arXiv:2601.05227v1 [stat.ML] 8 Jan 2026

随机深度学习：结构化时序数据不确定性建模的概率框架

James K Rice*

2026年1月9日

摘要

我提出了一种新颖的框架，将随机微分方程（随机微分方程）与深度生成模型相结合，以改进涉及结构化数据和时间序列数据的机器学习应用中的不确定性量化。该方法称为随机潜变量微分推断（SLDI），在变分自编码器的潜在空间中嵌入一个 Itˆo 随机微分方程，从而在保留严谨数学基础的同时，实现灵活的连续时间不确定性建模。随机微分方程的漂移项和扩散项由神经网络参数化，使数据驱动推断成为可能，并将经典时间序列模型扩展到能够处理不规则采样和复杂动态结构。

一项核心理论贡献是对伴随状态与专用神经网络进行协同参数化，形成一个耦合的前向-后向系统，不仅能够捕捉潜在状态的演化，还能刻画梯度动力学。我引入了路径正则化伴随损失，并从随机微积分的角度分析了方差缩减梯度流，为提高深度潜变量随机微分方程的训练稳定性提供了新工具。本文统一并拓展了变分推断、连续时间生成建模和控制理论优化，为随机概率机器学习的未来发展提供了严谨基础。

关键词：随机微分方程；变分推断；深度生成模型；连续时间模型；伴随灵敏度；不确定性量化。

1 引言

不确定性是现实世界系统的内在特征，存在于金融市场、气候动力学、神经活动和疾病进展等领域。传统机器学习方法虽然灵活性高、性能出色，却往往只提供点估计，几乎不对不确定性进行量化，甚至完全不予量化。这一局限可能导致预测过度自信，并降低关键应用中的可靠性。为了解决这一问题，概率机器学习

*博士生，埃塞克斯大学；james.k.rice@essex.ac.uk

方法应运而生，成为一种颇具吸引力的替代方案，旨在将严谨的统计推理与深度学习的表征能力相结合。

近期的进展推动了概率范式与深度学习范式的融合，包括贝叶斯神经网络（Blundell 等，2015）、高斯过程（Rasmussen 和 Williams，2005）以及变分推断框架（Kingma 和 Welling，2013）。然而，对兼具时间性 和结构性 的数据进行建模，仍是一项长期挑战——在这类数据中，不确定性会随时间演变，并跨越潜在空间或关系流形。标准方法可能会忽略噪声的动态变化（假设分布是静态的），也可能在整合进深度架构时面临可扩展性问题。

在这项工作中，我提出了一种新颖的概率框架，将随时间演变的不确定性明确纳入深度生成建模。我的方法以随机微分方程（随机微分方程）的数学理论为基础，尤其是伊藤型随机微分方程，它们能够自然地对随机条件下的连续时间过程进行建模。通过将随机微分方程嵌入变分自编码器（变分自编码器）结构中，我实现了对灵活、可解释且时间上连贯的潜在动力学的学习。这一“随机深度学习”框架兼顾随机微积分与深度学习，为经典随机建模和现代数据驱动推断搭建了桥梁。

我在多个真实世界和合成数据集上评估了该方法，结果表明，与确定性模型和标准变分方法相比，该方法在预测准确性和不确定性校准方面均表现更优。此外，我还证明，学习得到的潜在随机微分方程具有可解释性，并能揭示不确定性的时间演化。

2 背景与文献综述

2.1 概率机器学习

概率机器学习方法旨在捕捉预测建模中的认知不确定性和偶然不确定性。与确定性神经网络不同，概率模型会为参数或潜变量分配概率分布，从而更全面地理解模型置信度。

该领域的一个基础模型是贝叶斯神经网络（BNN），它通过学习权重的后验分布来建模，通常采用变分推断或马尔可夫链蒙特卡罗等近似推断技术（Neal，2012）。尽管理论上颇具吸引力，BNN 往往面临计算开销大、收敛不稳定等问题，因此人们转而采用基于丢弃法的近似（Gal 和 Ghahramani，2016）或深度集成（Lakshminarayanan 等，2017）等替代方法。

变分自编码器（VAEs）是概率深度学习的另一重要分支。VAEs 使用编码器—解码器架构来学习低维潜在空间，并通过变分后验捕捉不确定性（Kingma 和 Welling，2013）。然而，标准 VAE 通常使用离散潜在状态对时间进行建模，或忽略时间连贯性，因此在动态环境中的效用受到限制。

2.2 随机微分方程与机器学习

随机微分方程（SDEs）通过添加随机噪声分量扩展了常微分方程（ODEs），这些分量通常建模为维纳过程（布朗运动）。Itˆo 随机微分方程的一种标准形式为:

$$dX_t = \mu(X_t, t)\, dt + \sigma(X_t, t)\, dW_t,$$

其中， $X_t \in \mathbb{R}^d$ 是时间 t的状态， $\mu: \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$ 是描述确定性趋势的漂移函数， $\sigma: \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^{d\times m}$ 是扩散系数， $W_t \in \mathbb{R}^m$ 表示标准布朗运动。这些方程是对动力系统建模的基础，尤其适用于噪声发挥关键作用的情形；其经典应用涵盖 Black and Scholes 期权定价模型（Black and Scholes 1973年），以及使用随机流行病模型对传染病传播进行建模（Allen 2008）。

随机微分方程的数学理论以 Itˆo 积分为基础。Itˆo 积分对随机过程有明确的定义，并构成 Itˆo 引理的基础，而 Itˆo 引理是随机微积分的基石。与常微分方程不同，随机微分方程的解 X_t 并非确定性函数，而是一个随机过程，具有随时间变化的概率分布。例如，如果 σ 为常数且$\mu = 0$，那么解将简化为标准布朗运动，其满足标度性质$W_{ct} \sim \sqrt{c}W_t$ ，对任意 $c > 0$均成立。这意味着轨迹几乎必然处处不可微，这一特征与常微分方程的光滑解有本质区别。

随机微分方程的数值近似需要谨慎处理，以确保依分布收敛。欧拉–丸山方法是最常用的数值格式，可视为经典常微分方程欧拉方法的随机类比。它按如下方式更新状态:

$$X_{t+\Delta t} = X_t + \mu(X_t, t)\Delta t + \sigma(X_t, t)\sqrt{\Delta t} \cdot \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, I).$$

虽然这种方法简单且应用广泛，但其收敛速度仅为强收敛阶$\mathcal{O}(\sqrt{\Delta t})$。若要获得更高精度，可能需要采用更复杂的方法，例如米尔斯坦格式（Kloeden and Platen，1992），尤其是在模拟金融模型或具有刚性动力学的系统时。

随着神经随机微分方程（Kidger 等，2021）的出现，将随机微分方程融入机器学习逐渐受到关注。在这种方法中，漂移项和扩散项由神经网络参数化。这种形式能够对灵活的非参数连续时间动力学进行建模，因此适用于不规则时间序列建模和基于路径的生成建模等任务。这些方法通过允许系统动力学具有随机性，扩展了神经常微分方程（神经常微分方程），从而能够量化预测轨迹中的不确定性。

