利用变分贝叶斯神经网络从 N 体模拟中约束宇宙学参数

H´ector J. Hort´ua [1,2,*]，**Luz ´Angela Garc´ıa** [3] **和 Leonardo Casta˜neda C.** [4]

[1] *Grupo Signos*，数学系，*Universidad el Bosque*，波哥大，哥伦比亚。

[2] 数据科学硕士课程，*Universidad Escuela Colombiana de Ingenier´ıa Julio Garavito*，波哥大，哥伦比亚。
[3] *Universidad ECCI*。*Cra. 19 No. 49-20*，波哥大，哥伦比亚，邮政编码 *111311*。
[4]哥伦比亚国立大学国家天文台，波哥大，哥伦比亚。

通讯作者*:
hjhortuao@unal.edu.co

arXiv:2301.03991v1 [astro-ph.IM] 9 Jan 2023

摘要

深度学习方法凭借从复杂数据中提取信息的能力，近年来已应用于天体物理参数恢复。其中一种方法是近似贝叶斯神经网络（贝叶斯神经网络），它已证明能够在参数空间中产生一致的后验分布，有助于不确定性量化。然而，与所有现代神经网络一样，它们往往会给出过度自信的不确定性估计，并且在将贝叶斯神经网络应用于数据时可能引入偏差。在本研究中，我们采用乘性归一化流（混合归一化流）——一种针对贝叶斯神经网络参数的近似后验分布族，旨在提高变分后验分布的灵活性——从 QUIJOTE模拟中提取 Ω_m、 h 和 σ_8 。我们将此方法与标准贝叶斯神经网络及 Flipout 估计器进行了比较。结果发现，混合归一化流与贝叶斯神经网络结合后优于其他模型，其预测性能比标准贝叶斯神经网络高出近一个数量级， σ_8 提取结果精度很高（$r^2 = 0.99$），且不确定性估计精确。这意味着混合归一化流能够提供更现实、接近真实后验分布的预测分布，减轻变分近似引入的偏差，并支持使用校准良好的网络。

关键词：宇宙学，N 体模拟，参数估计，人工智能，深度神经网络

1 引言

宇宙学模拟为理解宇宙的初始条件以及深化我们对基础物理学的认识提供了最有力的方法之一（1）。它们还使我们能够充分探索线性阶段和非线性阶段的结构形成。目前，宇宙学标准模型 Λ-CDM 仅使用少量参数，就能准确描述从宇宙早期到晚期的大多数观测结果（2）。近期的宇宙微波背景（CMB）观测对宇宙学参数给出了如此精确的估计，从而引发了与局域尺度测量结果之间的张力，同时也揭示了总非相对论物质密度参数中一个众所周知的简并性（3，4，5）。通常，从

天文观测中提取信息的方法，是将数据中的汇总统计量与理论预测进行比较。然而，这会带来两个主要困难：首先，人们尚未充分了解应采用何种估计量，或者应将次序统计量近似到何种程度，才能从观测中提取最大信息。事实上，最常见的选择是功率谱（PS），事实证明它是进行推断的有力工具（2）。然而，众所周知，PS 无法充分刻画非高斯密度场的统计性质，因此不适用于即将开展的大尺度结构（LSS）或 21 厘米信号研究，因为这些信号具有高度非高斯性（6，7，8）。因此，如果仅使用这一统计量进行参数恢复，PS 将会遗漏相关信息（9）。其次，宇宙学家需要存储和处理大量数据，这可能代价高昂。显然，为了解释这些观测结果，必须开发先进的计算工具，并对数据的采集、存储和分析提出新的思路（10）。

近年来，人工智能（AI）和深度神经网络（DNN）已成为应对宇宙学领域上述难题的有力工具。这归功于它们能够从复杂数据中学习变量之间的关系、超越传统估计量的表现，并满足天体物理学和宇宙学中高昂的计算需求（10）。由于具有解决反问题的潜力，这些标准深度神经网络已被用于各种任务。然而，它们需要调整的参数过多，且无法解释针对特定实例所作的预测，因此容易过拟合（11）。后一问题对于宇宙学分析至关重要，因为评估模型预测的稳健性和可靠性必不可少。为深度神经网络赋予概率属性，可解决这一问题：这些属性能够量化模型输出的后验分布，并提供预测不确定性。其中一种方法是采用贝叶斯神经网络（BNN），其由能够捕捉神经网络参数（权重）不确定性的概率层构成，并通过贝叶斯推断进行训练（12）。已有多项研究将贝叶斯神经网络用于宇宙学场景：结合深度神经网络（通过卷积神经网络，即 CNN）与概率属性，可以构建适用于非高斯数据、无需预先选择汇总统计量的模型（9、13、14、15），同时量化预测不确定性（16、17、18、19、20、21）。事实上，贝叶斯神经网络能够推断权重的后验分布，而非给出点估计。这些分布刻画了参数不确定性；随后对其进行积分，便可获得与神经网络输出相关的不确定性。然而，后验分布难以直接求得，因此必须采用变分推断（VI）等近似方法才能将其付诸实践（22）。尽管 VI 中采用的权重近似后验分布显然能够加快推断任务的计算，但其也可能引入一定程度的偏差，具体取决于所选近似分布族的复杂程度（或简单程度）（23）。这一问题会导致不确定性预测过于自信，且与真实后验的接近程度不尽如人意。在（17、18）中，作者在贝叶斯神经网络之上加入了归一化流，以提高联合参数分布的灵活性。然而，该方法并未纳入贝叶斯框架，因此仍然存在偏差。

