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：基盤モデルの実社会応用を阻む課題「信頼性の問題」

昨年 2023 年に ChatGPT などの大規模言語モデルや StableDiffusion などの**基盤モデル**の登場を目撃してから，我々の人工知能（AI）に対する期待は増すばかりである．しかし筆者は，学部時代にデータサイエンティストとしてインターンする中，各企業の実際の要望に応える経験を通じて，最先端の AI 技術と**実際の需要との間には未だ埋められない溝がある**ことを思い知らされた.

それは**信頼性の問題**であった．例えば，LLM の出力には事実無根で全く信頼出来ないものが混ざっていることが多く，これは幻覚（hallucination）と呼ばれ問題とされている．他にも，モデルの精度が良くとも，その出力がいつ信頼に足るものかが分からなければ，結果として使用者の**不信感**を募らせ，うまく既存のシステムに組み込むことは出来ず採用には至らないということも多かった.

ベイズ法の採用による抜本的解決が悲願である

ここで筆者は**ベイズのアプローチによる信頼性問題の根本的な解決**が，今後 AI が**本格的に実社会に応用される段階を迎えるにあたって必要不可欠**であると主張したい.

ベイズ手法は不確実性の定量化を統一的に扱うことができ，上述の**信頼性問題**への抜本的解決が見込まれる．例えば幻覚の問題も，不確実な出力を回避するよう調整することで大幅に改善できることが判りつつある [1]．その他の場面でも，ベイズ手法を取り入れることにより，機械学習手法が実社会における意思決定システムの一環として採用しやすいものとなると期待されている.

実社会環境の複雑さが課す
新たな課題を乗り越えるには
双方の美点を活かすことが必要

- **信頼性問題**
  - **幻覚問題**
  - モデルの解釈可能性
- モデル誤特定へのロバスト性
- 変化するドメイン
  - 継続学習
  - 転移学習
  - ⋮

スケーラビリティ問題
により遅れをとっている

現状殆どの**基盤モデル**は
こちらのアプローチ

新たな課題	ベイズ流（確率論的）	非ベイズ流（非確率的・頻度論的）
出力の不確実性	自然に導ける	追加の手続きが必要
小規模データ	対応可能	大規模データは必須
⋮		

いずれもベイズからの
アプローチが期待されている [2]

ベイズ法を応用するに当たって最大にして唯一の課題「スケーラビリティ」

ベイズ法の**深層学習**への応用にはすでに約 30 年の歴史があり [3]，小規模なモデルや特定の用途においては成功を収めている．それにも拘らず，現在まで**基盤モデル**などの大規模なモデルへの応用が叶っていない最大の理由は，事後分布が高次元かつ多峰的である場合にも精度が落ちず，しかし同時に計算量が爆発しないような**スケーラブルな正確手法の欠如**による [2, 4]. [1]

サンプリングに基づいた正確な（バイアスフリーな）推定手法はこの 30 年で大きく発展し，広く統計と機械学習に応用されているが，**深層モデル**の事後分布が持つ高次元性と多峰性を特に苦手とし，現状**確率的勾配降下法（SGD）に匹敵する速度と精度を持つアルゴリズムの開発には至っていない**.

[1] MCMC や SMC などのサンプリング法に対し，Laplace 近似や変分近似などのバイアスを許した近似手法では SGD に匹敵する速度とスケーラビリティを持つものが多いが，分布の多峰性をうまく再現できずにいた．この問題は据え置きであるが，変分近似がサンプリング法と並んでベイズ深層モデルへの有力なアプローチの１つであると評せる結果が今年２月末に報告された [5].

【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３点を兼ね備えたスケーラブルな正確サンプリング手法の開発と実装を目指す

(i) **モデルスケーラビリティ**：高次元で多峰性を持つモデルでも精度と速度が落ちないこと
(ii) **データスケーラビリティ**：大規模なデータでも計算時間が爆発せず，SGD に匹敵する速度を保つこと

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 [5] や，実際の**深層モデル**の対称性を利用した効率化などは，扱わない．すなわち，**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目指す**のが特徴である．

	ベイズ機械学習（確率的）	非確率的機械学習
アイデア	モデルベース	推論ベース
実装	確率的プログラミング	最適化アルゴリズム
特徴	種々の不確実性を モデリング可能	アルゴリズムが高速 大データなら高精度

種々のモデルに統一的に使える手法の発展は，ベイズ手法が真に広く応用されるために必要不可欠な要素である．ベイズ機械学習の悲願とは，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．本研究には**開発したサンプラーの OSS パッケージへの実装**も含まれる．機械学習の不確実な環境への応用ではモデリングが不可欠であるから，本研究により確率的プログラミングの環境が前進した暁には，科学的データ解析や実世界応用などより複雑な環境への**信頼のおける形で**の応用が大きく進むだろう [6]．

実際，現在確率的プログラミング言語 Stan で標準に用いられているハミルトニアンモンテカルロ法（HMC）は，28 年前に**深層モデル**への応用を念頭に提案されたものであった [3]．それが現在汎用サンプラーとして用いられているのである．HMC が 30 年目を迎える前に，より効率的な手法の提案が待たれている．

研究計画（前半）：２つの有望な方向の検討と限界の明瞭化

現状，サブサンプリングによりバッチ実行可能であることが**スケーラブルなサンプリング手法**の有望な方向として考えられており，この性質を持つ代表的な手法として次の２つが挙げられる：