然而，标准的神经随机微分方程主要关注将初始条件映射为未来状态的概率分布，类似于黑箱随机模拟器。尽管它们在预测任务中表现有效，却往往缺乏潜变量模型中进行严谨推断所需的可解释性和结构。这一局限促使人们将随机微分方程纳入变分自编码器（VAE）框架，使潜变量通过随机动力学而非确定性转移演化。由此产生的模型，例如潜在随机微分方程（Li 等，2020），能够更丰富地表示所学习时间结构中的不确定性。

此外，将随机微分方程嵌入变分框架后，可以联合推断潜在轨迹的后验，并通过随机梯度下降优化模型参数。扩展到随机微分方程的重参数化技巧，在实现证据下界（ELBO）的低方差梯度估计方面发挥着关键作用。这为不确定性条件下的潜在动力学学习提供了一种灵活且可扩展的方法，兼具概率建模的可解释性与深度学习架构的适应性。

2.3 具有时间随机性的潜变量模型

将连续时间潜在动力学融入生成模型，推动了时间序列建模领域的一系列创新。经典状态空间模型（SSMs）假设线性转移和高斯噪声，这限制了其在复杂非线性环境中的表达能力。神经状态空间模型试图通过使用神经网络对转移函数和发射函数进行参数化来放宽这些假设。然而，这些模型通常以离散化方式处理时间；对于不规则采样数据，或在底层过程本质上是连续的情况下，这种做法可能会产生问题。

为了解决这些局限，近期有多项研究通过常微分方程和随机微分方程引入了连续时间潜在动力学。神经常微分方程框架（Chen 等，2018）将残差网络解释为常微分方程求解器，使隐藏状态能够随连续时间演化。尽管神经常微分方程具有优雅性，但其施加了确定性结构，缺乏捕捉路径不确定性的能力——在对含噪真实世界序列进行建模时，这是一个关键局限。

潜在神经随机微分方程（潜在随机微分方程）试图通过使用随机微分方程定义潜在轨迹来弥合这一差距，其中漂移项和扩散项均由神经网络参数化（Rubanova 等人，2019 年）。由此形成了一个连续时间变分推断框架，能够学习潜在路径上的分布。然而，由于路径 KL 散度难以计算，这些模型中的后验推断往往需要复杂的近似，尤其是在扩散项为非对角形式或数据采样不规则时。

在此基础上，Tzen 和 Raginsky（2019 年）提出将随机过程视为受控随机微分方程的解，其中控制输入通过优化进行学习。他们的框架将随机微分方程建模与控制理论联系起来，并弥合了基于采样的推断与路径式随机优化之间的差距。其他进展，例如利用布朗桥对潜在轨迹进行正则化（Archer 等人，2015 年），展现了结合物理约束与概率生成模型的潜力；这在神经科学和机器人学等可利用动力学先验知识的领域尤为重要。

另一个值得关注的方向是使用基于分数的生成模型和扩散过程（Song 等人，2021 年）。这些模型定义了一个随机过程，通过逆转正向扩散，将噪声转换为数据。尽管它们通常不被表述为潜变量模型，但实际上能够学习复杂数据分布随时间演化的结构；从随机控制视角来看，它们与神经随机微分方程密切相关。将其整合到概率推断流程中，是一个很有前景的研究前沿。

重要的是，学习连续随机潜在表示的能力，能够稳健地应对不规则采样、长程依赖和多模态不确定性——这些方面往往是传统变分自编码器或循环神经网络的短板。我提出的框架立足于

这一新兴领域，提供了一种有原则的方法，可联合学习潜在动力学、量化不确定性并重构高维序列。该框架将推断和生成均表述为随机微分方程驱动的过程，并通过随机变分推断进行端到端训练，从而推广了先前的方法。

从理论角度来看，我的方法也与粗糙路径理论和基于签名的学习方面的近期研究相关（Lyons 和 Qian，2007年；Kidger 等人，2020年），这些研究为从不规则的连续时间路径中学习提供了其他视角。这些联系指向了对现代深度学习系统中路径依赖推断的更深入数学理解，并促使我们进一步探索随机潜在空间的几何结构。

2.4 贡献

本研究首次将 Itô 随机微分方程（随机微分方程）融入变分自编码器的潜在结构，建立了一个有原则且富有表现力的框架，用于在连续时间情境下对不确定性建模。通过将随机微分方程直接嵌入生成模型，我能够构建时间上连贯、路径依赖的潜在动力学，从而捕捉复杂序列中的偶然不确定性和认知不确定性。我的编码器、解码器和转移函数均由神经网络参数化，而扩散组件则增强了对演变中的随机系统进行建模的灵活性和真实性。这一形式扩展了标准变分自编码器和神经常微分方程范式（Kingma 和 Welling，2013年；Chen 等，2018），并提供了一种自然的方法来处理不规则采样数据（Rubanova 等人，2019 年）。

我开发了一种基于随机微分方程伴随灵敏度方法的高效、可扩展学习算法。该算法无需存储完整的潜在路径即可估计梯度，从而显著降低训练期间的内存使用量（Kidger 等，2021）。为提高稳定性和准确性，我引入了路径正则化伴随损失、谱范数正则化和自适应方差控制机制。这些创新共同构成了一种稳健且具有坚实数学基础的训练策略，可用于在高维环境中学习潜在随机微分方程。通过异方差解码器、重参数化变分推断和结构化似然模型等计算技术，可以进一步增强模型的灵活性；这些技术均有助于提升重构质量和预测性能。

最后，我通过一个简单的证明展示了该方法的有效性，证明其在预测准确度、轨迹连贯性和不确定性校准方面理想情况下表现更优。我的推导证实，SLDI 框架能够学习可解释且具有时间结构的潜在空间，同时保持推断的可处理性和优化的高效性。所提出的模型对文献中的多个研究脉络进行了泛化与统一——从深度生成建模和变分推断，到随机控制和扩散过程——为概率机器学习、序贯决策和时间序列建模的未来研究提供了灵活的基础。

3 方法论

在本节中，我将介绍一种名为随机潜在微分推断 （SLDI）的新颖框架。该框架将随机微分方程（SDEs）融入变分自编码器（VAE）结构，用于建模序列数据中随时间演变的不确定性。该方法结合了深度神经网络的表征能力与连续时间随机过程的概率表达能力，从而构建出能够捕捉潜在动力学并进行有原则的不确定性估计的模型。

3.1 潜在随机动力学

我首先将连续时间演化的潜变量 $\{z_t\}_{t=0}^{T}$ 建模为 Itˆo 随机微分方程（SDE）的解。具体而言，令 $z_t \in \mathbb{R}^d$ 表示时间 t的潜在状态。 z_t 的演化由下式支配：

$$dz_t = \mu_\theta(z_t, t)\, dt + \Sigma_\theta(z_t, t)\, dW_t, \tag{1}$$

其中， μ_θ： $\mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$ 是漂移函数， Σ_θ： $\mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^{d\times m}$ 是扩散系数， $W_t \in \mathbb{R}^m$ 是一个 m 维布朗运动。漂移项和扩散项由神经网络参数化，其参数为 θ ，并通过变分推断进行优化。