本文尝试通过采用乘性归一化流，提升神经网络权重的近似后验分布的灵活性，从而使贝叶斯神经网络能够提供准确且精确的不确定性估计。我们将这一方法应用于取自 QUIJOTE 数据集 (24) 的 N 体模拟，以展示贝叶斯神经网络不仅能够利用非高斯信号，而无需指定汇总统计量（如功率谱），还能够提高后验复杂度，从而带来更显著的性能提升。本文结构如下：第 2 节概述贝叶斯神经网络框架，并详细介绍归一化流的实现。第 3 节介绍本文使用的数据集和分析工具。第 4 节介绍贝叶斯神经网络的数值实现与配置。

第 5 节展示了我们通过采用不同方法训练贝叶斯神经网络所获得的结果，并给出宇宙学参数的推断结果。本节还介绍用于确定不确定性估计准确性的校准图。最后，第 6 节总结本文的主要结论，并探讨贝叶斯神经网络在宇宙学中的应用方向。

2 变分贝叶斯神经网络

这里我们将详细介绍贝叶斯神经网络（BNN）及其在参数推断中的应用。我们先简要介绍相关内容，然后重点探讨如何改进变分近似。更多细节请参阅文献（25、26、22）。

2.1 近似贝叶斯神经网络

贝叶斯神经网络的目标是在观测到数据 $\mathcal{D} = (X,\ Y)$ 后，推断神经网络权重 w 的后验分布 $p(w|\mathcal{D})$。该后验分布可根据贝叶斯定律求得：$p(w|\mathcal{D}) \sim p(\mathcal{D}|w)p(w)$，其中给定似然函数 $p(\mathcal{D}|w)$ 以及权重的先验分布 $p(w)$。计算出后验分布后，新测试样本 x^* 上的概率分布为

$$p(y^*|x^*, \mathcal{D}) = \int_w p(y^*|x^*, w)p(w|\mathcal{D})dw, \tag{1}$$

其中，$p(y^*|x^*,\ w)$ 表示给定权重值时的预测分布。然而，对于神经网络而言，计算精确的后验分布是不可行的，因此必须采用近似贝叶斯神经网络进行推断（26）。一种常用的后验分布近似方法是变分推断（VI）（22）。设$q(w|\theta)$为一族由 θ参数化的简单分布。因此，VI 的目标是选择一个分布$q(w|\theta^*)$，使得 θ^* 最小化 $KL\big[q(w|\theta)\big\|p(w|\mathcal{D})\big]$，其中 $KL[\cdot\|\cdot]$ 为库尔贝克-莱布勒散度。该最小化等价于最大化证据下界（ELBO）（26）

$$\mathrm{ELBO}(\theta) = \mathbb{E}_{q(w|\theta)}\big[\log p(Y|X, w)\big] - \mathrm{KL}\left[q(w|\theta)\big\|p(w)\right], \tag{2}$$

其中，$\mathbb{E}_{q(w|\theta)}[\log\ p(Y|X,\ w)]$ 表示相对于变分后验的期望对数似然，而 $\mathrm{KL}[q(w|\theta)||p(w)]$ 表示变分后验与先验分布之间的散度。由公式 2 可知，KL 散度充当正则化项，促使变分后验向先验分布的众数移动。变分后验的一种常见选择是相互独立的（即均值场）高斯分布，对于神经网络中的每个参数 w 分别对应一个分布（25）

$$q(w|\theta) = \prod_{ij} \mathcal{N}(w; \mu_{ij}, \sigma_{ij}^2) \tag{3}$$

其中，i 和 j 分别是前一层和当前层神经元的索引。应用重参数化技巧后，我们得到 $w_{ij} = \mu_{ij} + \sigma_{ij} * \epsilon_{ij}$，其中$\epsilon_{ij}$ 取自标准正态分布。此外，如果先验分布也是相互独立的高斯分布的乘积，则先验分布与变分后验之间的 KL 散度可以通过解析方式计算，从而使该方法具有较高的计算效率。

2.1.1 Flipout

如果从 $q(w|\theta)$ 中采样时，小批量中不同样本的采样并非完全独立，我们将得到方差较大的梯度估计。Flipout方法提供了一种替代方案，通过对

每个样本隐式采样伪独立权重，从而使小批量内的梯度去相关（27）。该方法对 $q(w|\theta)$ 的性质有两个假设：关于零对称，且神经网络的权重相互独立。在这些假设下，分布在与随机符号矩阵 $\hat{}r$ 逐元素相乘后保持不变，即 $\hat{}w = w \circ \hat{}r$ 意味着 $w \sim q(w) \approx \hat{}w \sim \hat{}q(\hat{}w)$。因此，针对各个样本计算所得梯度的边缘分布，与使用共享权重样本计算所得的分布相同。由此，Flipout 在对一个小批量求平均时，能够实现方差低得多的更新。我们通过与乘性归一化流进行比较，在实验中验证了这一方法。

