試行回数を考慮した不確実性下での意思決定モデルの検討
Experimental investigation of decision-making models under uncertainty, considering the number of trials.

石倉 圭悟 †, 横須賀 天臣 †, 中村 紘子 †‡, 高橋 達二 †
Keigo Ishikura, Takaomi Yokosuka, Hiroko Nakamura, Tatsuji Takahashi
†東京電機大学, ‡日本学術振興会
Tokyo Denki University, Japan Society for the Promotion of Science
tatsujit@mail.dendai.ac.jp

概要

本研究では同期待値で試行回数が異なる選択肢の選好を問う二者択一課題において, 試行回数が人間の選択に与える影響について検討した. 本研究の参加者はスロットマシンの期待値が小さい時は試行回数が少ない選択肢を, 大きい時は試行回数の多い選択肢を選好した. また, 実験結果から RS モデル, Q 学習モデル, IBL モデルのパラメータ推定し, モデルの予測と実験結果の比較を行った. その結果, Q 学習モデルの予測が最も良い結果を示した.

キーワード：ギャンブル，二者択一課題，モデル比較，強化学習

1. 序論

人間の意思決定研究では, 主に記述形式の課題と経験形式の課題という 2 つのアプローチが用いられてきた. 記述形式の課題とは, 選択肢が引き起こす結果やその結果が起こる確率が明示的に記述された課題である. 記述形式の課題により, 人々が損失を回避しようとする傾向（損失回避）や, 小さな確率を過大評価し, 大きな確率を過小評価する傾向（確率加重関数）が示された.

経験形式の課題は, 選択肢の結果や確率が明示的に提示されない. 代わりに, 意思決定者は選択肢を繰り返し試行することで, その特性を学習する. 具体的には, コンピュータ画面上の複数ボタンから選択し, 結果をサンプリングしていく. この過程から, 各選択肢の結果分布を経験的に理解し, 最終的な意思決定を行う. 経験形式の課題は現実世界の多くの意思決定状況により近いと考えられており, 記述形式と経験形式での意思決定のギャップが示されている (Hertwig & Erev, 2009) . 例えば, 記述形式では稀少事象の影響が過大評価される傾向があるのに対し, 経験形式では過小評価される傾向が見られる.

従来の意思決定研究では, 主に期待値の異なる選択肢間の比較が行われてきた. そのため, 期待値は同じでも試行回数が異なる選択肢（例：10 回中 1 回当たったスロット, 100 回中 10 回当ったスロット）間でどちらが好まれるかについては, 十分な検討がなされていない. 横須賀ら（横須賀・石倉・中村他, 2023）は, こうした期待値が同じで試行回数が異なる選択肢では, 期待値が正の場合は試行回数が多い選択肢, 期待値が負の場合は試行回数が少ない選択肢が選ばれやすいことを示した. この結果は RS (Risk-sensitive Satisficing)（高橋・甲野・浦上, 2016）と呼ばれるモデルの予測を支持するものであった. しかし, こうした課題において記述と経験のギャップが生じるかも明らかではない.

本研究では記述と経験の両方の性質を持つ課題を用いて, 試行回数が人間の意思決定に与える影響を検証するとともに, 記述と経験のギャップが生じるか, どのようなモデルが試行回数の異なる選択肢間の選好をよく予想できるかを検討する. 課題では, 表 1 のようにスロット A と B の結果が逐次的に提示され, 参加者にどちらのスロットが良いかを選択するよう求める. この課題は, どのような選択がされたかが逐次的に提示されるという経験形式の課題と, 結果が明示されるという記述形式の課題の双方の特徴を持っているといえる.

表 1　二者択一課題の提示例

順番	選択	スロット A	スロット B
1	B		アタリ ＋ 1 万円
2	B		ハズレー 1 万円
3	A	ハズレー 1 万円	
4	・	・・・・・	・・・・・

2. 意思決定モデル

本研究では, RS (Riskk-sensitive Satisficing) モデル, Q 学習モデル, IBL (Instance-based learning) モデル (Lejarraga & Dutt & Gonzalez, 2012) の 3 つの意思決定モデルを用いて実験結果のシミュレーションを行った.

