

2. 【研究計画】 適宜概念図を用いるなどして、わかりやすく記入してください。なお、本項目は1頁に収めてください。様式の変更・追加は不可。

(1) 研究の位置づけ

特別研究員として取り組む研究の位置づけについて、当該分野の状況や課題等の背景、並びに本研究計画の着想に至った経緯も含めて記入してください。

着想経緯：データサイエンスの発展にはベイズ手法の発展が重要

申請者は学部時代にデータサイエンティストとしてのインターンを通じ、ベイズ手法は解析結果を推定値の一点だけでなく誤差の広がりを持って視覚化できることにポテンシャルを感じた（図1参照）。

しかし、現状のベイズ推論は万人にとって使いやすいものであると言えない。後述のHMCアルゴリズムに習熟していない場合は、最適な調節を見つけるまで何度も試行錯誤を必要とすることもある。そのことから、ベイズ手法が敬遠されて安易な代替手法が選好されがちな現状には苦い思いを感じていた。

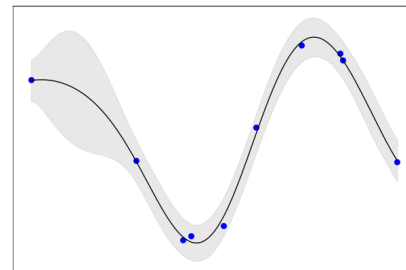


図1: ベイズ手法（Gauss 過程回帰）のプロット。推定値（実線）だけでなく95%信用区間（灰色）も自然に得られる。データが少ない領域で予測が不確実である（＝モデルに自信がない）ことがわかる。

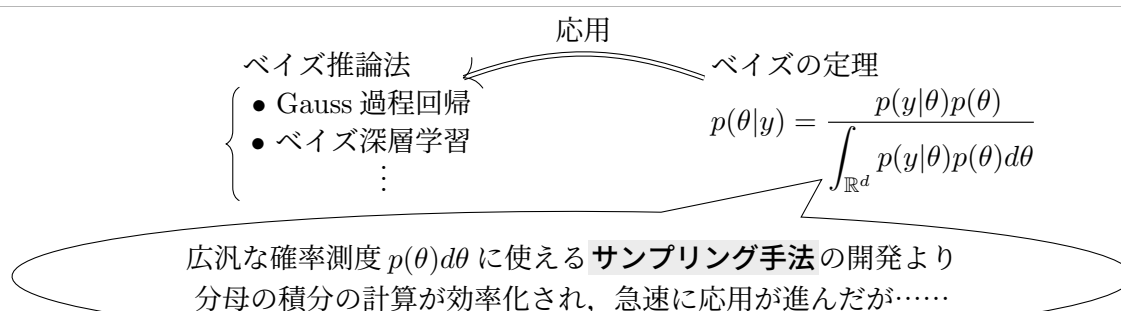
当該分野の状況：ベイズ手法の発展にはサンプリング法の発展が重要

ベイズ推論とは、機械学習や統計学において、ベイズの定理に基づく手法群の総称である。ベイズ推論は最適である（＝どの手法よりも悪くない）ことが示されている上に、事前情報を取り入れた柔軟なモデリングが可能であるため大変魅力的であるが、実際の計算が困難であることが広い応用を阻んできた。

ベイズ推論は数学的には確率測度 $P \in \mathcal{P}(\mathbb{R}^d)$ に関する積分の数値計算の問題に帰着する。決定論的な近似は高々 $d \leq 3$ までの場合にしか現実的な時間内で実行できないため、収束レートは落ちるが P に従う乱数を生成（**サンプリング**という）して Monte Carlo 推定により積分を近似する：

$$Pf := \int_{\mathbb{R}^d} f(x) P(dx) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N f(X_n), \quad X_n \stackrel{\text{i.i.d.}}{\sim} P, f \in \mathcal{L}^1(P).$$