2.2 贝叶斯神经网络中的不确定性

贝叶斯神经网络为从后验分布中纳入两类不确定性提供了基础：数据本身固有的不确定性（偶然不确定性），以及由于训练数据有限而导致的模型参数不确定性（认知不确定性）（28）。根据文献（16），假设贝叶斯神经网络的顶层由均值向量 $\mu \in \mathbb{R}^N$ 和协方差矩阵 $\Sigma \in \mathbb{R}^{N(N+1)/2}$[1] 组成，对于给定的固定输入 x^*，对神经网络进行 T 次前向传播，得到每次传播对应的均值 μ_t 和协方差矩阵 Σ_t。于是，预测协方差的近似估计量可写为

$$\widehat{\mathrm{Cov}}(y^*, y^*|x^*) \approx \underbrace{\frac{1}{T}\sum_{t=1}^{T}\Sigma_t}_{\text{Aleatoric}} + \underbrace{\frac{1}{T}\sum_{t=1}^{T}(\mu_t - \overline{\mu})(\mu_t - \overline{\mu})^{\mathrm{T}}}_{\text{Epistemic}}, \tag{4}$$

其中 $\overline{\mu} = \frac{1}{T}\sum_{t=1}^{T}\mu_t$。注意，当 Σ 为对角矩阵，且 $\sigma^2 =$ 为 diag(Σ)时，最后一个方程可简化为文献（29，30）中的结果

$$\widehat{\mathrm{Var}}(y^*|x^*) \approx \underbrace{\frac{1}{T}\sum_{t=1}^{T}\sigma_t^2}_{\text{Aleatoric}} + \underbrace{\frac{1}{T}\sum_{t=1}^{T}(\mu_t - \bar{\mu})^2}_{\text{Epistemic}}. \tag{5}$$

在这种情况下，可以使用贝叶斯神经网络学习目标值之间的相关性，并估计其不确定性。遗憾的是，根据公式 4、5 计算得到的不确定性往往校准不佳，即预测的不确定性（同时考虑认知不确定性和偶然不确定性）被低估，因而无法在推断时可靠地检测出不确定的预测结果。因此，在使用贝叶斯神经网络进行预测推断之前，必须采用校准图以及联合校准偶然不确定性和认知不确定性的方法（31）。我们将在第 5 节回到这一点。

2.3 乘法归一化流

如前所述，贝叶斯神经网络中变分后验最常用的分布族，是式 3 中定义的均值场高斯分布。这种简单的分布无法捕捉真实后验的复杂性。因此，我们认为，增加变分后验的复杂性能够使贝叶斯神经网络显著提升性能，因为这样我们便能够从更复杂、也更接近真实后验的分布中进行采样。当然，对变分后验进行变换必须兼顾快速计算与数值可处理性。现在，我们详细介绍乘性归一化流（MNFs）方法；该方法能够以一种灵活的方式表示后验分布，并在

[1] 目标值 $y \in \mathbb{R}^N$。

高效地运用文献（32）提出的辅助随机变量和归一化流。MNFs 提出，变分后验可以表示为无穷混合分布

$$q(w|\theta) = \int q(w|z,\theta)q(z|\theta)dz \tag{6}$$

其中，θ 是可学习后验参数，$z \sim q(z|\theta) \equiv q(z)$[2] 是一个与输入层维度相同的向量，充当辅助潜变量。此外，通过允许局部重参数化，全连接层的变分后验可改写为公式 3 的一种形式，即

$$w \sim q(w|z) = \prod_{ij} \mathcal{N}(w; z_i\mu_{ij}, \sigma_{ij}^2). \tag{7}$$

请注意，通过提高 $q(z)$的复杂度，我们可以增加变分后验的灵活性。由于 z 的维度远低于权重的维度，因此可以使用归一化流来实现这一点。从公式 3 中完全因子化高斯分布的样本 $z_0 \sim q(z_0)$ 出发，通过在 z_0上依次应用 K 个可逆变换 f_K，可以得到丰富的分布 $q(z_K)$

$$z_K = \mathrm{NF}(z_0) = f_K \circ \cdots \circ f_1(z_0); \quad \log q(z_K) = \log q(z_0) - \sum_{k=1}^{K} \log \left| \det \frac{\partial f_k}{\partial z_{k-1}} \right|. \tag{8}$$

遗憾的是，公式 2 中的 KL 散度通常变得难以处理，因为后验分布 $q(w)$如公式 6 所示是一个无限混合分布。文献（33）也通过调用贝叶斯定律 $q(z_K)q(w|z_K) = q(w)q(z_K|w)$并引入一个由 ϕ参数化的辅助分布 $r(z_K|w, \phi)$，解决了这一问题，其目的是近似原始变分参数的后验分布 $q(z_K|w)$，从而进一步对 KL 散度项给出下界。因此，KL 散度项可界定如下