(1) **区分的確定的マルコフ過程（PDMP）**と呼ばれる，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，連続時間 MCMC とも呼ばれる手法群
(2) **確率的勾配マルコフ連鎖モンテカルロ法（SG-MCMC）**と呼ばれる，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 (i),(ii) のうち片方を満たさないため，**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) データスケーラビリティ**の条件を満たし，高次元での収束速度も期待されている [4] が，適用可能なモデルが限られており **(i) モデルスケーラビリティ**が課題である．一方で，SG-MCMC は SGD と同じく勾配の情報を用いるもので，適用可能なモデルは広いが，SGD に匹敵する **(ii) データスケーラビリティ**が未だ達成されていない．

(研究目的・内容等の続き)

研究計画（後半）：連続時間極限をキーワードに，SMC と最適化数理の知見を流入させる

これら２つの手法の持つ課題はいずれも困難で，すぐに HMC より効率的なサンプラーになるとは思えない．２つの手法の強みを活かした上で，**新たな知見**が必要と考える．**ここからが本研究の真の独自性である**．

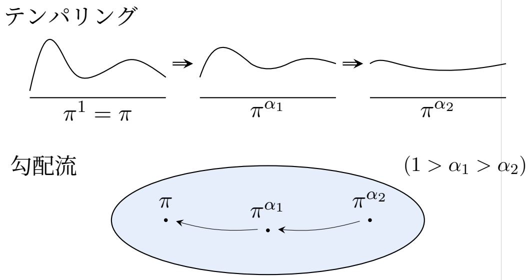
そもそも SG-MCMC は拡散過程の離散化であり，PDMP も元は MCMC の**連続時間極限**の考察から着想された．そこで自然に逐次モンテカルロ法（SMC）の**連続時間極限**はどうなるかという興味が生じるが，これを考察してサンプリングに応用する研究は現状１つ [7] しかなく，それも拡散過程に問題を絞っている．これをテンパリングの考え方と組み合わせて一般の分布からサンプリングする試みはまだ見られない．

この連続時間 SMC からのアプローチが現状を打開する極めて有望な方向である．それは SG-MCMC と PDMP の弱みを補い合う新たな方向を提供するためである．特に次の２つのアイデアを検証する．

(a) SG-MCMC は拡散過程をシミュレートする際に入るバイアスを補正する手続きが計算を遅くしている．PDMP では誤差は入らない（右図参照）．ここに**連続時間 SMC** [7] を応用すると，サブサンプルからの重点荷重の不偏推定量を通じて，PDMP をシミュレートすることを通じて拡散過程からの正確なシミュレーションが可能になる．これにより PDMP と SG-MCMC の美点を組み合わせる方向が考えられる．

PDMP は区分的確定的なので
正確なシミュレーションが可能

(b) テンパリングを用いた SMC サンプラーは，**連続時間極限**において確率測度の空間 $\mathcal{P}(X)$ 上のある勾配流とみなせる（右図参照）この**最適化**からの再解釈を通じて，機械学習研究の蓄積を SMC サンプラーに応用できる [8]．元々 SMC サンプラーが MCMC より分散実行に向いていることもあり，PDMP や SG-MCMC とも違うサブサンプリングに依らない**新たなスケーラビリティ**が連続時間 SMC を基礎として提案される可能性も十分にある．



以上のアイデアは全て**連続時間極限**というキーワードを中心に MCMC, SMC, $\mathcal{P}(X)$ 上の**最適化**の３つの知見を融合させて生まれたものである．PDMP や SMC の**連続時間極限**の理解が進み，３分野間の交流が活発になると，一気に**新時代のサンプラー**に近づけるだろうというのが応募者の計画である．

このアプローチが現在まで試みられていない理由の１つに**連続極限**という確率過程論のトピックと，$\mathcal{P}(X)$ 上の**最適化**のトピックという，従来希薄であった２つの分野の交流が必要である点が挙げられる．

研究計画（在外研究）：確率過程論と最適化数理との交流

以上の研究テーマの一部は，応募者の 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最適化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いる [8] 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．

参考文献

[1] Mohri, C. & Hashimoto, T. Language models with conformal factuality guarantees (2024). 2402.10978.

[2] Papamarkou, T. *et al.* Position paper: Bayesian deep learning in the age of large-scale ai (2024). 2402.00809.

[3] Neal, R. M. *Bayesian Learning for Neural Networks*, vol. 118 of *Lecture Notes in Statistics* (Springer New York, 1996).

[4] Fearnhead, P., Bierkens, J., Pollock, M. & Roberts, G. O. Piecewise Deterministic Markov Processes for Continuous-Time Monte Carlo. *Statistical Science* **33**, 386 – 412 (2018).

[5] Shen, Y. *et al.* Variational learning is effective for large deep networks (2024). 2402.17641.

[6] Ghahramani, Z. Probabilistic machine learning and artificial intelligence. *Nature* **521**, 452–459 (2015).

[7] Fearnhead, P., Latuszynski, K., Roberts, G. O. & Sermaidis, G. Continious-time importance sampling: Monte carlo methods which avoid time-discretisation error (2017). 1712.06201.

[8] Chopin, N., Crucinio, F. & Korba, A. A connection between tempering and entropic mirror descent (2024). 2310.11914.