広汎な確率分布 P に使える汎用**サンプリング手法**であるハミルトニアンモンテカルロ法（HMC）が開発[1]されてから、ベイズ手法は爆発的に応用され始めた。しかし問題は、このHMCを40年近くも改善できないまま現在でも用いている点にある。前述の通りHMCは万人に使いやすい訳ではないため、ベイズ手法の普及を妨げる要因となってしまう。



当該分野の課題：サンプリング法の発展では拡張性（スケーラビリティ）の解決が重要

課題は大きく分けて2つある。本研究（次ページ）では②に重点を置きつつ、①、②両方の解決を図る。

① **性能比較** サンプリング法の**性能比較**や理論解析が遅れている。新たな手法は次々提案されており、数値実験では多くの改善が確認されている。本当はHMCより効率的なサンプリング法はすでに得られているかもしれないが、我々はそのことさえ確認できていないのである。

② **拡張性** $d \gg 1$ の際でも広い範囲の分布 $P \in \mathcal{P}(\mathbb{R}^d)$ に汎用的に使えるサンプリング法が得られていない。現代のモデルは自然言語処理・画像解析・経済学・疫学をはじめとして殆どの領域で $d = 10^n$ ($n \geq 4$) などの極めて高次元になることが日常茶飯事である。

【研究計画】(続き) 適宜概念図を用いるなどして、わかりやすく記入してください。なお、各事項の字数制限はありませんが、全体で2頁に収めてください。様式の変更・追加は不可。

(2) 研究目的・内容等

- ① 特別研究員として取り組む研究計画における研究目的、研究方法、研究内容について記入してください。
- ② どのような計画で、何を、どこまで明らかにしようとするのか、特別研究員奨励費の応募区分（下記（※）参照）に応じて、具体的に記入してください。
- ③ 研究の特色・独創的な点（先行研究等との比較、本研究の完成時に予想されるインパクト、将来の見通し等）にも触れて記入してください。
- ④ 研究計画が所属研究室としての研究活動の一部と位置づけられる場合は申請者が担当する部分を明らかにしてください。
- ⑤ 研究計画の期間中に受入研究機関と異なる研究機関（外国の研究機関等を含む。）において研究に従事することも計画している場合は、具体的に記入してください。

(※) 特別研究員奨励費の研究期間が3年の場合の応募総額は（A区分）が240万円以下、（B区分）が240万円超450万円以下（DC1のみ）。2年の場合は（A区分）が160万円以下、（B区分）が160万円超300万円以下。1年の場合は（A区分）が80万円以下、（B区分）が80万円超150万円以下。（B区分については研究計画に必要な場合のみ記入）

研究目標：次の2点を兼ね備えた拡張性に優れたサンプリング手法の開発と実装を目指す

- (i) **モデル拡張性**：高次元空間上の多峰性分布 P でも精度と速度が落ちないこと
- (ii) **データ拡張性**：大規模なデータでも計算時間が爆発せず、現実的な時間内で実行可能であること

研究の特色とインパクト：サンプリングに注目することにより、モデル中心のパラダイムに資する

本研究では変分推論アルゴリズムや、具体的なモデル $\{P_i\} \subset \mathcal{P}(\mathbb{R}^d)$ に特化した手法よりもむしろ、**モデルに依存せず普遍的に使える汎用サンプラーの構成をめざす**のが特徴である。

種々のモデルに汎用的に使える手法の発展は、ベイズ手法が広く応用されるために必要不可欠な要素である。

ベイズ統計の悲願は、解析者が具体的な推論手法に囚われず、**モデルの設計に集中出来る枠組み**の構成にある。

	ベイズ手法	従来手法
アイデア	モデルベース	推論ベース
推論エンジン	サンプリング法	推論手法ごとに構成
長所	推論手法が統一なのでモデリングに集中可能	専用に作っているのでアルゴリズムが高速

また、本研究では**開発したサンプラーを Python パッケージなどの形で一般公開**することも目標である。本研究により大規模データでも効率よく推論できる手法が、非専門家でも簡単に利用できるようになれば、ベイズ手法を複雑な実世界の現象に応用することが、産学両面で促進されるだろう。