$$\begin{aligned}
-\mathrm{KL}\left[q(w)\middle\|p(w)\right] &= -\mathbb{E}_{q(w)}\left[\log\left(\frac{q(w)}{p(w)}\right)\right] \\
&\geq -\mathbb{E}_{q(w)}\left[\log\left(\frac{q(w)}{p(w)}\right) + \mathrm{KL}\left[q(z_K|w)\middle\|r(z_K|w,\phi)\right]\right] \\
&= -\mathbb{E}_{q(w)}\left[\log\left(\frac{q(w)}{p(w)}\right) + \mathbb{E}_{q(z_K|w)}\left[\log\left(\frac{q(z_K|w)}{r(z_K|w,\phi)}\right)\right]\right] \\
&= -\mathbb{E}_{q(w)}\left[\mathbb{E}_{q(z_K|w)}\left[\log\left(\frac{q(w)}{p(w)}\right)\right] + \mathbb{E}_{q(z_K|w)}\left[\log\left(\frac{q(z_K|w)}{r(z_K|w,\phi)}\right)\right]\right] \\
&= -\mathbb{E}_{q(w,z_K)}\left[\log\left(\frac{q(w)}{p(w)}\right) + \log\left(\frac{q(z_K|w)}{r(z_K|w,\phi)}\right)\right] \\
&= \mathbb{E}_{q(w,z_K)}\left[-\log\left(q(w)q(z_K|w)\right) + \log r(z_K|w,\phi) + \log p(w)\right] \Rightarrow \\
-\mathrm{KL}\left[q(w)\middle\|p(w)\right] &\geq \mathbb{E}_{q(w,z_K)}\left[-\mathrm{KL}\left[q(w|z_K)\middle\|p(w)\right] + \log q(z_K) + \log r(z_K|w,\phi)\right],
\end{aligned} \tag{9}$$

其中我们考虑了 $\mathrm{KL}[P\|Q] \geq 0$，且当且仅当 $P = Q$ 时等式成立。在最后一行中，第一项可以进行解析计算，因为它是两个高斯分布之间的 KL 散度；第二项则由 f_K 生成的归一化流给出，正如我们在

[2] 为简化符号表示，本节将省略参数 θ。

公式 8。最后，a 辅助后验分布项通过逆归一化流参数化如下（34 ）

$$z_0 = \mathrm{NF}^{-1}(z_K) = g_1^{-1} \circ \cdots \circ g_K^{-1}(z_K); \quad \log r(z_K|w,\phi) = \log r(z_0|w,\phi) + \sum_{k=1}^{K} \log \left| \det \frac{\partial g_k^{-1}}{\partial z_k} \right|, \quad (10)$$

其中可以将其参数化为 g_K^{-1}另一个归一化流。在论文（32）中，作者还提出了将辅助后验分布灵活参数化为

$$z_0 \sim r(z_K|w,\phi) = \prod_i \mathcal{N}(z_0; \tilde{\mu}_i(w,\phi), \tilde{\sigma}_i^2(w,\phi)). \quad (11)$$

我们将采用原论文中均值 $\tilde{\mu}$ 和方差 $\tilde{\sigma}^2$ 的参数化方式，并选择掩码 RealNVP（35）作为归一化流。

3 N 体模拟数据集

在本研究中，我们使用了取自 Quijote 项目 (24) 的 2000 个超立方体模拟。这些模拟使用 TreePM 代码 Gadget-III (36) 运行，其初始条件在$z = 127$ 处通过二阶拉格朗日微扰理论 (37) 生成。本研究选用的是具有不同随机种子的标准模拟，旨在模拟宇宙方差。每个样本对应一个大小为 64^3 的三维密度场分布。宇宙学参数 $\Omega_m \in [0.1, 0.5]$、$\Omega_b \in [0.03, 0.07]$、 $h \in [0.5, 0.9]$、 $n_s \in [0.8, 1.2]$和 $\sigma_8 \in [0.6, 1.0]$各不相同，而中微子质量（$M_\nu = 0$eV）和状态方程参数（$w = -1$）保持不变。数据集被划分为训练集（70%）、验证集（10%）和测试集（20%）；超立方体经过对数变换，宇宙学参数则归一化到 0 至 1 之间。本文将构建贝叶斯神经网络，使其能够预测上述五个参数中的三个：Ω_m、 σ_8 和 h。

4 贝叶斯神经网络的实现

根据第 2 节的讨论，我们将考察三种不同的贝叶斯神经网络架构：标准贝叶斯神经网络（先验和变分后验均定义为平均场正态分布） **[随机贝叶斯神经网络]**；采用 Flipout 估计器的贝叶斯神经网络 **[Flipout 贝叶斯神经网络]**；以及乘性归一化流 **[变分贝叶斯神经网络]**。实验使用 TensorFlow v:2.9 和 TensorFlow-probability v:0.19(38) 实现。本文设计的所有贝叶斯神经网络均由三部分组成。首先，所有实验都以一个 64^3 体素输入层开始，该层对应于归一化的三维密度场，后接全卷积 ResNet-18 骨干网络，如表1所示。所有残差块均采用完全预激活结构，其表示形式见图 1。我们使用 Classification models 3D 仓库构建贝叶斯神经网络的骨干网络 (39)。其次，贝叶斯神经网络的第二部分用于表示神经网络的随机性。该部分仅由一层组成，具体取决于所用贝叶斯神经网络的类型。对于随机贝叶斯神经网络，我们采用稠密变分层，通过变分推断对核矩阵和偏置项的分布拟合近似后验分布。这里，我们将正态分布用作后验分布和先验分布（不可训练）。例如，采用*Flipout* 贝叶斯神经网络 的实验使用 Flipout 稠密层，同样利用平均场正态分布对这些分布进行参数化。这两种层已在 TF-probability 软件包中实现 (38)。另一方面，对于变分贝叶斯神经网络，我们将 TF-MNF 、MNF-VBNN (32) 仓库中实现的 DenseMNF 类适配到我们的模型中。这里，我们为掩码 RealNVP 归一化流使用 50 层，层权重的最大方差约为 1。最后一部分