该随机过程在潜在空间的样本路径上诱导出一个分布，其在路径空间 $\mathcal{C}([0, T],\ \mathbb{R}^d)$ 上的概率测度记为 $\mathbb{P}_\theta$。与神经常微分方程（Chen 等，2018）生成的确定性轨迹不同，方程 1 的解构成一个具有路径依赖不确定性的马尔可夫过程。对于固定的 θ，边缘分布 $p_\theta(z_t)$ 按照福克–普朗克方程（Risken，1996）演化，该方程描述了 z_t 的概率密度函数 $\rho(z, t)$ 随时间的演化：

$$\frac{\partial \rho}{\partial t} = -\nabla_z \cdot (\mu_\theta(z, t)\rho) + \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2}{\partial z_i \partial z_j}\left(D_{ij}(z, t)\rho\right), \tag{2}$$

其中 $D(z,\ t) = \Sigma_\theta(z,\ t)\Sigma_\theta(z,\ t)^\top$ 是扩散矩阵。这个偏微分方程揭示了确定性漂移项和随机扩散如何共同塑造潜在空间中不确定性的演变。在实践中，我并不直接求解方程 2，但它提供了一个理论视角，帮助我们理解潜在随机微分方程的行为，尤其是在研究平稳分布和熵产生时（Pavliotis，2014）。

为了对潜在过程的样本路径进行数值模拟，我采用欧拉-丸山离散化（Kloeden 和 Platen，1992）。令 Δt 为一个较小的时间步长，离散化后的更新公式为：

$$z_{t+\Delta t} = z_t + \mu_\theta(z_t, t)\Delta t + \Sigma_\theta(z_t, t)\sqrt{\Delta t} \cdot \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, I_m), \tag{3}$$

其中， ϵ_t 是 $\mathbb{R}^m$中的标准高斯噪声向量。此随机更新规则既用于生成过程中的前向模拟，也用于近似推断中的变分后验。虽然欧拉–丸山方法仅是一阶方法（其强收敛阶为

$\mathcal{O}(\sqrt{\Delta t})$，但它高效且足以满足许多无法进行精确模拟的学习应用的需求。

过程 z_t 的一阶和二阶矩满足各自的确定性方程。对式 1 取期望，可得：

$$\frac{d}{dt}\mathbb{E}[z_t] = \mathbb{E}[\mu_\theta(z_t, t)], \tag{4}$$

$$\frac{d}{dt}\mathrm{Cov}(z_t) = \mathbb{E}[\Sigma_\theta(z_t, t)\Sigma_\theta(z_t, t)^\top] + \mathrm{Cov}(\mu_\theta(z_t, t)), \tag{5}$$

如 S¨arkk¨a 和 Solin（2019年）所述。这些方程表明，即使漂移项的期望为零，扩散仍会导致协方差随时间线性或非线性增长。不确定性的这种时间演化对于区分均值相似但噪声特性不同的系统至关重要——确定性潜变量模型（如神经常微分方程）无法捕捉这一特征。

从变分推断的角度来看，潜在随机微分方程充当路径上的随机先验，随后与根据观测数据推断出的后验相匹配。通过神经网络对 μ_θ 和 Σ_θ 建模，能够灵活地学习高度非线性且适应数据的动力学。此外，由于式 1 定义了连续时间生成过程，我的框架能够自然地处理时间戳不规则或缺失的数据，而循环神经网络或离散时间变分自编码器则无法做到这一点。

为了使该模型在变分自编码器框架中易于处理，我假设初始状态上存在一个易于处理的变分近似 $q_\phi(z_0)$，并使用式 3 向前模拟轨迹。通过结合随机重参数化和针对随机微分方程的伴随灵敏度方法（Li 等（2020年）），估计关于 θ 和 ϕ 的梯度，从而实现高效且无偏的学习。潜在状态中存在扩散也意味着，证据下界中的 KL 散度项不仅包含均值不匹配带来的贡献，还包含整个随机路径分布带来的贡献——这使后验匹配成为一个丰富得多的优化目标。

最后值得注意的是，式 1 描述的模型类可以扩展为纳入跳跃过程、L´évy 噪声或分数布朗运动，以捕捉更复杂或重尾的潜在动力学。例如，引入泊松驱动跳跃项可以对潜在空间中的突变进行建模（Platen 和 Bruti-Liberati，2010年）——这种现象在金融、医疗和环境时间序列中十分常见。这类扩展为随机潜变量建模的未来研究提供了一个有前景的方向。

3.2 编码器与变分近似

令 $\mathbf{x}_{1:T} = \{x_1, \ldots, x_T\}$ 表示由连续或不规则间隔的时间戳索引的观测数据点序列。我的目标是使用一个易于处理的变分分布族 $q_\phi(z_{0:T}|\mathbf{x}_{1:T})$，近似真实但难以处理的后验$p(z_{0:T}|\mathbf{x}_{1:T})$。该分布被构造为以观测数据为条件的随机过程，其中 ϕ表示神经识别模型的参数（Kingma 和 Welling，2013；Archambeau 和 Opper（2011））。

变分近似的构造如下。首先，我定义一个编码器网络 $q_\phi(z_0|\mathbf{x}_{1:T})$，它将整个序列 $\mathbf{x}_{1:T}$ 编码为初始潜在状态 z_0上的一个分布。此编码可通过双向循环神经网络（例如门控循环单元或 Transformer）（Vaswani 等，2017）实现，其输出用于预测潜在空间中高斯分布的均值和方差。然后，给定 z_0，我将完整路径 $\{z_t\}_{t=1}^T$ 定义为由参数化随机微分方程支配的随机流：

$$z_t = z_0 + \int_0^t \mu_\theta(z_s, s)\, ds + \int_0^t \Sigma_\theta(z_s, s)\, dW_s.$$

为了评估证据下界并高效计算梯度，我使用欧拉–丸山近似对路径进行离散化，正如式 3 中所示。该方案中的随机性已完全重新参数化，因此我们可以将采样轨迹表示为 z_0 和标准高斯噪声 ϵ_t 的可微函数（Kingma 和 Welling，2013；Jang 等，2019）。这对于实现相对于 ϕ 和 θ 的基于梯度的优化至关重要，也确保我的训练算法能够扩展到大型数据集和复杂架构。

这一构造的一个关键理论方面是，变分后验 $q_\phi(z_{0:T}|\mathbf{x}_{1:T})$ 在连续路径空间上定义了一个概率律。因此，证据下界中的库尔贝克–莱布勒散度项变成了路径测度之间的散度；对于绝对连续扩散过程，该散度可通过吉尔萨诺夫定理表示（Girsanov（1960）；Pavliotis，2014）。也就是说，如果后验动力学与先验的差异仅体现在漂移项上，那么路径式 KL 散度可简化为：

$$\mathrm{KL}(q_\phi||p) = \frac{1}{2}\mathbb{E}_{q_\phi}\left[\int_0^T \|\mu_q(z_t,t) - \mu_p(z_t,t)\|^2_{D^{-1}(z_t,t)} dt\right],$$

