込もうというアプローチから生まれた．それに対してLS等の主観確率は，確率という定量的な判断に定性的な性質を与えようというアプローチである．

主観確率モデルであるLSは確率の公理を満たすものである．ファジィ集合にはいくらか特殊な規定を行う必要があるのに対して，LSは確率を扱うシステムであれば大きな変更なく扱える．前述のとおりLSは情報獲得が制限される場合に有用に働き，また情報が十分に獲得できている場合には客観的な確率と等しく収束する．即ち，広い分野で単純に確率の表記をLSやその改良モデルに変更するだけで複雑な環境に適応する主観確率を扱えるようになると考えられる．

このようにLSは偶然の影響，不確定な情報に適応する人間の基本的認知能力を表現する良いトイモデルとなる可能性がある．この点については意思決定係数の分離による分析(表4,図5)等も行い，今後，確率に対する人間の主観的な値付けを実験的に研究して行く．

参考文献

[1] 篠原修二, 田口亮, 桂田浩一, 新田恒雄 (2007) “因果性に基づく信念形成モデルとN本腕バンディット問題への適用,” 人工知能学会論文誌, Vol. 22, No. 1, pp. 58–68.

[2] Sutton, R. S., Barto, A. G. (2000) 強化学習. 森北出版. (三上, 皆川 訳)

[3] Takahashi, T., Nakano, M., Shinohara, S. (2010) “Cognitive symmetry: Illogical but rational biases,” *Symmetry: Culture and Science* , Vol. 21, No. 1-3, pp. 275–294.

[4] Hattori M., Oaksford M. (2007) “Adaptive non-interventional heuristics for covariation detection in causal induction: Model comparison and rational analysis,” *Cognitive Science*, Vol. 31, No. 5, pp. 765–814.

[5] 大用庫智,高橋達二 (2010) “因果帰納と意思決定を結ぶ緩い対称モデル ,”日本認知科学会第*27*回大会発表論文集, pp. 799–800.

[6] Tatsuji Takahashi, Kuratomo Oyo, Shuji Shinohara : “ A Loosely Symmetric Model of Cognition ”, *Lecture Notes in Computer Science*, No. 5778, Springer, pp. 234–241 (2011).

[7] 大用 庫智, 甲野 佑, 高橋 達二 (2011) “n本腕バンディット問題に対する LS モデルの一般化, ” *2011*年度人工知能学会全国大会*(*第*25*回*)*予稿集, 1G1-2in.

[8] 清水 隆宏, 横川 純貴, 甲野 佑, 高橋 達二 (2011) ,“ , 認知バイアス調整機構 LS の Q 学習への実装とその機能 ”, *2011*年度人工知能学会全国大会*(*第*25*回*)*予稿集, 1P2-12in.