ResNet-18 骨干网络		
层名称	**输入形状**	**输出形状**
批归一化	(N_{batch}, 64,64,64,3)	(N_{batch}, 64,64,64,3)
三维卷积	(N_{batch}, 70,70,70,3)	(N_{batch}, 32,32,32,64)
批归一化+ReLU	(N_{batch}, 32,32,32,64)	(N_{batch}, 32,32,32,64)
三维最大池化	(N_{batch}, 34,34,34,64)	(N_{batch}, 16,16,16,64)
批归一化+ReLU	(N_{batch}, 16,16,16,64)	(N_{batch}, 16,16,16,64)
残差块1	$\left[(N_{\text{batch}}, 16,16,16,64)\ (N_{\text{batch}}, 16,16,16,64)\right]$	(N_{batch}, 16,16,16,64)
批归一化+ReLU	(N_{batch}, 16,16,16,64)	(N_{batch}, 16,16,16,64)
残差块 2	$\left[(N_{\text{batch}},\ 16,\ 16,\ 16,\ 64)\ (N_{\text{batch}},\ 8,\ 8,\ 8,\ 128)\right]$	(N_{batch}, 8,8,8,128)
批归一化+ReLU	(N_{batch}, 8,8,8,128)	(N_{batch}, 8,8,8,128)
残差块 3	$\left[(N_{\text{batch}},\ 8,\ 8,\ 8,\ 128)\ (N_{\text{batch}},\ 4,\ 4,\ 4,\ 256)\right]$	(N_{batch}, 4,4,4,256)
批归一化+ReLU	(N_{batch}, 4,4,4,256)	(N_{batch}, 4,4,4,256)
残差块 4	$\left[(N_{\text{batch}},\ 4,\ 4,\ 4,\ 256)\ (N_{\text{batch}},\ 2,\ 2,\ 2,\ 512)\right]$	(N_{batch}, 2,2,2,512)
批归一化+ReLU	(N_{batch}, 2,2,2,512)	(N_{batch}, 2,2,2,512)
全局平均池化	(N_{batch}, 2,2,2,512)	(N_{batch}, 512)

表1。配　置，用于本文所有实验的骨干贝叶斯神经网络 per.

所有贝叶斯神经网络的最后一部分负责神经网络的输出，该输出取决于偶然不确定性的参数化。我们使用一个三维多元高斯分布，其中有九个待学习参数（三个均值 μ 对应宇宙学参数，六个元素对应协方差矩阵 Σ）。

训练过程中需要优化的损失函数由证据下界2给出，其中第二项与负对数似然（NLL）相关。

$$-\,\text{NLL} \sim \frac{1}{2}\log|s\cdot\Sigma| + \frac{1}{2}(\mathbf{y}-\mu)^{\top}(s\cdot\Sigma)^{-1}(\mathbf{y}-\mu), \tag{12}$$

在小批量上取平均。标量变量 s 在训练过程中等于 1，并在训练后成为可训练变量，以重新校准概率密度函数（16, 31）。用于最小化目标函数的算法是 Adam 优化器，其一阶和二阶矩指数衰减率分别为 0.9 和 0.999（40）。学习率从 10^{-3} 开始；如果连续 10 个 epoch 未观察到任何改进，则将其乘以 0.8。此外，我们应用了预热期，使模型逐步启用公式 2 中的 KL 项。具体而言，我们在证据下界中引入了一个 β 变量，即 $\beta\cdot\text{KL}\left[q(w|\theta)\middle\|p(w)\right]$；因此，该参数初始为 0，并在 10 个 epoch 内线性增长至 1（41）。贝叶斯神经网络使用 32 个批次进行训练，并采用早停回调以避免过拟合。所使用的基础设施是谷歌云平台（GCP），配置为一块 16 GB GDDR6 显存的 NVIDIA Tesla T4，运行于 N1 机器系列共享核心实例上。

