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，ベイズ統計の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**ことに魅力を感じた（図 1 参照）.

しかし，現状のベイズ法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，ベイズ法が敬遠されたり，安易な近似手法が選好されたりしがちな現状には苦い思いを感じていた.

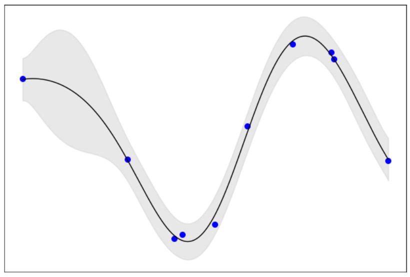


図 1: ベイズ法（Gauss 過程回帰）の結果．推定値（実線）だけでなく 95% 信用区間（灰色）も自然に得られる．データが少ない領域で予測が不確実である（＝モデルに自信がない）ことがわかる.

当該分野の状況：ベイズ法の発展にはサンプリング法の発展が重要

ベイズ法とは，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．ベイズ法は最適である（＝どの他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応用範囲を狭めてきた.

ほとんどすべてのベイズ法は確率測度 $P\in\mathcal{P}(\mathbb{R}^d)$ に関する積分の数値計算の問題に帰着する．決定論的な数値計算法は高々 $d\le 3$ 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）してモンテカルロ推定により積分を近似する：

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

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，ベイズ法の普及と応用を妨げる要因となってしまっている.

応用

ベイズ法 ← **ベイズの定理**

- Gauss 過程回帰
- ベイズ深層学習
- ⋮

本研究 ⇑ **課題**

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\displaystyle\int_{\mathbb{R}^d} p(y|\theta)p(\theta)d\theta}$$

サンプリング法 HMC の開発により分母の積分の計算が効率化されて普及

① 性能比較のための理論が遅れている
② 拡張性のある手法が不足している

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．たとえその中に HMC より効率的なサンプリング法があったとしても，我々はそのことさえ確認できないのである.

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 $d=10^n$ $(n\ge 4)$ などの極めて高次元になることが多い．このような高次元の問題ではベイズ法が採用されにくい状況にある.

【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上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リング法の構成をめざす**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，**モデルの設計だけに集中出来る枠組み**の構成にある．

	ベイズ法	従来法
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**（共通）	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**開発したサンプリング法を 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．

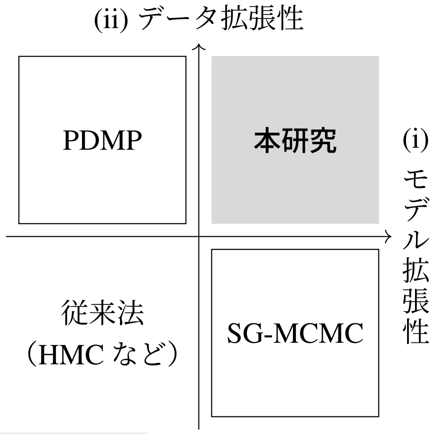
本研究は上述の研究目標に向けた研究計画 ①, ②, ③ の３部構成からなり，各アプローチに１年ずつかける．

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) **区分確定的マルコフ過程（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

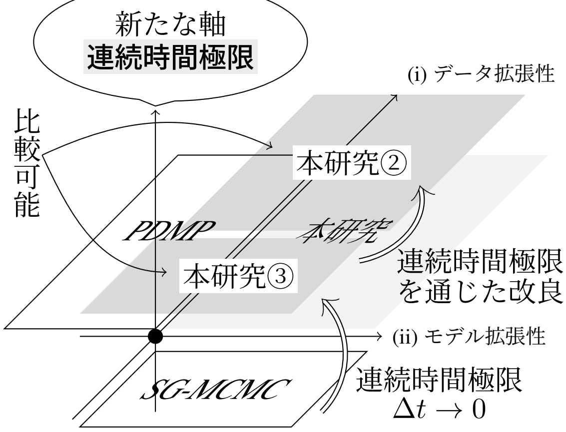


そこで**本研究はこの２つの手法を検討するところから始める**．いずれの手法も **２つの拡張性** (i),(ii) のうち片方を満たさないため（右図参照），**その弱点をどのように克服できるかの限界を探る**．