本研究は上述の研究目標に向けた研究計画 ①, ②, ③ の **3部構成**からなり、各アプローチに1年ずつかける。

研究計画①：2つの有望な方向の検討と限界の明瞭化

現状、**拡張性**を持つ次世代のサンプリング法として、次の2つが候補に挙がっている：

(1) 区分確定マルコフ過程 (PDMP)

ランダムな時刻にランダムに変化する以外は確定的な動きをする連続時間マルコフ過程を用いて空間を探索する手法群

(2) 確率的勾配マルコフ連鎖モンテカルロ法 (SG-MCMC)

サブサンプルから計算した対数尤度の勾配の推定量を用いてエルゴード的なマルコフ連鎖を構成する MCMC 手法群

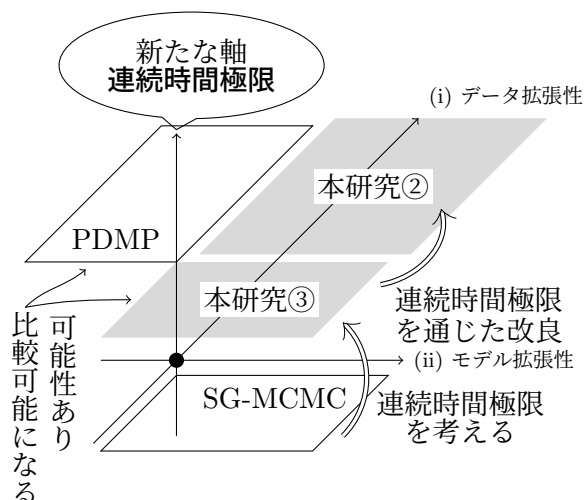
そこで**本研究はこの2つの手法を検討するところから始める**。いずれの手法も **2つの拡張性** (i),(ii) のうち片方を満たさないため（右図参照）、**その弱点をどのように克服できるかの限界を探る**。

(ii) データ拡張性		(i) モデル拡張性
PDMP	本研究	
従来法 (HMC など)	SG-MCMC	

PDMP もバイアスを導入しない効率的なサブサンプリングが可能で **(ii) データ拡張性** の条件を満たし、高次元での収束速度も期待されている [2] が、適用可能なモデルが限られており **(i) モデル拡張性** が課題である。一方で、SG-MCMC は勾配さえ推定可能であれば使える手法であり適用可能なモデルは広いが、高次元空間での収束レートが速くなく、**(ii) データ拡張性** が未だ達成されていない。

本研究では、前述の2つの手法の強みを組み合わせる。その際、数学的に自然な拡張として**連続時間極限の性質をアルゴリズムに取り入れることを考える**。**連続時間極限の手法改善・比較への利用は本研究の大きな独自性である**。

従来殆どのサンプリング法は離散時間ベースである。だが、SG-MCMCは拡散過程の離散化であり、PDMPも従来法の連続時間極限から着想された[3]。そこで数学的には、**連続時間極限こそがサンプリング手法の数学的本体であるのではないか**という理解がありえる。しかし、サンプリング法の連続時間極限を通じて比較を試みる研究はSMC法に対する1例[4]しかなく、それも拡散過程に問題を絞っている。



研究計画②：連続時間極限をキーワードに既存のアルゴリズムを改良する

サンプリング法には複数の極限が考えられ、サンプルサイズ無限大 $N \rightarrow \infty$ や実行時間無限大 $T \rightarrow \infty$ の極限はよく調べられてきたが、**タイムステップ極限 $\Delta t \rightarrow 0$ の連続時間極限**は、確率過程の収束と連続時間のエルゴード理論という最先端の数学が必要である。その数学的な困難さの割に、他の極限 $N \rightarrow \infty, T \rightarrow \infty$ に比べてアルゴリズムの性能に対する直接的示唆が得にくく、見過ごされてきたものと思われる。