2.1 Risk-sensitive Satisficing (RS) モデル

RS モデルは Simon が提唱した満足化の原理 (Simon, 1956) に基づいており, 選択肢が目標とされる希求基準を満たすか否かにより, 探索を行うか活用を行うかを決定する逐次的意思決定モデルである.

RS 価値関数は期待値 E_i と希求水準 $\aleph$ の差である δ_i と, その係数である信頼度によって表現される. δ_i は式 (1) で与える.

$$\delta_i = E_i - \aleph \tag{1}$$

RS 価値関数は式 (2) で与える.

$$\mathrm{RS}_i = \frac{n_i}{N}\delta_i = \frac{n_i}{N}(E_i - \aleph) \tag{2}$$

ここで n_i は行動 a_i を選択した回数であり, N は総試行回数である. タイムステップ t での選択肢 i の期待値である期待値 $E_{i,t}$ は以下の式 (3) で計算することができる.

$$E_{i,t} = E_{i,t-1} + \frac{r_t - E_{i,t-1}}{t} \tag{3}$$

r_t はタイムステップ t で得た報酬である.

RS モデルは式 (2) で計算した RS_i 値が最大の選択肢 i を選択する.

RS モデルに基づく予測： RS モデルは, 満足状態と非満足状態で異なる振る舞いを示す.

- 満足状態（少なくとも 1 つの選択肢 i の δ_i が正）: RS 値が最も大きい選択肢の期待値が他の選択肢より小さくなるまで選択を続ける. 期待値が等しい選択肢間では, 試行回数が多い選択肢を選択する（リスク回避）.
- 非満足状態（すべての選択肢 i の δ_i が負）: 基準値 $\aleph$ を超える期待値を持つ選択肢を発見するまで探索を続ける. 期待値が等しい選択肢間では, 試行回数が少ない選択肢を選択する（リスク志向）.

2.2 Q 学習モデル

Q 学習モデルは古典的条件づけのモデルである Rescorla-Wagner モデルを拡張した強化学習モデルである. 今回のシミュレーションでは状態遷移を考えないため以下の式 (4) で更新される Q 値を用いて意思決定を行う.

$$\mathrm{Q_{t+1}(a_t)} = \mathrm{Q_t(a_t)} + \alpha(r_t - \mathrm{Q_t(a_t)}) \tag{4}$$

r_t はタイムステップ t で得た報酬である. 式 (4) の α は学習率であり, 新しい経験から Q 値をどの程度更新するかを決定するパラメータである. 本研究では二者択一課題を用いるので選択肢を A, B とすると行動選択確率は以下の式 (5) に従って計算される.

$$P(a_t = A) = \frac{\exp(\beta \cdot Q_t(A))}{\exp(\beta \cdot Q_t(A)) + \exp(\beta \cdot Q_t(B))} \tag{5}$$

式 (5) の β は選択確率に影響を与えるパラメータである.

Q 学習モデルに基づく予測： Q 学習モデルでは学習率に基づいて更新された Q 値を用いて次の行動を選択する. 本実験で用いる課題の期待値は選択肢間で等しいため Q 値はほぼ同じ値として計算される. そのため, β が大きいほど高い Q 値を持つ選択肢を β が小さいほどランダムに選択を行う.

2.3 Instance-based learning (IBL) モデル

二つのギャンブルの内どちらを好むかといった, 経験形式の二者択一問題に対して一般的に用いられる意思決定アルゴリズムとして IBL モデルが提案されている.