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．

(研究目的・内容等の続き)

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．この**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 と PDMP はいずれも連続時間の要素がある [3]．そこで従来手法も**連続時間極限を考えることで効率的化できるのではないか**と考えられる．現状，サンプリング法の連続時間極限を通じて改良・比較を試みる研究は限定的である [4]．さらに一般の手法の連続時間極限はジャンプを持つが，現状はジャンプのない拡散過程しか考察されていない．

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 $N \to \infty$ や実行時間無限大 $T \to \infty$ の極限はよく調べられてきたが，**タイムステップ極限 $\Delta t \to 0$** 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 $N \to \infty, T \to \infty$ 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，近年，**連続時間極限 $\Delta t \to 0$ も重要だと示す証拠も見つかりつつある**．

例えば，連続時間極限から得られるジャンプ測度から，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．同様にして，この**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することを考える．したがって，本研究が狙う連続時間極限に基づくサンプリング法の創出は，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，確率過程論と機械学習数理の稀有な交流が見られ，近年急速に知見の交換が進んで結果 [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．

サンプリング法の発展は広くベイズ法に貢献し，ベイズ法の発展は多くの応用科学分野に貢献する．そのため，統計と機械学習をはじめとした**極めて広い分野の研究者がサンプリング法の発展状況に強い興味を抱いている**ことが，修士を通じて，多くの学会へ出席・発表した経験からわかった．

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, Continious-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）の広い素養
(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年生時点で，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．

加えて，サークルの勉強会で Hennessy & Patterson (2013) *Computer Organization and Design* を学び，計算機科学の基礎を体得したことは，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．通常の数学科生活では絶対得られなかったはずの知識でも，**申請者が主体的に学習できる証左でもある.**

主体性

申請者は，統計計算のでは理論と実践の結びつきが重要だと考え，学部時代に自ら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のインターンを志願**し，実際に１年に渡り継続的に関わった．その際の**実体験がベイズ法への理論的な興味を支えている.** このように，主体的に自身のスキルアップを理論・実践の両面で志向し，理論を実践に還元するだけでなく，実践の場での問題を理論的な興味に還元し，**双方向にシナジーを生み出すことができる.**

このように申請者は「ベイズ法は自分にぴったりな研究テーマだ」と深く感じているため，**周りを巻き込んで主体的に研究・勉強を進めることができる.** 学部３年時には Lebesgue 積分・複素解析学などの数学科必修科目で好成績を収めつつ，分野をまたいだ勉強会を開催し，さらに経済学部の数理ファイナンスのゼミに他学部履修をして交流を深めた．その結果，自身の持つ数学的な強みを提供しつつ，自分は確率論の応用を学べるという，ギブアンドテイクの良い研究協力関係を育てることができた．

(b) 数学について

確率過程・関数解析

申請者は数学科３年次に Pedersen (1989) *Analysis Now* を，４年次の１年をかけて Nualart & Nualart (2018) *Introduction to Malliavin Calculus* を読み，毎週 90 分を指導教員の前で発表した．初めは失敗ばかりであったが必死に準備するうちに，何も見ずに定義・定理・例・証明を説明する力が身についた．本研究で扱うサンプリング法の分野においては，アルゴリズムを確率過程とみなして関数解析の方法で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数学的基礎体力となっている.**

発想力・問題解決能力

申請者が上述のセミナーで涵養した数学的な基礎力は，本研究を遂行するにあたって莫大な財産である．以前な不慣れであった分野の専門的な数学的文献も，読み解いて自身の研究に活かすことが出来る．大学院進学以降も，確率過程の収束やさらに一般の積分に関す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* などを参照し，適切な数学的結果を利用することが出来た．

(研究遂行力の自己分析の続き)

(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も主体的に活動し，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むツイートを自動収集し[1]形態素分析を通じて分析するという**複数のパイプラインを持ったツールを一人で開発した**．このパイプラインはチームの大きな役に立ち，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，会社で種々のタスクに応用された．