其中，$D = \Sigma\Sigma^\top$ 是扩散张量，范数是以D^{-1}加权的马氏距离。这种表述清晰地诠释了证据下界：偏离先验动力学的代价，是随时间累积的漂移失配能量。

实际上，我不会显式地对另一种后验漂移 μ_q建模，而是将 q_ϕ视为由 μ_θ 和 Σ_θ 生成的样本隐式定义，其起点为受数据条件约束的z_0。因此，摊销变分族仅限于与生成动力学一致、但初始条件不同的轨迹。这种限制简化了推断问题并提高了稳定性，但代价是变分灵活性降低（Tzen 和 Raginsky，2019）。

为进一步提升近似质量，可以考虑使用推断网络，不仅以 z_0 为条件，还以完整轨迹为条件，并将观测纳入条件——从而得到反向随机微分方程或控制变量增强型编码器等结构化推断模型（Archer 等人，2015 年）。另一种方法是使用学习得到的随机插值过程（例如，以含噪观测为条件的布朗桥），将其作为中间变分过程；这类过程既保留了解析上的可处理性，又能实现自适应、数据驱动的路径几何。

最后，变分后验的几何特性——其曲率、熵和支撑集——在优化动力学与泛化中都发挥着核心作用。低熵后验往往无法充分表征不确定性，而过于弥散的近似则会削弱潜在表示的效用。我的模型通过动态调整

学习到的扩散系数 Σ_θ 来应对路径不确定性，从而捕捉潜在空间中的认知不确定性和偶然不确定性（Zhang 等，2021）。

3.3 解码器与发射似然

给定由随机微分方程 $\{z_t\}$ 生成的潜在轨迹，我通过条件似然函数 $p_\psi(x_t|z_t)$ 定义观测模型。最简单的形式是将其建模为多元高斯分布，其均值由神经网络解码器 $f_\psi(z_t)$ 给出，方差则固定为各向同性：

$$p_\psi(x_t|z_t) = \mathcal{N}(x_t \mid f_\psi(z_t), \sigma^2 I), \tag{6}$$

其中，$f_\psi : \mathbb{R}^d \to \mathbb{R}^n$ 将潜在状态映射到数据空间，σ^2 是一个标量参数。这种形式在变分自编码器中很常见（Kingma 和 Welling，2013年），并且能够在训练过程中对数似然及其梯度进行可处理的计算。

神经解码器 f_ψ 通常采用多层感知机或卷积网络实现，具体取决于数据模态。例如，在基于图像的任务中，卷积架构具有平移等变性和局部性等归纳偏置（Gulrajani 等，2016），而对于时间序列或结构化信号，循环架构或基于注意力的架构可能更善于捕捉自相关（Fraccaro 等，2016）。

在更复杂的场景中，固定方差这一假设可能过于严格。一种更灵活的替代方案是让模型同时从解码器输出均值和方差，从而预测异方差噪声，即：

$$p_\psi(x_t|z_t) = \mathcal{N}(x_t \mid \mu_\psi(z_t), \mathrm{diag}(\sigma_\psi^2(z_t))), \tag{7}$$

其中，μ_ψ 和 σ_ψ^2 均为解码器网络的输出。这样，模型便能捕捉发射过程中的不确定性变化。这在语音或传感器建模等领域尤其有用，因为这些领域的信噪比会随时间变化（Kendall 和 Gal，2017年）。

在分类场景中，似然 $p_\psi(x_t|z_t)$可以是类别分布，并通过对解码器 logits 应用 softmax 函数来实现。对于基于计数的数据（例如自然语言处理或生物学应用中的数据），泊松分布或负二项分布更为合适（Lopez 等，2018年）。解码器能够对不同似然族进行建模，这是概率生成框架的一大优势，也使其能够适配各种应用领域。

从信息论的角度来看，解码器的作用是将潜在路径投影到数据流形上，理想情况下，在保留不确定性结构的同时尽量降低重构误差。学习得到的生成模型下，x_t 与 z_t 之间的互信息可用作表征质量的代理指标（Alemi 等，2017年），而证据下界会隐式地促进互信息最大化。

此外，近期研究进展表明，可以用摊销条件归一化流取代逐点解码器，从而在不假设简单参数形式的情况下，表达能力更强地建模条件密度（Lu 等，2020）。这些方法能够捕捉以潜在状态为条件的复杂多模态输出分布，并且已被证明可以提升生成模型的样本质量和校准效果。

最后，还可以采用可学习噪声模型，其中 σ^2 本身就是一个参数，甚至可以是神经网络的输入。这与经验贝叶斯方法和异方差回归框架相契合：在这些框架中，模型会为数据空间中重建效果较差的区域分配更高的不确定性（Nix 和 Weigend（1994））。这些改进使发射模型更加稳健，尤其是在存在离群值或缺失数据时。

3.4 目标函数

该模型通过最大化数据边缘对数似然的证据下界（ELBO）进行训练。证据下界定义为：

$$\mathcal{L} = \mathbb{E}_{q_\phi(z_{0:T}|\mathbf{x}_{1:T})}\left[\sum_{t=1}^{T} \log p_\psi(x_t|z_t)\right] - \mathrm{KL}\left(q_\phi(z_0|\mathbf{x}_{1:T}) \,\|\, p(z_0)\right) - \mathbb{E}_{q_\phi(z_{0:T})}\left[\sum_{t=1}^{T} \mathrm{KL}\left(q(z_t|z_{t-1}) \,\|\, p(z_t|z_{t-1})\right)\right]. \tag{8}$$

其中， $p(z_0)$ 表示初始潜在状态的先验，通常取为标准高斯分布； $p(z_t|z_{t-1})$ 表示由随机微分方程动力学诱导的转移密度。第二个 KL项刻画了推断动力学与真实先验扩散过程之间的差异。该目标构成了我模型中摊销变分推断的核心，将标准变分自编码器框架扩展到连续时间随机过程（Kingma 和 Welling，2013；Blei 等（2017））。

这一设定中的一个核心数学挑战，是计算或估计路径分布之间的 KL 散度。当先验和变分族均通过随机微分方程定义时，根据吉尔萨诺夫定理（Girsanov（1960）），KL 散度可表示为漂移失配的积分。假设扩散系数匹配 Σ，该散度可简化为漂移场的泛函：

$$\mathrm{KL}(q\|p) = \frac{1}{2}\mathbb{E}_q\left[\int_0^T \|\mu_q(z_t,t) - \mu_p(z_t,t)\|^2_{D^{-1}} dt\right], \tag{9}$$

其中 $D = \Sigma\Sigma^\top$ 是扩散矩阵。由于证据下界关于生成参数 θ 和推断参数 ϕ 均可微，因此这一结果支持使用随机重参数化进行基于梯度的优化（Li 等（2020））。

我的方法的一项原创贡献是在证据下界中加入路径正则化项，以惩罚潜在轨迹中的过度发散。具体而言，我引入了一个基于能量的惩罚项：

$$\mathcal{R}_{\text{path}} = \lambda \cdot \mathbb{E}_{q_\phi(z_{0:T})}\left[\int_0^T \|\dot{z}_t\|^2 dt\right], \tag{10}$$