しかし、**連続時間極限**から得られる Poisson 測度から、アルゴリズムの最適なイベント時刻（方向転換、リサンプリングなど）への示唆を得るというアイデアは極めて普遍的なポテンシャルを持つと期待できる[3, 4]。加えて、連続時間確率過程の方が幾何学的な発想からアルゴリズムのダイナミクスを解析しやすく[5]、より効率的な動きを導入することで新たな**サンプリング法のアイデアが得やすい**と期待される[6]。

研究計画③：連続時間極限をキーワードに提案アルゴリズムを性能比較する

研究計画②では、**連続時間極限**を一種の理想極限とみなし、最適なアルゴリズム設計について示唆を得ることを目指した。同様に、理想極限を性能比較に用いることも有望なアイデアだと思われる。

「研究の位置づけ」で述べた通り、従来手法を改良したアルゴリズムが提案されても、理論的な性能比較が進まなければ、従来手法に慣れた者を転向させるには至らないことも多い。そのため、新手法の提案と同時に性能比較の新たな枠組みを提供できたならば、**分野全体に大きく貢献できる**。従って、本研究が狙う**連続時間極限**に基づくサンプリング手法の創出は、分野に大きく貢献できるポテンシャルを持っている。

在外研究の計画：確率過程論と機械学習数理との交流

本研究では海外研究グループとの交流が重要になると予想される。実際、研究計画②の一部は、申請者の MLSS2024 でのポスター発表（業績1）から始まった Omar Chehab 氏との**ディスカッションを通じて着想**された。その際、彼の所属する CREST-ENSAE の研究室では、確率過程論のグループと機械学習数理のグループの稀有な交流が見られ、近年知見の交流が進んで結果[5]が生まれたのだということを教わった。そこで、研究計画の後半へ移る段階で CREST-ENSAE での滞在研究を計画している。

サンプリング手法の発展は広くベイズ推論に貢献し、ベイズ推論の発展は広く統計と機械学習に貢献する。そのため、統計と機械学習の極めて広い分野の研究者がサンプリング手法の発展状況に強い興味を抱いていることが、修士を通じた発表・研究経験からわかった。

本研究計画では、着想からそうであったように**広い分野の研究者との交流**が思わぬ着想をもたらすことが多いと予想される。そのため、本研究では数学だけでなく広い応用分野（統計関連学会連合大会など）での学会発表や、国内だけでなく**国際学会（数学の ICIAM や機械学習の ICML など）での発表も目指す**。

参考文献 (1) Duane, S., et al. 1987, Physics Letters B, 195, 216 (2) Fearnhead, P., et al. 2018, Statistical Science, 33, 386 (3) Peters, E. A. J. F. & de With, G. 2012, Physical Review E, 85 (4) Fearnhead, P., et al. 2017, Continuous-time Importance Sampling: Monte Carlo Methods which Avoid Time-discretisation Error (5) Chopin, N., et al. 2024, A connection between Tempering and Entropic Mirror Descent (6) Bierkens, J., et al. 2020, The Boomerang Sampler

3. 人権の保護及び法令等の遵守への対応

本項目は1頁に収めてください。様式の変更・追加は不可。

本欄には、「2. 研究計画」を遂行するにあたって、相手方の同意・協力を必要とする研究、個人情報の取り扱いの配慮を必要とする研究、生命倫理・安全対策に対する取組を必要とする研究や安全保障貿易管理を必要とする研究など指針・法令等（国際共同研究を行う国・地域の指針・法令等を含む）に基づく手続が必要な研究が含まれている場合、講じる対策と措置を記入してください。

例えば、個人情報を伴うアンケート調査・インタビュー調査、行動調査（個人履歴・映像を含む）、国内外の文化遺産の調査等、提供を受けた試料の使用、侵襲性を伴う研究、ヒト遺伝子解析研究、遺伝子組換え実験、動物実験、機微技術に関わる研究など、研究機関内外の情報委員会や倫理委員会等における承認手続が必要となる調査・研究・実験などが対象となりますので手続の状況も具体的に記入してください。