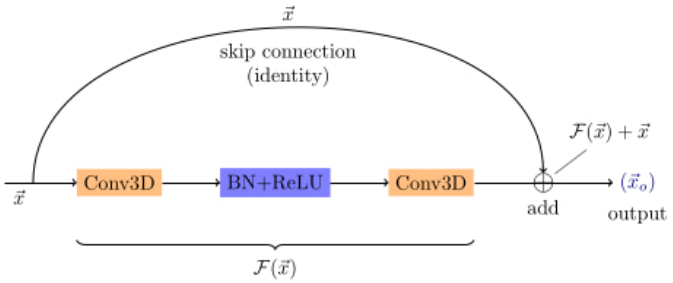


图 1a。 残差块中第一个跳跃连接的示意图。

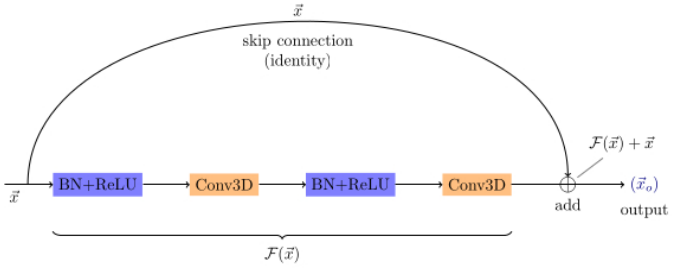


图 1b。 残差块中第二个跳跃连接的示意图。

图 1。 每个残差块都包含两种跳跃连接配置。（**A**）残差块首先将此配置应用于输入张量。（**B**）上一配置的输出被输入此连接。

4.1 度量指标

我们根据性能比较所有贝叶斯神经网络的结果，即通过均方误差（MSE）、证据下界以及绘制真实值与预测值并计算其决定系数，来量化其对宇宙学参数的预测精度。此外，量化不确定性估计的质量也十分重要。诊断不确定性估计质量的方法之一是使用可靠性图。遵循（31, 11），我们可以将回归不确定性的完美校准定义为

$$\mathbb{E}_{\hat{\sigma}^2}\left[\left|\left(\mathbb{E}[(\boldsymbol{y}-\boldsymbol{\mu})^2]\,\middle|\,\hat{\sigma}^2=\alpha^2\right)-\alpha^2\right|\right] \quad \forall\left\{\alpha^2\in\mathbb{R}\,\middle|\,\alpha^2\geq 0\right\}. \tag{13}$$

因此，预测的不确定性 $\hat{}\sigma^2$ 被划分为宽度相等的 K 个区间，每个区间的方差定义为

$$\mathrm{var}(B_k) := \frac{1}{|B_k|}\sum_{i\in B_m}\frac{1}{N}\sum_{n=1}^{N}\left(\boldsymbol{\mu}_{i,n}-\boldsymbol{y}_i\right)^2, \tag{14}$$

通过 N 次随机前向传播。另一方面，每个区间的不确定性定义为

$$\mathrm{uncert}(B_k) := \frac{1}{|B_k|}\sum_{i\in B_k}\hat{\sigma}_i^2. \tag{15}$$

利用这两个量，我们可以通过绘制 $\mathrm{var}(B_k)$ 与 $\mathrm{uncert}(B_k)$ 的关系图来生成可靠性图，从而评估估计不确定性的质量。此外，我们还可以计算预期不确定性

度量指标 Flipout贝叶斯神经网络						变分贝叶斯神经网络					随机贝叶斯神经网络				
	Ω_m	σ_8	h	$\Omega_m h^2$	$\sigma_8\Omega_m^{0.25}$	Ω_m	σ_8	h	$\Omega_m h^2$	$\sigma_8\Omega_m^{0.25}$	Ω_m	σ_8	h	$\Omega_m h^2$	$\sigma_8\Omega_m^{0.25}$
MSE	0.063					0.057					0.190				
ELBO	20.85					19.71					31.57				
r^2	0.82	0.98	0.2	0.03	0.93	0.85	0.99	0.4	0.56	0.95	0.75	0.85	0.01	0.23	0.80
UCE	0.109	8.10	0.26			0.0008	0.0008	0.010			$>$1.0				

表 2. 所有贝叶斯神经网络架构在测试集上的度量指标结果。较高的不确定性校准误差值表示校准不佳。均方误差和证据下界仅针对宇宙学参数计算。

校准误差（UCE），以量化校准偏差

$$\mathrm{UCE} := \sum_{k=1}^{K} \frac{|B_k|}{m} \big|\mathrm{var}(B_k) - \mathrm{uncert}(B_k)\big|, \tag{16}$$

输入数量为 m，输入索引集合为 B_k，其不确定性落入区间 k。文献 (16) 提出了一种更通用的方法，即计算期望覆盖概率：参数真实值落在联合后验分布所定义的 x% 置信区域内的样本所占百分比。显然，这种方法更精确，因为它通过完整的后验分布捕捉了高阶统计量。不过，为了简便起见，我们将采用不确定性校准误差方法。

5 贝叶斯神经网络参数推断的分析与结果

本节将讨论三种不同贝叶斯神经网络版本的比较结果：采用混合归一化流的版本、标准贝叶斯神经网络，以及使用 Flipout 作为估计量的第三种版本。本节报告的结果均基于测试数据集计算。表 2 展示了各贝叶斯神经网络方法得到的度量指标。如前所述，均方误差、证据下界和 r^2 能够为评估模型精度提供良好的估计，而不确定性校准误差则用于衡量校准情况。在这里可以看到，变分贝叶斯神经网络在所有实验中表现最佳，不仅平均误差更低，而且对每个宇宙学参数的估计都更精确，同时其不确定性预测也具有良好的校准效果。其次是变分贝叶斯网络，然后是 Flipout 贝叶斯神经网络。不过，尽管这种方法能够较好地估计宇宙学参数，却低估了其不确定性。因此，变分贝叶斯神经网络确实避免了在机器学习流水线中额外应用与校准相关的训练后步骤。请注意，在所有实验中，h 都很难被模型预测。图 2 展示了 Ω_m、ω_m（而非 h）、σ_8 的预测值与真实值的对比，以及由 $\sigma_8\Omega_m^{0.25}$ 定义的简并方向。误差棒表示贝叶斯神经网络预测的认知不确定性与偶然不确定性之和，体现了此类概率模型的优势：模型能够给出确定性预测，而传统深度神经网络只能提供点估计。此处的不确定性取自协方差矩阵的对角部分。