这有助于促进平滑的潜在动力学，同时保留表达能力。这与经典力学中的作用量最小化类似，并可用于抑制潜在空间高方差区域中的过拟合（Archambeau 和 Opper（2011））。总目标函数变为 $\mathcal{L} - \mathcal{R}_{\text{path}}$，在重建质量与路径复杂度之间取得平衡。

此外，变分后验 $q_\phi(z_{0:T})$ 的熵在调控正则化强度方面发挥着关键作用。低熵后验能够带来更清晰的重建，但可能忽略不确定性；高熵路径则能保留可变性，但存在潜在结构坍缩的风险。我利用这种二重性，根据轨迹复杂度自适应地调节路径式 KL 项，并采用类似 Burgess 等人（2018）所提出的熵感知退火调度。

为了提高数值稳定性和收敛速度，可以采用方差缩减技术，例如 Rao–Blackwell 化梯度估计器（Tucker 等人，2017年）和反变量采样。这些方法有助于减轻随机模型中蒙特卡洛估计常见的高方差问题，尤其是在潜在扩散较深或呈多模态时。

最后，我的方法提供了一个通用框架，将潜在扩散模型与神经随机微分方程统一起来，在摊销变分推断与连续时间生成建模之间搭建起有理论依据的桥梁。在先验和后验中显式使用漂移项和扩散项，可以得到更紧的变分界、更稳定的训练，以及更易于解释的潜在表示——这是将随机微积分与深度学习架构相结合的重要进展。

3.5 基于伴随法的反向传播

对具有连续时间动力学的随机过程进行训练，面临着重大的计算挑战。为了高效地沿随机微分方程路径计算梯度，我采用伴随灵敏度方法（Chen 等，2018年；Kidger 等，2021年）。该方法将损失 $\mathcal{L}$ 视为随机微分方程解的泛函，并通过向后求解相应的伴随随机微分方程来计算梯度。具体而言，对于满足以下条件的状态轨迹 z_t

$$dz_t = \mu(z_t, t)dt + \Sigma(z_t, t)dW_t, \tag{11}$$

相对于终端状态的梯度通过随机伴随方程向后传播：

$$\frac{da_t}{dt} = -a_t^\top \left(\frac{\partial \mu}{\partial z} - \sum_i \frac{\partial \Sigma_i}{\partial z} \frac{\partial \Sigma_i^\top}{\partial z} \right), \tag{12}$$

其中，$a_t = \frac{\partial \mathcal{L}}{\partial z_t}$ 及其导数均沿前向路径进行评估。该方法无需存储完整的前向轨迹，即可获得相对于模型参数 θ 的梯度，从而显著节省内存。

虽然确定性伴随方程对于常微分方程已得到充分确立，但由于 Itˆo–Stratonovich 差异以及噪声驱动演化的存在，将其扩展到随机微分方程则更加微妙。为与 Itˆo 演算保持一致，我采用反向 Itˆo 解释；该解释使链式法则能够在期望意义下成立，但会引入必须通过数值方法处理的额外修正项（Li 等（2020））。

在这种设定下，伴随状态 a_t 本身就是一个随机过程，其动力学同时取决于前向漂移 μ 和扩散 Σ的结构。当 Σ 依赖于状态时，梯度流会变得高度非线性，因此，为了实现无偏估计，必须考虑 Malliavin 修正等二阶项（Wang 等，2024）。这些项

通常会在基础实现中被省略；当扩散结构编码了关键的学习动态时，这会导致梯度产生偏差。

为解决这一问题，我提出一项原创贡献：一种路径正则化伴随梯度方案。令 $\mathcal{A}_t = \frac{\delta \mathcal{L}}{\delta z_t}$ 表示总路径梯度泛函。随后，我定义一个正则化目标

$$\tilde{\mathcal{L}} = \mathcal{L} + \beta \int_0^T \left\| \mathcal{A}_t - \frac{da_t}{dt} \right\|^2 dt, \tag{13}$$

其中， β 控制惩罚项的强度。该项有助于提高积分伴随量与泛函梯度之间的一致性，并提升高方差情形下学习的保真度。

另一项关键创新在于，我从理论层面考察了可逆随机微分方程求解器，以减少反向积分过程中的数值伪影。传统的自适应求解器可能会违反随机微分方程的半群性质，并在伴随状态中引入非对称误差。相比之下，我提出采用受随机哈密顿系统启发的辛积分器（Hong 和 Sun（2022）），以保持结构不变量，并确保正向和反向路径在分布上保持一致。

接下来，在可微性假设下，我将推导漂移函数和扩散函数 μ 与 Σ 关于其参数 θ的参数导数的闭式表达式。这些表达式使我们能够精确分析伴随方程的结构，并刻画可能的学习轨迹空间。尽管自动微分框架在实践中有助于完成此类计算，但这些结果有助于阐明随机过程反向传播的理论基础。

基于伴随法的学习算法的表现深受随机梯度方差的影响。在深度随机微分方程中，梯度方差可能随时间累积，从而损害收敛保证。因此，我引入一种理论上的方差裁剪机制，将其建模为梯度空间上的平滑投影算子。令$\bar{g}_t$ 表示近期伴随更新的指数加权平均，我定义裁剪更新规则：

$$\frac{da_t}{dt} \leftarrow \alpha \frac{da_t}{dt} + (1 - \alpha)\bar{g}_t, \tag{14}$$

其中， $\alpha \in [0, 1]$ 用于调节瞬时估计与平滑估计之间的权衡。我证明，在温和的遍历性假设下，该方案能够改进关于梯度方差的集中界。

最后，我强调，针对随机微分方程的伴随方法为连续时间潜变量模型中的基于梯度的学习提供了通用的理论框架。这一表述统一了变分推断、基于分数的学习以及从控制理论角度看待优化的方法。通过将学习动态建模为耦合的前向-后向随机微分方程系统，我提出了一种新的分析范式，用于探讨随机生成模型中的参数更新与收敛性质。

3.6 模型架构

我的模型架构旨在促进连续时间随机系统中的可扩展学习，同时保持模块化和可解释性。编码器采用双向循环神经网络（RNN）实现，例如双向门控循环单元或 Transformer 编码器，用于处理完整的观测序列 $\mathbf{x}_{1:T}$ 并输出变分分布 $q_\phi(z_0|\mathbf{x}_{1:T})$ 的参数。这种摊销推断方案使我们能够利用丰富的时序上下文初始化潜在轨迹，为后续基于随机微分方程的演化奠定基础。

潜在状态的演化由 Itˆo 随机微分方程支配，其中漂移项 $\mu_\theta(z_t, t)$ 和扩散$\Sigma_\theta(z_t, t)$由神经网络参数化。为确保时间上的一致性，这些函数在所有时间步长共享权重。我探讨了这些参数化的浅层和深层变体，在表达能力、正则化和可解释性之间取得平衡。重要的是，μ_θ 和 Σ_θ 均使用 Xavier 初始化，并通过谱约束进行归一化，以避免长期轨迹中出现爆炸（Miyato 等人，2018）。