なお、該当しない場合には、その旨記入してください。

本研究の遂行上最も関連するものは個人情報の取り扱いについてであるが、アンケート調査・インタビュー調査や行動調査等を実行することも極めて考えにくく、該当しないと言って良いと思われる。

4. 【研究遂行力の自己分析】 各事項の字数制限はありませんが、全体で2頁に収めてください。様式の変更・追加は不可。

本申請書記載の研究計画を含め、当該分野における(1)「研究に関する自身の強み」及び(2)「今後研究者として更なる発展のため必要と考えている要素」のそれぞれについて、これまで携わった研究活動における経験などを踏まえ、具体的に記入してください。

(1) 研究に関する自身の強み

先述の【研究計画】を遂行するために重要な資質は、以下の通りである：

- (a) 統計計算：効率的なサンプリング手法を創出するためのアルゴリズム論・計算機科学の広い見識
- (b) 数学：サンプリング法を解析する数学（確率過程・Malliavin 解析・関数解析）の広い知見
- (c) 社会実装：提案手法を万人が利用できるパッケージへ実装する技術力と主体性
- (d) コミュニケーション力：多様な研究者と交流して問題意識を共有し、共に問題解決へと向かう力

この4分野の全てにおいて高い適性を持つ者は、相当に限られるものと思われる。

以下、(a) 統計計算、(b) 数学、(c) 社会実装 そして (d) コミュニケーション力 を順に分析する。

(a) 統計計算について

知識の幅・深さ、技量

申請者は学部時代にコンピュータサークルで活動した経験があり、計算機科学やセキュリティに関する深い知識がある。実際、学部1年時から世界最大のセキュリティーコンテストである SECCON 決勝大会への出場経験をもつ。

また、サークルを通じて Hennessy & Patterson *Computer Organization and Design* を学び、計算機のアーキテクチャから計算機科学を学べたことは、通常の数学科生活では絶対得られなかったはずの体験であり、現在でも統計計算の研究をする上で大きな糧となっている。

主体性

前述の通り、申請者は学部時代に経営コンサルティング会社にてデータサイエンティストとしてインターンを経験し、その際の実体験がベイズ機械学習への理論的な興味を支えている。製造ラインの異常検知システムを提供するにあたり、熱心にベイズ手法での解決を提案したが、当時は力不足で別の手法を取るようになった。その結果、数上の予測ミスが現場の不信感を招き、既存の意思決定システムへの採用には至らずに終わった。ベイズ手法で提供していたならば、結果は変わっていただろうという信念があり、当時の過ちを二度と犯すまいとベイズ法の周知に尽力するようになった。

申請者は信念を持って周りを巻き込んで主体的に研究を進めることができる。学部3年時には Lebesgue 解析論・確率統計学などの数学科必修授業で好成績を収めるだけでなく、分野をまたいだ勉強会を開催し、経済学部の数理ファイナンスのゼミにも参加して交流を深めた。自身の持つ数学的な強みを提供しつつ、実際に確率論が種々の分野でどう応用されているかを学ぶことができた。

(b) 数学について

確率過程・確率解析

申請者は数学科4年次の1年をかけて Nualart & Nualart (2018) *Introduction to Malliavin Calculus* を読み、毎週90分を指導教授の前で発表した。初めは失敗ばかりであったが必死に準備するうちに、最終的には何も見ずに定義と定理の主張と例から証明までを説明する力が身についた。

計算統計の分野、特に本研究のサンプリングの分野においては、アルゴリズムを確率過程とみなして解析することが常套手段であり、このセミナーの経験が大きな基礎体力となっている。

発想力・問題解決能力

申請者は上述のセミナーを経て涵養した数学的な基礎力を用いて、多くの分野の専門的な数学的文献を読み解き、自身の研究に活かすことが出来る。大学院進学以降も、確率過程と関数解析の更なる知識が必要であったため、指導教員とのセミナーを通じ Jacod & Shiryaev (2003) *Limit Theorems for Stochastic Processes* や Dunford & Schwartz (1958) *Linear Operators I* などにも何度も立ち戻り、適切な数学的結果を利用することが出来た。