5.1 校准指标

在图 3 中，我们使用校准图分析不确定性测量的质量。我们展示了模型在测试集上的预测不确定性与观测不确定性的对比。性能更好的不确定性估计应与虚线更加准确地吻合。可以看出，变分贝叶斯神经网络对不确定性的估计更能反映真实的不确定性。此外，变分贝叶斯神经网络的尺度比 Flipout贝叶斯神经网络低两个数量级，这也说明了根据其预测结果判断该模型的可靠程度。请注意，即使我们将方差划分为 $K = 10$ 个等宽区间，Flipout 贝叶斯神经网络和随机贝叶斯神经网络仍会低估不确定性（许多样本集中在较低区间值内）。因此可以看到，变分贝叶斯神经网络在校准图中覆盖了全部十个样本，而

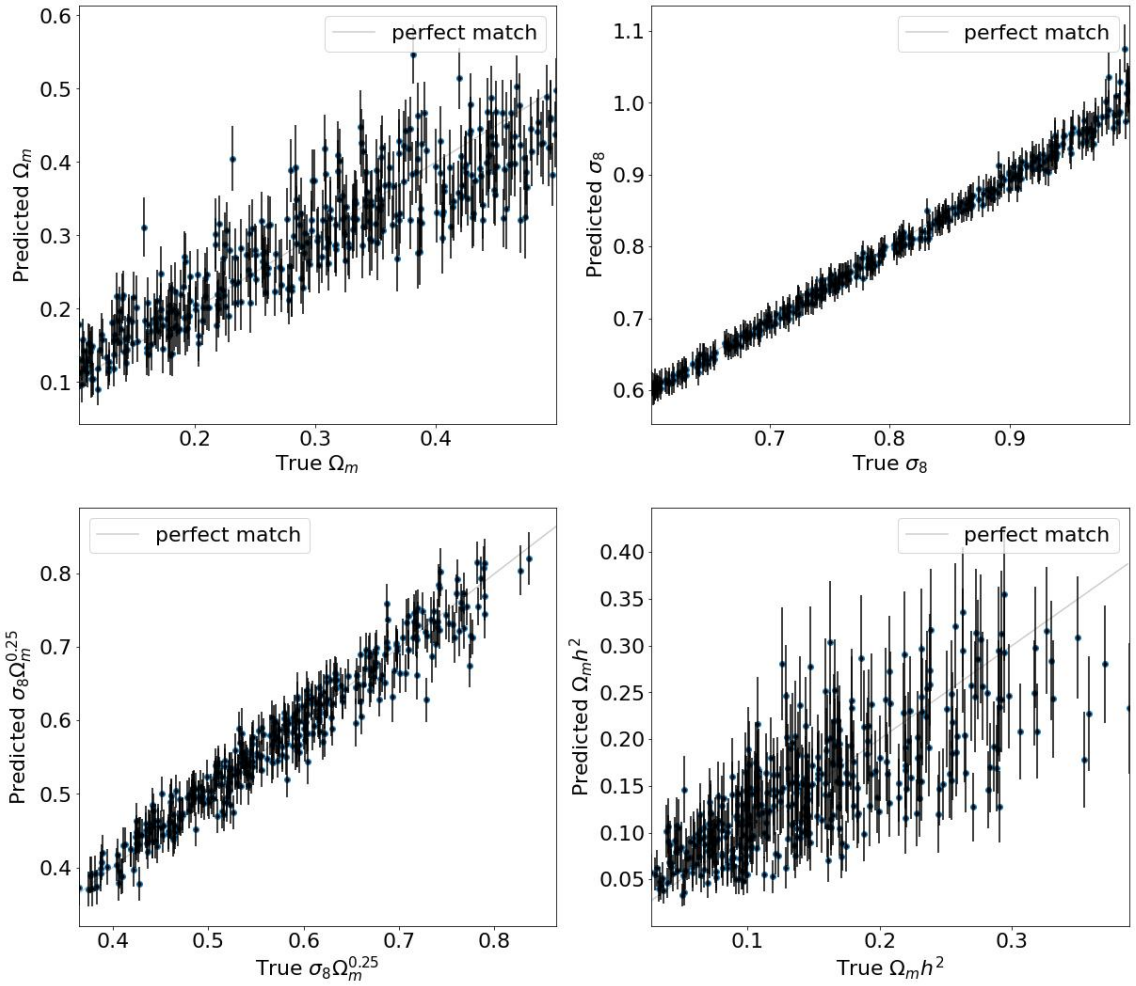


图 2. 最佳实验变分贝叶斯神经网络给出的 Ω_m, σ_8 及部分导数参数真实值与预测值的对比图。数据点表示预测分布的均值，误差棒表示在 1σ处与认知不确定性和偶然不确定性之和相关的异方差不确定性。

其他方法只有 3–4 个样本。接下来，我们采用 σ-缩放方法对 Flipout 贝叶斯神经网络的预测进行校准（31）。为此，我们仅优化式 12 中描述的损失函数，并冻结所有与贝叶斯神经网络相关的参数，即唯一可训练的参数是 s。训练后，我们得到 $s \sim 0.723$，不确定性校准误差仅降低了 10%，校准图中的样本数量增加到 4–5 个。这种微小的性能提升表明，σ-缩放并不适合校准所有贝叶斯神经网络；要构建可靠区间，必须考虑其他重新校准技术。此时，我们注意到，采用在训练步骤后便已校准良好的神经网络方法具有优势（17）。