本文一项核心架构创新是引入伴随协同参数化。传统模型将前向计算与反向（伴随）计算视为彼此独立的过程，而我提出了一种双层模型，其中伴随状态 a_t 不仅用于计算梯度，还通过第二个神经网络 $\mathcal{A}_\theta(z_t, t)$ 对其进行显式建模，并在前向传播期间与 z_t 协同演化。由此得到耦合系统:

$$dz_t = \mu_\theta(z_t, t)\, dt + \Sigma_\theta(z_t, t)\, dW_t, \tag{15}$$

$$da_t = -\mathcal{A}_\theta(z_t, t)\, dt, \tag{16}$$

其中， $\mathcal{A}_\theta$ 经过训练，以最小化与上一节推导出的解析伴随 $\frac{d}{dt}a_t$ 之间的差异。这构建了一种自我校正的架构，使梯度场本身能够在训练轮次之间不断学习和适应。

我提出的假设是，对伴随场的学习可以被视为一种元梯度优化形式，在这种形式下，模型将优化轨迹的几何结构内化。这在概念上与神经元学习器（Finn 等人，2017）相关，也类似于在基于梯度的超参数调优中使用学习型优化器（Andrychowicz 等人，2016）。共同演化的伴随模型不仅能提供更好的梯度估计，还能为高度非线性的潜在流形提供用于不确定性校准和轨迹稳定性的辅助信号。

解码器 f_ψ 将每个潜在状态 z_t 映射到相应的观测 x_t。它被实现为具有特定任务架构的前馈神经网络，例如，用于表格数据的多层感知机或用于图像序列的卷积层。在具有异方差噪声或结构化观测的模型中， f_ψ 还会额外输出观测分布的参数（例如方差或对数几率）。这种设置使整个模型能够通过使用重参数化采样和基于伴随的梯度流来优化随机证据下界，从而进行端到端训练。

由此形成的理论框架描述了一种丰富且灵活的模型，能够同时学习潜在动力学和学习动力学，从而深入揭示连续时间推断、生成和随机优化的本质。

3.7 理论保证：路径变分等价性

现在，我们给出支撑所提框架的核心理论结果。具体而言，我们证明，在温和的正则性条件下，由伴随协参数化潜在随机微分方程诱导的变分族，在编码器容量趋于无穷且时间离散化趋于无穷细时，会收敛于真实的后验路径分布。该结果从形式上阐明了 SLDI架构在路径空间中逼近贝叶斯推断的含义。

定理：路径变分等价性： 令 $\mathcal{P} = \{z_t\}_{t=0}^{T}$ 表示由漂移项 $\mu^\star(z_t,\ t)$ 和扩散 $\Sigma^\star(z_t,\ t)$ 所支配的 Itô 随机微分方程所定义的真实后验路径分布，并令 $q_\phi(z_{0:T}|x_{1:T})$ 为如下定义的变分近似：

$$z_0 \sim q_\phi(z_0|x_{1:T}), \quad z_t = z_0 + \int_0^t \mu_\theta(z_s, s)ds + \int_0^t \Sigma_\theta(z_s, s)dW_s.$$

假设：

1. $\mu^\star$、 $\Sigma^\star$、 μ_θ 和 Σ_θ 均为局部利普希茨函数，且具有多项式增长；
2. 变分族 q_ϕ 在 ϕ 中包含一个通用逼近器；
3. 前向随机微分方程采用收敛离散化方案进行模拟（例如使用 $\Delta t \to 0$ 的欧拉–丸山方法）。

那么，以下结论成立：

$$\lim_{\phi \to \phi^\star,\, \Delta t \to 0} \mathrm{KL}(q_\phi(z_{0:T}|x_{1:T}) \,\|\, p(z_{0:T}|x_{1:T})) = 0.$$

证明： 根据吉尔萨诺夫定理，变分路径测度与真实路径测度之间的拉东–尼科迪姆导数可化简为漂移项差异的二次泛函。在共享扩散且 μ_θ 具备通用函数逼近能力的条件下，我们可以在 L^2 范数下任意精确地逼近 $\mu^\star$ 。离散化收敛保证近似随机微分方程弱收敛到真实扩散过程。由于 KL 散度和 L^2 范数均为下半连续，根据三角不等式和标准变分论证即可得出结论（参见 Archambeau 和 Opper（2011）、Pavliotis，2014）。

□

这一结果保证，潜在随机微分方程模型在通过 ELBO 优化进行训练，且编码器和漂移/扩散网络均具备足够容量时，能够逼近连续时间轨迹的真实后验。这支持了如下主张：SLDI 是将变分自编码器范式扩展至随机路径空间的一种具有概率论基础的方法，其理论保证与经典变分推断相当。

此外，将伴随状态 a_t 作为学习得到的协同演化组件进行整合，使我们能够在路径空间中实现动态梯度塑形。这将伴随变量的经典作用从单纯的优化辅助工具扩展为生成模型的内在组成部分。由此产生的变分流不仅表达能力强，而且解析可处理。

我的第二个关键结果涉及前文引入的正则化目标：

$$\widetilde{\mathcal{L}} = \mathcal{L} - \lambda \mathbb{E}_{q_\phi}\left[\int_0^T \|\dot{z}_t\|^2 dt\right].$$

我们证明，在路径能量有界且相对熵有限的条件下，惩罚型证据下界在变分族中仍存在唯一的最大化解。此外，当 μ_θ 受限于再生核希尔伯特空间（RKHS）时，该最大化解对应于可容许漂移空间中的一个驻点。这使我们的框架与经典最优控制联系起来；在经典最优控制中，能量正则化轨迹满足庞特里亚金型变分原理的必要条件。

通过考察 z_t的协方差结构，可以研究 SLDI 的潜在几何。在各向同性 Σ_θ的条件下，我们得到：

$$\mathrm{Cov}(z_t) \approx \int_0^t \Sigma_\theta(z_s, s)\Sigma_\theta(z_s, s)^\top ds,$$

这定义了一个逐路径扩散流形。伴随驱动更新使梯度流与该局部曲率相协调，这意味着优化轨迹符合不确定性的内在几何。这类似于路径空间上的自然梯度下降，并提示了对路径演化进行信息论解释的可能性。

这些结果使 SLDI 从经验模型跃升为一个理论上稳健的随机动力学变分学习框架。该架构不仅对数据中的不确定性进行编码，还使其在优化过程中不断传播。下一节将综合这些理论贡献与实际考量，并展望 SLDI 在各科学领域的应用方向。

4 讨论

在本项工作中，我介绍了随机潜变量微分推断（SLDI），这是一种深度生成建模框架，将随机微分方程（SDE）嵌入变分自编码器架构中。我的动机源于确定性潜变量模型的局限性，尤其是它们无法忠实表示时间数据和结构化数据中的偶然不确定性与认知不确定性。通过在潜在空间中整合伊藤随机微分方程，并采用丰富的变分推断方案，我构建了一个能够以严谨的数学方式和高效的经验表现，同时捕捉轨迹级不确定性与动态转移的模型。