IBL モデルは各選択肢における V 値を式 (6) を用いて計算し最も大きい選択肢を選択する. V 値はその選択肢が取りうる n 個全ての結果について, それぞれ結果の大きさである x_i とその結果が起きる確率 p_i の積の合計として計算される.

$$V_j = \sum_{i=1}^{n} p_i x_i \tag{6}$$

繰り返しの意思決定課題におけるタイムステップが t である時の結果が起きる確率 $p_{i,t}$ は式 (7) で定義される.

$$p_{i,t} = \frac{e^{A_{i,t}/\tau}}{\sum_j e^{A_{j,t}/\tau}} \tag{7}$$

$A_{i,t}$ はタイムステップ t での選択肢の i 番目の結果の確率の重みづけに使われる変数である. $A_{i,t}$ はその出来事をどれだけ活性化させるかの変数である.

$$A_{i,t} = \sigma \ln\left(\frac{1-\gamma_{i,t}}{\gamma_{i,t}}\right) + \ln\left(\sum_{t_p \in \{1,\ldots,t-1\}} (t - t_p)^{-d}\right) \tag{8}$$

式 (7) における $\tau = \sigma$・2 でありランダムノイズである. σ は自由なノイズパラメータである. 式 (8) における γ は 0 から 1 までの一様分布からサンプリングされた値であり, d は自由減衰パラメータである. t_p は結果 x_i が起きた一連のタイムステップである. 式 (7) では人間が記憶から思い出す時の不確実性を表現しており, 式 (8) では記憶の活性化が経験の頻度と親近性に影響を受けていることを表現している. また, IBL モデルは最初のタイムステップでは default_utility (U_d) と呼ばれる事前に設定される V 値を持つ. この U_d は各選択肢の期待値よりも大きい値を持つハイパーパラメータであり, 探索を促す効果を持つ.

IBL モデルに基づく予測：繰り返しの二者択一問題に対して IBL モデルは選択肢が取りうる結果と, そ

の結果が起きる確率の積の合計が最も大きいものを選択する．選択肢の期待値が同じである時, その選択時に結果が起きた回数が多くタイムステップが現在に近いほど, 計算される結果が想起される確率値が大きくなり, その選択肢が選ばれやすくなる．

3. 実験

3.1 参加者

実験は Web 上で実施し, クラウドソーシングサービス (CrowdWorks: https ://crowdworks.jp/) を用いて, 実験参加者を 300 名募集した．回答に不備のあるデータを除外し，244 名 (女性 103 名，男性 141 名，未回答 0 名，平均年齢 42.8 歳，標準偏差 9.45) を分析に使用した．

3.2 手続き

課題は Qualtrics (https://www.qualtrics.com) を用いて提示し, はじめにインフォームドコンセントを行い, 実験参加者に同意を得た．続いて, 二者択一課題の練習課題 1 問を行った後, 本課題 9 問を行った．最後に, 参加者が課題に注意を向けていたかを確認するために Instructional Manipulation Check (Oppenheimer & Meyvis & Davidenko, 2009) を行なった．

3.3 実験材料

二者択一課題は, 表 1 で示すような経験形式と記述形式の両方の性質を持つような意思決定課題であり, 課題では期待値が同じスロットマシン A と B の 2 台について, A を 8 回プレイした結果と B を 24 回プレイした結果の, 合計 32 試行をランダムな順序で提示した．スロットマシンは「アタリ」が出た場合は 1 万円を獲得し，「ハズレ」が出た場合は 1 万円を損失するとした．その後, 参加者には 1 回だけプレイするとしたら, A と B のどちらを選択するかを尋ねた．スロットマシンの当たり確率は 0 から 1 まで 0.125 刻みで 9 段階を設けており, 期待値は - 1 万円から 1 万円まで 2500 円刻みで 9 段階とした．それぞれの確率（期待値）ごとに二者択一課題を作成し, 9 種類の課題をランダムな順で参加者に提示し, 回答を求めた．

4. 結果

4.1 行動データ

二者択一課題において, 試行回数の多いリスク回避的選択 (24 回試行) がされた割合を図 1 に示す．

二者択一課題の選択について, ベイジアン・ロジスティック回帰を行った．従属変数は選んだ選択肢とし, 基準カテゴリをリスク志向的選択肢とした．独立変数はスロットの結果（アタリ，ハズレ）, 結果の表示回数