(c) 社会実装について

実装力：中規模開発と継続運用の経験

データサイエンティストとしての業務の一環で、公開ライブラリ OSS を有効活用して Twitter から特定の単語を含んだツイートを自動収集し¹形態素分析を通じてトレンドを分析・検知するという、**複数のパイプラインを持ったツールを一人で開発した**。このパイプラインの一部はその後、LLM を活用した感情分析や LLM エージェントを用いたシミュレーションなど、その他のタスクにも利用した。

問題解決能力：データから統計分析を通じ経営提言まで

データサイエンティストとしてのインターン中には、埼玉大学の経営学教授にメンタリングを受ける機会を得た。その助言の下、レポートの形で顧客に報告・還元する業務を行うために、**実際のものでくり企業において制作工程から収集されたデータの分析から施策提言までを一人で実行した**。

コミュニケーションと異分野交流

東京大学を卒業後も先端科学技術研究センター連携研究員という立場で継続的な共同研究や意見交換を継続している。実際、AI の信頼性や知的財産に関する問題など、10 を超える（英語）シンポジウムと研究会を主催し、登壇者とのディスカッションや運営に貢献した。

(d) コミュニケーション力

さらに、本研究を遂行するにあたっては、国内での協業に加えて**海外交流も極めて大事な要素になる**。その理由は、国内で「サンプリング法」または「ベイズ機械学習」を専門とする研究者は少なく、この2つにまたがる研究となるとさらに少ないためである。

申請者は**日本語だけでなく中国語も母語レベルに話せ、また英語も TOEFL iBT で 100 点の語学力を持つ**。また、学部時代に大学設置の Globalization Office 後援の国際交流サークルで、交換留学生の日本での生活のサポートの役割だけでなく、留学生と日本人学生との交流イベントも複数企画し、自ら司会も務めた。大学院進学後も、現在の指導教員の下への訪問研究員のサポート・交流や、英語でのポスター発表を行った経験（業績1）もある。本研究も、その後のディスカッションから多くのアイデアを得たものである。

(1) Shiba, H. *A Recent Development of Particle Filter*. MLSS 2024, Mar. OIST, Okinawa. ポスター発表（査読なし）。本発表では研究計画 ① の一部である PDMP と、その連続時間極限というアイデアの応用の可能性を紹介した。本発表を通じたディスカッションから研究計画 ③ のアイデアが着想された。

(2) 今後研究者として更なる発展のため必要と考えている要素

成果の少なさ → 落ち着いて深い数理的素養の獲得を目指す

申請者はまだ研究成果が少ない状態である。これは、申請者の研究分野が、手法開発だけでなく、これを数学的に理解することで統一的な枠組みを提供しようとするものであるためである。研究対象こそ応用志向であれど、必要であれば数学的結果も創出することが必要である。そのためには、高度な数学（確率過程・確率解析・関数解析・偏微分方程式）への熟練が必要であり、時間はかかるかもしれないが腰を落着けた習得が肝心だと考えている。

高度な数学的概念もわかりやすく説明する力 → 幅広い背景の研究者と研究交流をする

本研究分野は物理学とも深い関わりを持つ。実際、最も代表的なサンプリング法である MCMC は物理シミュレーションのために開発されたものである。そこで申請者は物理学・機械学習など幅広い分野の学会に出席した。その結果、自分の研究成果を他の分野の研究者に説明するためには、自身の分野の枠組みの中で理解をするだけでなく、他分野の問題意識と専門用語を理解する必要があると悟った。

このことは、他分野からのアイデアを積極的に取り入れたり、自身の成果をいろんな人に説明したりする際に、今後とも極めて重要な課題であることは間違いない。数学的な形式を知るだけでなく、分野固有の問題意識の把握のため、多くの交流と積極的な相互理解を努めたい。