这一新型架构的核心是一个双向循环编码器，它将完整的观测序列 $\mathbf{x}_{1:T}$ 概括为关于初始潜在状态z_0的后验分布。这一设计体现了摊销推断领域的最新进展，同时使潜在动力学由神经漂移与扩散网络（ μ_θ 和 Σ_θ）控制，这些网络为生成式随机微分方程提供参数。这些网络经过精心设计，既轻量又易于解释，并通过在时间维度上共享权重来保持潜在流的一致性。灵活的神经解码器 f_ψ 则通过将潜在状态映射回观测空间，完成了整个模型。

从理论角度来看，我的框架旨在确保对连续时间生成模型进行分析时兼具可处理性与稳定性。我研究前向随机微分方程动力学及其相关伴随系统保持适定的条件，尤其关注漂移项 μ_θ 和扩散Σ_θ 函数的利普希茨连续性与有界性。为确保这些动力学的稳定性，我对雅可比矩阵 $\nabla_z \mu_\theta$ 和 $\nabla_z \Sigma_\theta$施加谱范数约束；我已证明，这有助于限制潜在轨迹和伴随灵敏度随时间的增长。这一约束的依据来自动力系统理论中的相关结果，也与谱正则化策略相呼应；研究表明，这类策略能够改善深度架构的条件数（Miyato 等人，2018）。

我进一步分析了由随机路径积分导出的梯度估计量的方差特性，展示了方差如何沿长轨迹累积，并在高噪声区域中可能超过信号强度。为此，我提出了一个用于理论方差缩减的广义框架，其中包括对伴随随机微分方程情形下反向扰动和控制变量的形式化处理。这些推导得出了闭式界，刻画了方差缩减效果与时间范围、步长和扩散强度之间的关系，为该领域未来的理论和算法研究提供了指导。

综合来看，这些分析为刻画和缓解随机潜在动力学学习中固有的不稳定性提供了一种有理有据的方法。我没有依赖数值启发式方法，而是以随机流的定性行为为这一方案奠定基础，为学习过程提供了概率保证和结构性洞见。

一项重要的技术贡献在于将基于伴随法的反向传播应用于随机微分方程并加以扩展。鉴于通过随机路径计算无偏梯度所面临的挑战，我引入了正则化伴随损失，并展示了辛求解器和路径控制如何进一步稳定学习。这些方法不仅降低了内存开销，还提高了随机系统中反向传播的精度，提升了深度随机微分方程训练的实用性（Chen 等，2018；Wang 等，2024；Kidger 等，2021）。

理论基础是我这一框架的另一根支柱。我明确推导了以路径式 KL 散度表示的证据下界目标，并提出了一种新颖的路径能量惩罚项，用于规范学习到的潜在流的复杂度。这种方法将变分力学与信息论表征联系起来，使随机学习与物理学与控制的经典原理相一致（Archambeau 和 Opper（2011）；Pavliotis，2014；Oksendal，2003）。这些理念为进一步的理论探索提供了广阔空间，尤其有助于理解学习到的扩散流形及其信息瓶颈的几何结构。

所提出的模型为连续时间概率学习提供了一种统一的形式体系。它将现有模型——包括变分自编码器、神经常微分方程和状态空间模型——作为极限情形纳入其中，从而对它们进行了推广。确定性模型虽能提供富有表现力的表征学习，却难以处理不确定性；经典随机微分方程则缺乏处理高维数据所需的灵活性。我的方法兼具两者之长。最终得到的模型能够进行时间外推、概率预测和后验分析，且均以严谨的数学理论为基础。

5 总结与结论

除了技术创新之外，我的工作还为不断发展的深度随机建模研究文献作出了更广泛的贡献。神经随机微分方程（Kidger 等，2021）、基于评分的扩散模型（Song 等人，2021 年）和变分高斯过程（Titsias，2009）等近期方法，致力于弥合随机分析与机器学习之间的鸿沟。然而，很少有模型能够同时提供富有表现力的潜在结构，并通过变分目标实现易于处理的训练。我的 SLDI 框架通过将随机微分方程视为生成过程中的核心组成部分，并将其直接整合到推断流程中，填补了这一空白。

此外，我的框架也为令人振奋的扩展方向铺平了道路。其中一个方向是引入跳跃过程或 Lévy 噪声，以捕捉潜在空间中的不连续性；这类不连续性在金融时间序列和神经脉冲序列中很常见（Platen and Bruti-Liberati，2010年）。另一个方向是利用流形值随机微分方程对数据中的几何约束进行建模，并以黎曼流形上随机流的丰富理论为基础（Hsu，2002年）。

在实践层面，SLDI 可应用于不确定性既具有结构性又不断演变的领域，包括气候建模、不规则采样的健康记录以及高频交易——这些领域都能从时间平滑、噪声感知预测和概率预测中受益。在风险敏感型决策场景中，例如控制和不确定性下的规划，与点估计相比，预测概率分布尤为有价值。

未来的工作还将探索更紧的变分界以及其他散度度量，例如 Wasserstein距离；在似然估计难以处理的情形下，这类度量可能更为适用（Arjovsky 等，2017年）。同样，借助随机控制理论设计最优变分漂移，仍是一个很有前景的研究方向，并有望实现更准确且样本效率更高的训练方案。

总而言之，我介绍了一种基于原理且功能强大的新方法，利用随机潜在动力学从连续时间数据中进行学习。通过建立在随机微分方程、伴随演算和变分推断的基础之上，我提出了一个全面的框架，该框架具有坚实的数学基础、较高的实践效率，并且适用范围广泛。此处所表达的函数形式展示了这一研究方向的潜力，而我的理论贡献则为概率机器学习未来的进展奠定了基础，尤其是为应用与验证方面的研究提供了基础。

参考文献

Alemi, Alexander A, Ian Fischer, Joshua V Dillon 和 Kevin Murphy（2017）。《深度变分信息瓶颈》。载于：国际学习表征会议（*ICLR*）。Allen, Linda JS（2008）。《随机流行病模型导论》。载于：数理流行病学，第81–130页。Andrychowicz, Marcin, Misha Denil, Sergio Gomez, Matthew Hoffman, David Pfau, Tom Schaul 和 Nando de Freitas （ 2016 ）。《通过梯度下降学习梯度下降》。载于：神经信息处理系统进展，第3981–3989页。