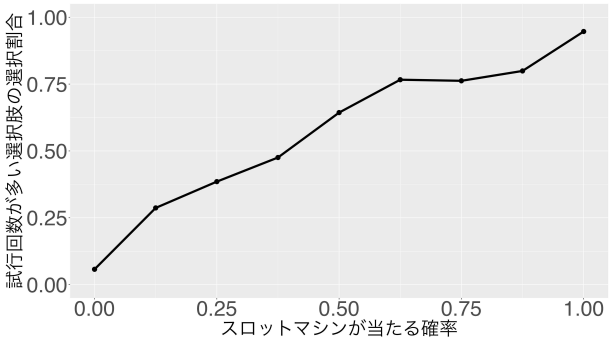

図 1 二者択一課題におけるスロットマシンの当たり確率とリスク回避的選択 (24 回試行) の選択割合

（結果が表示される個数が 11 個以上の場合は High , 結果が表示される個数が 5 個以下の場合は Low, それ以外の個数は Mid）, 提示順序（提示順序の前半 16 試行に含まれる場合は 1st, 後半 16 試行に含まれる場合 2nd）, 表の提示位置（表の左側を Left, 右側を Right）とし, 基準カテゴリは Lose, High, 1st , Left とした．また, 参加者を変量効果とした．

分析では R の brmspackage (Bürkner,2018) の無情報事前分布をもとに, 4 つのチェーンを 4000 回反復し, 最初の 2000 回はウォームアップ期間とみなして破棄した．すべての変数の $\hat{R} \leq 1.0$ となり, 分析が収束したといえる．カテゴリカル回帰分析の結果, 95% 信用区間 (credible interval) に 0 を含まなかった変数を表 2 に示す．結果と結果の表示回数の交互作用について下位検定を行った．その結果, 全ての条件の組み合わせの結果は有意であり, 表示された勝ちの回数が多い場合はリスク回避的, 少ない場合はリスク志向的な選択が増えることが示された．

表 2 回帰分析の推定値, 推定誤差, および 95% 信用区間

変数	推定値	推定誤差	95% 信用区間
Win	4.30	0.35	3.63 – 5.01
Low	2.55	0.22	2.13 – 2.99
Mid	1.51	0.26	1.02 – 2.02
Win:Low	-5.10	0.36	-5.81 – -4.42
Win:Mid	-2.09	0.42	-2.91 – -1.31

4.2 モデルのシミュレーション

RS モデル, Q 学習モデル, IBL モデルによるシミュレーションを, 以下の手順で行った：実験で参加者に提示した履歴を元に各モデルで学習を行い, 参加者と同様の提示された履歴の後にモデルがどちらの選択肢を選択するかシミュレーションを行った．各モデルでシュミレーションに用いたパラメータは, 9 課題において参加者とモデル間の試行回数が多い選択肢を選んだ割合の最小二乗誤差を目的

関数として, 目的関数を最小化するように Optuna (https://optuna.readthedocs.io) を用いてハイパーパラメータを探索した. 探索に用いた trial 数は 300 として設定した. Optuna によって求めた, ハイパーパラメータの推定値を表 3 に示す.

表 3　各モデルのパラメータの推定値

モデル	パラメータ 1	パラメータ 2	パラメータ 3
RS	$\aleph$:0.1624		
IBL	U_d:1.011	σ:0.9644	d:0.6444
Q 学習	α:0.01377	β:339.7	

表 3 のハイパーパラメータ設定時の最小二乗誤差は RS：1.028, IBL：0.6382, Q 学習：0.1413 であった. 図 2 に表 3 のパラメータを用いてシミュレーションをした結果を示す.

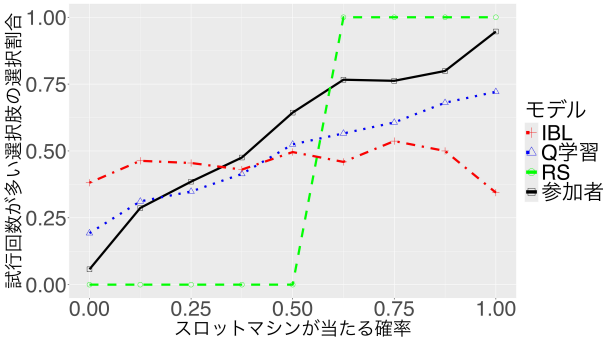

図 2　探索したハイパーパラメータを用いたシミュレーション結果

5.　考察

記述形式での先行研究と同様に, スロットマシンの期待値が高くなるにつれ試行回数が多いリスク回避的な選択が増えることが示された. これらの結果は RS モデルに基づく予測と一致していた. 本研究では, 逐次的な試行の結果を提示していたが, 最終的には全ての結果が明示的に示されていることから, 記述形式の課題と同様の結果となったと考えられる.