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，実際のものづくり企業において**制作工程から収集されたデータの分析から事業改善の提言までを一人で実行**した．

異分野交流への強い主体性

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「ベイズ法」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．本研究も一部，その後のディスカッションから着想を得たものである．

(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ち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，自身の分野に詳しくなるだけでなく，他分野の問題意識と専門用語を深く学ぶ必要もあると悟った．

申請者は他分野との相互理解と意見交換は今後とも極めて重要な課題であると考えており，他分野を数学的な形式として学ぶだけでなく，分野固有の問題意識を把握し，より深い交流と積極的な相互理解に繋げることが重要だと考えている．

[1] 2023 年 2 月当初は可能であったが，Twitter のサービスが 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・継続運用はされていない．

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。

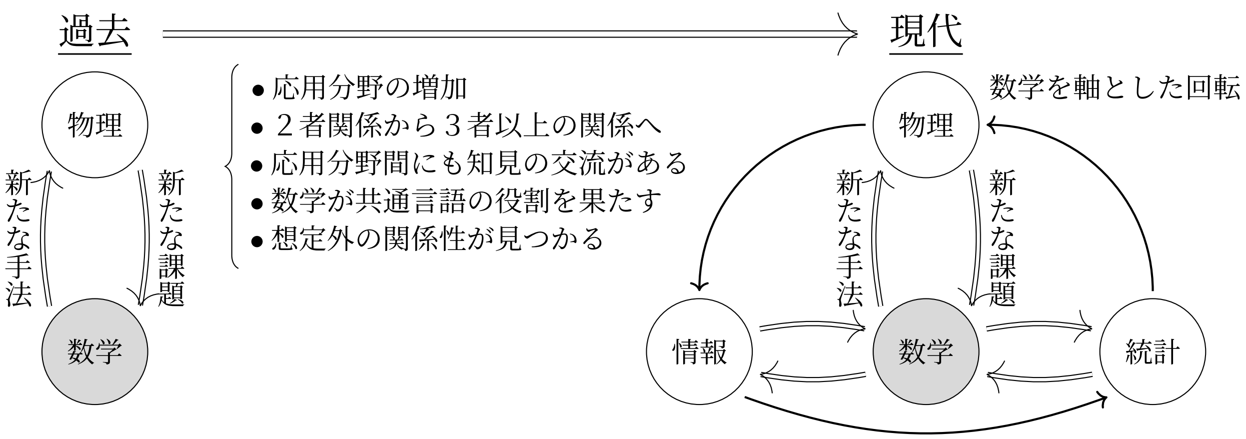
目指す研究者像：数学に軸足を置いた応用ができる「真の意味の数学者」

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．これについては，日本の数学者岡潔が「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」という言葉を残している．

申請者は，この岡潔のいう「**真の意味の数学者**」に当たる資質を身につけた応用数学者となることで，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，現代の応用数学では，**数学を軸とした「回転」**が起こっていると考える．



そこで申請者は，「真の意味の数学者」として次の２つの資質が必要と考え，己に課している：

(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：相互理解の場や共通言語としての数学と，その理解を深める数学力

(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

本研究は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

申請者は数学を軸としつつも，種々の応用分野の学習を欠かさなかった．その中で，現在の指導教員を通じて，統計計算とサンプリング法という，**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．現代のサンプリング法は統計や機械学習で必要不可欠な技術であるだけでなく，もともと物理学で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．まさに多くの応用分野のハブとしての役割を果たしていると言っても過言でないテーマである．

(b) 更なる数学的理解が必要な共通言語がある

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．マルコフ過程を分析する際は $\mathcal{P}(\mathbb{R}^d)$ 上の（決定論的な）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 $\mathcal{P}(\mathbb{R}^d)$ 上の力学系にはまだ理解が及ばない部分が多い．近年，機械学習や最適輸送の分野でも $\mathcal{P}(\mathbb{R}^d)$ の解明が重要になっており，サンプリング法との関係性の解明が切望されている．つまり，$\mathcal{P}(\mathbb{R}^d)$ は**共通理解の場を提供する言語として更なる理解が待たれている．**