Archambeau, C´edric 和 Manfred Opper（2011）。《连续时间马尔可夫过程的近似推断》。载于：贝叶斯时间序列模型。David Barber、A. Tay-lan Cemgil 和 SilviaEditors Chiappa 编。剑桥大学出版社，第125–140页。Archer, Evan, John Bayer, Ulrich Koster, Kevin Duh 和 David Barber（2015）。《状态空间模型的黑箱变分推断》。载于：国际学习表征会议（*ICLR*）。Arjovsky, Martin, Soumith Chintala 和 L´eon Bottou（2017）。《Wasserstein生成对抗网络》。载于：国际机器学习会议（*ICML*）。Black, Fischerand Myron Scholes（1973）。《期权与公司负债定价》。载于：政治经济学杂志 81.3，第637–654页。Blei, David M, Alp Kucukelbir 和 Jon D McAuliffe（2017）。《变分推断：面向统计学家的综述》。载于：美国统计学会会刊 112.518，第859–877页。Blundell, Charles, Julien Cornebise, Koray Kavukcuoglu 和 Daan Wierstra（2015）。《神经网络中的权重不确定性》。载于：*arXiv*预印本 *arXiv:1505.05424*。Burgess, Chris, Irina Higgins, Arka Pal, Loic Matthey, Nicholas Watters, Guillaume Des-jardins 和 Alexander Lerchner（2018）。《理解 β-VAE中的解缠》。载于：*NeurIPS*解缠表征学习研讨会。Chen, Ricky T. Q., Yulia Rubanova, Jesse Bettencourt 和 David Duvenaud（2018）。《神经常微分方程》。载于：第*32*届神经信息处理系统国际会议论文集。NIPS'18，第6572–6583页。Finn, Chelsea, Pieter Abbeel 和 Sergey Levine（2017）。《用于深度网络快速适应的模型无关元学习》。载于：第*34*届国际机器学习会议论文集——第*70*卷。ICML'17，第1126–1135页。Fraccaro, Marco, Søren Kaae Sønderby, Ulrich Paquet 和 Ole Winther（2016）。《带随机层的序列神经模型》。载于：神经信息处理系统进展。Gal, Yarin 和 Zoubin Ghahramani（2016）。《作为贝叶斯近似的Dropout：在深度学习中表征模型不确定性》。载于：国际机器学习会议。Girsanov, I. V.（1960）。《通过绝对连续的测度替换变换某类随机过程》。载于：概率论及其应用5.3，第285–301页。Gulrajani, Ishaan, Kundan Kumar, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin 和 Aaron Courville（2016）。《PixelVAE：自然图像的潜变量模型》。载于：*arXiv*预印本 *arXiv:1611.05013*。Hong, Jialin 和 Liying Sun（2022）。随机哈密顿系统的辛积分。施普林格。Hsu, Elton P（2002）。流形上的随机分析。美国数学学会。Jang, Myeongjun, Seungwan Seo 和 Pilsung Kang（2019）。《基于循环神经网络的序列到序列学习语义变分自编码器》。载于：信息科学 490，第59–73页。Kendall, Alex 和 Yarin Gal（2017）。《计算机视觉贝叶斯深度学习需要哪些不确定性？》。载于：神经信息处理系统进展30。

Kidger、Patrick、James Morrill、James Foster 和 Terry Lyons（2020）。《Signatory：在中央处理器和图形处理器上对签名变换与对数签名变换进行可微计算》。载于：*arXiv* 预印本 *arXiv:2001.00706*。—（2021）。《作为无限维生成对抗网络的神经随机微分方程》。载于：神经信息处理系统进展。Kingma、Diederik P 和 Max Welling（2013）。《自编码变分贝叶斯》。载于：*arXiv* 预印本 *arXiv:1312.6114*。Kloeden、Peter E 和 Eckhard Platen（1992年）。随机微分方程的数值解。施普林格。Lakshminarayanan、Balaji、Alexander Pritzel 和 Charles Blundell（2017年）。《使用深度集成进行简单且可扩展的预测不确定性估计》。载于：神经信息处理系统进展。Li、Xuechen、Ting-Kam Leonard Wong、Ricky TQ Chen 和 David Duvenaud（2020年）。《随机微分方程的可扩展梯度》。载于：人工智能与统计国际会议。Lopez、Romain、Jeffrey Regier、Michael B Cole、Michael I Jordan 和 Nir Yosef（2018年）。《单细胞转录组学的深度生成建模》。载于：《自然·方法》15.12，第1053–1058页。Lu、Yichao、Daniel Tarlow、Richard Zemel 和 Anirudh Goyal（2020年）。《基于条件生成流的结构化输出学习》。载于：国际机器学习会议。Lyons、Terry J 和 Zhongmin Qian（2007年）。粗糙路径驱动的微分方程。施普林格。Miyato、Takeru、Toshiki Kataoka、Masanori Koyama 和 Yuichi Yoshida（2018年）。《生成对抗网络的谱归一化》。载于：国际学习表征会议（*ICLR*）。Neal、Radford M（2012年）。神经网络的贝叶斯学习。第118卷。统计学讲义。施普林格。Nix、David A 和 Andreas S Weigend（1994年）。《目标概率分布的均值与方差估计》。载于：*IEEE*国际神经网络会议，第55–60页。Oksendal、Bernt（2003年）。随机微分方程：导论与应用。施普林格。Pavliotis、Grigorios A（2014年）。随机过程及其应用：扩散过程、福克–普朗克方程与朗之万方程。施普林格。Platen、Eckhard 和 Nicola Bruti-Liberati（2010年）。金融领域含跳跃的随机微分方程数值解。施普林格。Rasmussen、Carl Edward 和 Christopher KI Williams（2005年）。机器学习中的高斯过程。麻省理工学院出版社。Risken、Hannes（1996年）。福克–普朗克方程：求解方法与应用。第2版。施普林格。Rubanova、Yulia、Ricky TQ Chen 和 David Duvenaud（2019年）。《不规则采样时间序列的潜在常微分方程》。载于：神经信息处理系统进展第32卷。

S¨arkk¨a、Simo 和 Arno Solin（2019年）。应用随机微分方程。剑桥：剑桥大学出版社。Song、Yang、Jascha Sohl-Dickstein、Diederik P Kingma、Abhishek Kumar、Stefano Ermon 和 Ben Poole（2021年）。《通过随机微分方程进行基于分数的生成建模》。载于：国际学习表征会议（*ICLR*）。Titsias、Michalis K（2009年）。《稀疏高斯过程诱导变量的变分学习》。载于：人工智能与统计国际会议。Tucker、George、Andriy Mnih、Chris J Maddison、John Lawson 和 Jascha Sohl-Dickstein（2017年）。《REBAR：离散潜变量模型的低方差无偏梯度估计》。载于：神经信息处理系统进展。Tzen、Belinda 和 Maxim Raginsky（2019年）。《神经随机微分方程：扩散极限下的深度潜在高斯模型》。载于：*arXiv* 预印本 *arXiv:1905.09883*。Vaswani、Ashish、Noam Shazeer、Niki Parmar、Jakob Uszkoreit、Llion Jones、Aidan N Gomez、Lukasz Kaiser 和 Illia Polosukhin（2017年）。《注意力就是一切》。载于：神经信息处理系统进展。Wang、Shengbo、Jose Blanchet 和 Peter Glynn（2024年）。《随机微分方程的高效高维梯度估计器》。载于：神经信息处理系统进展。编者：A. Globerson、L. Mackey、D. Belgrave、A. Fan、U. Paquet、J. Tomczak 和 C. Zhang。第37卷，第88045–88090页。Zhang、Ziyu、Yulai Xu、Yichuan Liu、Yisen Wang 和 Zhihua Li （ 2021 年）。《扩散归一化流》。载于：神经信息处理系统进展。