モデルのシミュレーション結果から, RS モデルのパラメータである希求水準 $\aleph$ の値が 0.1624 であり, 参加者はスロットマシンの当たり確率として平均して 0.1624 を期待していたことが示唆された. また, 実験結果とモデルの予測の最小二乗誤差は Q 学習モデル, IBL モデル, RS モデルの順番に大きかった. 図 2 を見ると Q 学習モデルが最もよく参加者の傾向を捉えていることがわかる. また, 推定したパラメータ β の値が 339.7 と大きな値になっていることから, Q 値が最も高い行動を高い確率で選んでいることがわかる. IBL モデルは図 2 から全ての選択においてどちらの選択肢もほぼ同じ割合で選択していることがわかる. そのため, スロットマシンの期待値に応じて選好を変化させる傾向は IBL モデルにおいて捉えられていない. 従ってモデルを改善するためには, 試行回数を考慮する必要があることが考えられる. RS モデルはスロットマシンが当たる確率 0.5 と 0.625 間で試行回数が少ない選択肢から試行回数が多い選択肢に選択を切り替えていることがわかる. しかし, RS モデルは決定論的であるため他のモデルと比較して実験結果とモデルの予測の最小二乗誤差は最も大きくなった. これらのことから, RS モデルは参加者全体の意思決定の予測に応用するためには, 満足状態と非満足状態に確率的に選択をする必要があることが示唆された.

6.　結論

本研究では試行回数が人間の選択に与える影響について検討した. 参加者はスロットマシンの期待値が小さい時は試行回数が少ない選択肢を, 大きい時は試行回数の多い選択肢を選好した. 今回の表形式の課題では最終的に全ての結果が明示的に示されていることから記述形式に近い結果となったと考えられる. そのため, 今後試行回数と課題形式についての関係を明らかにするためには, 表の履歴が残らない課題において調査をする必要があると考えられる.

また, 実験結果から RS モデル, Q 学習モデル, IBL モデルのパラメータを推定し, それぞれのモデルの予測と実験結果の比較をした. その結果, 参加者全体の意思決定傾向の予測においては Q 学習モデルの予測が最も良く, RS モデルで参加者全体の意思決定傾向の予測を行うためにはモデルの改善が必要であることが示唆された.

文献

Ralph Hertwig, Ido Erev. (2009), The description–experience gap in risky choice,Trends in Cognitive Sciences,Volume 13, Issue 12,2009,Pages 517-523,ISSN 1364-6613

高橋達二, 甲野佑, & 浦上大輔. (2016). 認知的満足化 限定合理性の強化学習における効用. 人工知能学会論文誌, 31(6), AI30-M_1.

横須賀天臣, 石倉圭悟, 中村紘子, 高橋達二, (2023), 不確実性下の意思決定におけるリスク態度と認知的満足か. 2023 年度日本認知科学会第 40 回大会. P1-064A

Lejarraga, T., Dutt, V., & Gonzalez, C. (2012). Instance-based learning: a general model of repeated binary choice. Journal of Behavioral Decision Making, 25(2), 143–153.

Simon, H. A. (1956). Rational choice and the structure of the environment. Psychological review, 63(2), 129.

Bürkner, P.-C. (2018) Advanced Bayesian Multilevel Modeling with the R Package brms, The R Journal, Vol. 10, No. 1, pp. 395–411

Daniel M. Oppenheimer and Tom Meyvis and Nicolas Davidenko, (2009), Instructional manipulation checks: Detecting satisficing to increase statistical power, Journal of Experimental Social Psychology, 45(4), Pages 867-872, 0022-1031