¹2023 年当時は可能であったが、サービスが X に変更後はスクレイピングは不可能になり、現在はスクリプトとして公開されていない。

5.【目指す研究者像等】 各事項の字数制限はありませんが、全体で1頁に収めてください。様式の変更・追加は不可

日本学術振興会特別研究員制度は、我が国の学術研究の将来を担う創造性に富んだ研究者の養成・確保に資することを目的としています。この目的に鑑み、(1)「目指す研究者像」、(2)「目指す研究者像に向けて特別研究員の採用期間中に行う研究活動の位置づけ」を記入してください。

(1) 目指す研究者像 ※目指す研究者像に向けて身に付けるべき資質も含め記入してください。

目指す研究者像：数学に軸足を置いた応用ができる数学者

申請者は数学に軸足を置いた応用ができる数学者になりたい。

岡潔が「数学の応用には、真の意味の数学者をじかに使うのが最も簡単で、最も先鋭で、しかも適用範囲が比較にならないほど広い」と述べたように、「真の意味の数学者」に当たる資質を身につけることで、広い範囲の分野に貢献できる仕事をする研究者を目指している。

身に付けるべき資質：(a) 相互理解を促進する (b) 共通言語を提供できる 力

申請者は、現代の数理科学では、数学を軸とした「回転」が起こっていると考え、すなわち、申請者は数理科学の発展の歴史を、次の3ステップとして理解している：

- (i) 初期の応用数学分野（確率統計学やエルゴード理論など）は基本的に数学と応用分野の2者関係であり、「数学的に豊かな問題の提供」と「数学手法の応用」との交換関係があった。
- (ii) 計算機の発展に後押しされ、応用分野が爆発的に増えた（計算機科学、数値計算など）。すると、数学と応用分野の間の2者関係にとどまらず、応用分野の間で数学を共通言語として知見の交換が進み、相互に発展するという3者以上の関係性が増えた。
- (iii) 現代では、数学を軸として、はじめは関係があるとは思わなかった応用分野の間にも、深い繋がりと発見されることも増えている（統計力学と情報科学、ベイズ推論と認知科学など）。

従って、現代では数学が「共通言語」としてより中心的な役割を果たすことが重要である。そのためには、

- (a) 相互理解：各分野でどのような相互理解と共通言語が必要とされているかを見抜く洞察力
- (b) 共通言語：相互理解の場としての共通言語の提供と、その数学的な理解を深める数学力

の2点が必要だと考え、己に課している。

(2) 上記の「目指す研究者像」に向けて、特別研究員の採用期間中に行う研究活動の位置づけ

次に述べる通り、本研究は「目指す研究者像」に必要な2つの資質 (a) 相互理解の促進 と (b) 共通言語の提供 を鍛えるための格好のテーマとなっている。

(a) 本研究は相互理解のために極めて肝心なテーマである

申請者は大学入学後すぐ、統計学が新たな応用数学の舞台になるという直感に従って学修を進めてきた。その中でも現在の指導教員を通じて、統計計算とモンテカルロ法という、極めて多くの分野が交差する魅力的な研究テーマに出会うことができた。統計学や機械学習で必要不可欠な技術であるだけでなく、モンテカルロ・サンプリングはもともと物理学シミュレーションのために開発されたものであり、現在では物性科学一般で広く使われる手法になっている。まさに諸科学のハブとしての役割を果たしていると言える。

(b) 更なる数学的理解が必要な共通語彙がある

モンテカルロ・サンプリング法においてマルコフ過程が必要不可欠な役割を果たす。それだけでなく、サンプリング法としてはこれが $\mathcal{P}(\mathbb{R}^d)$ 上に定める力学系として捉える見方が極めて自然になる。有限次元の幾何学には多くの理論の蓄積があるが、無限次元空間内の凸集合である $\mathcal{P}(\mathbb{R}^d)$ 上の力学系の解析には理解が足りない部分も多い。近年、 $\mathcal{P}(\mathbb{R}^d)$ を通じてサンプリング法と最適化や最適輸送などの諸分野との繋がりが見えてきており、共通理解の場を提供する語彙として更なる理解と解明が待たれている。