5.2 宇宙学参数的联合分析

为了展示 N 体模拟得到的参数区间和轮廓，我们从测试集中随机选取了一个样本，其真实值见表3。宇宙学参数的二维后验分布如图 4 所示，参数的 95% 区间列于表3。我们可以观察到，与 sBNNs 相比，VBNNs 对所有参数的约束都显著更紧、更好

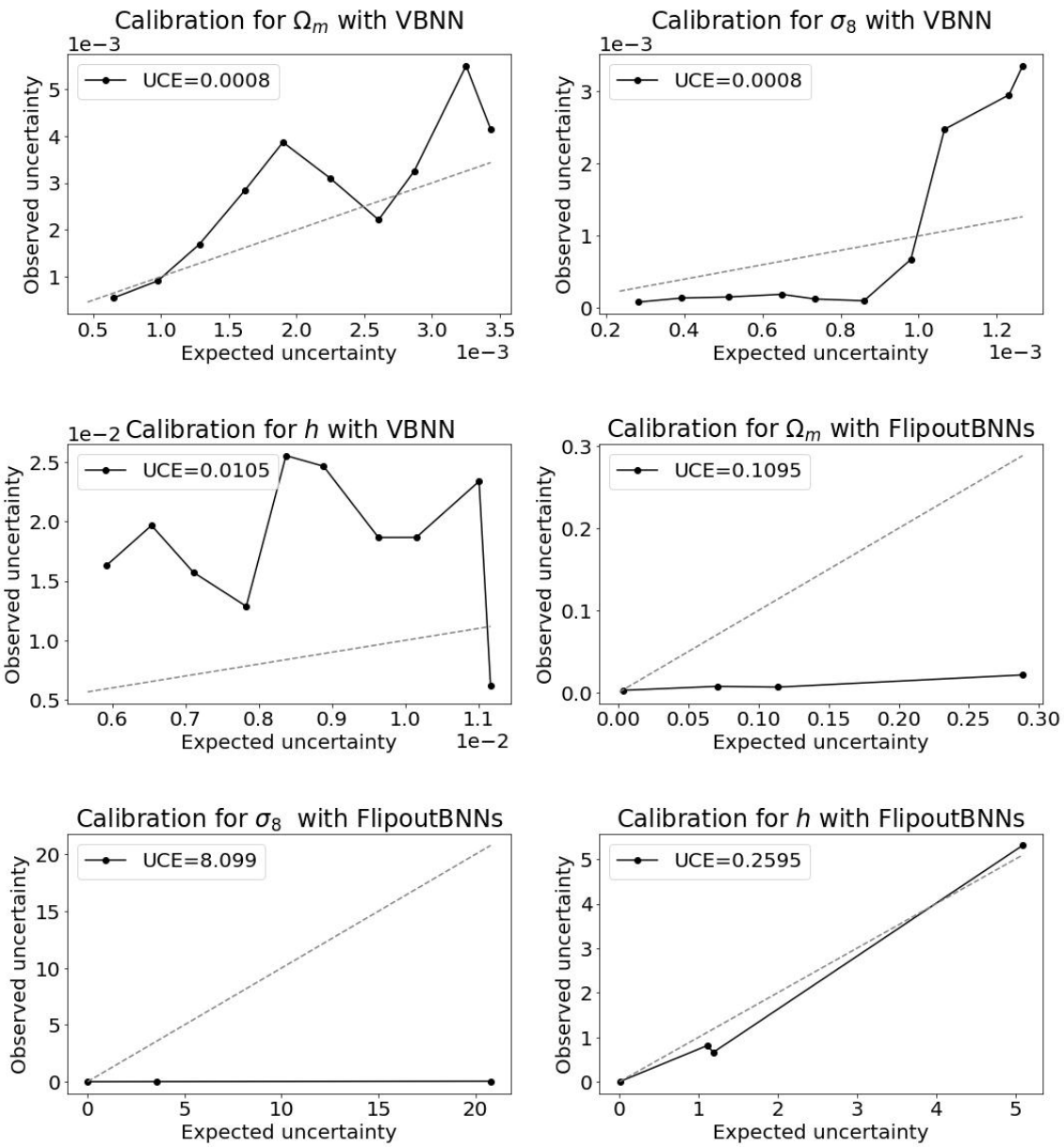


图 3. 最佳实验、变分贝叶斯神经网络和 Flipout贝叶斯神经网络的校准图。不确定性校准误差（UCE）值越低，模型的校准效果越好。虚线表示完美校准，因此与这条恒等曲线的偏差揭示了校准失准的程度。

（18）。最重要的是，这种技术还能够给出参数之间的相关性，并衡量模型预测的可靠程度。

6 结论

N 体模拟是了解宇宙初始条件、增进我们对基础物理学认识的最有力方法之一。本文使用 QUIJOTE 数据集，展示卷积深度神经网络如何在无需指定汇总统计量（如功率谱）的情况下捕捉非高斯模式。此外，我们还展示了如何构建概率深度神经网络，以获得能够反映其预测可靠性的不确定性。本文的主要目标之一是

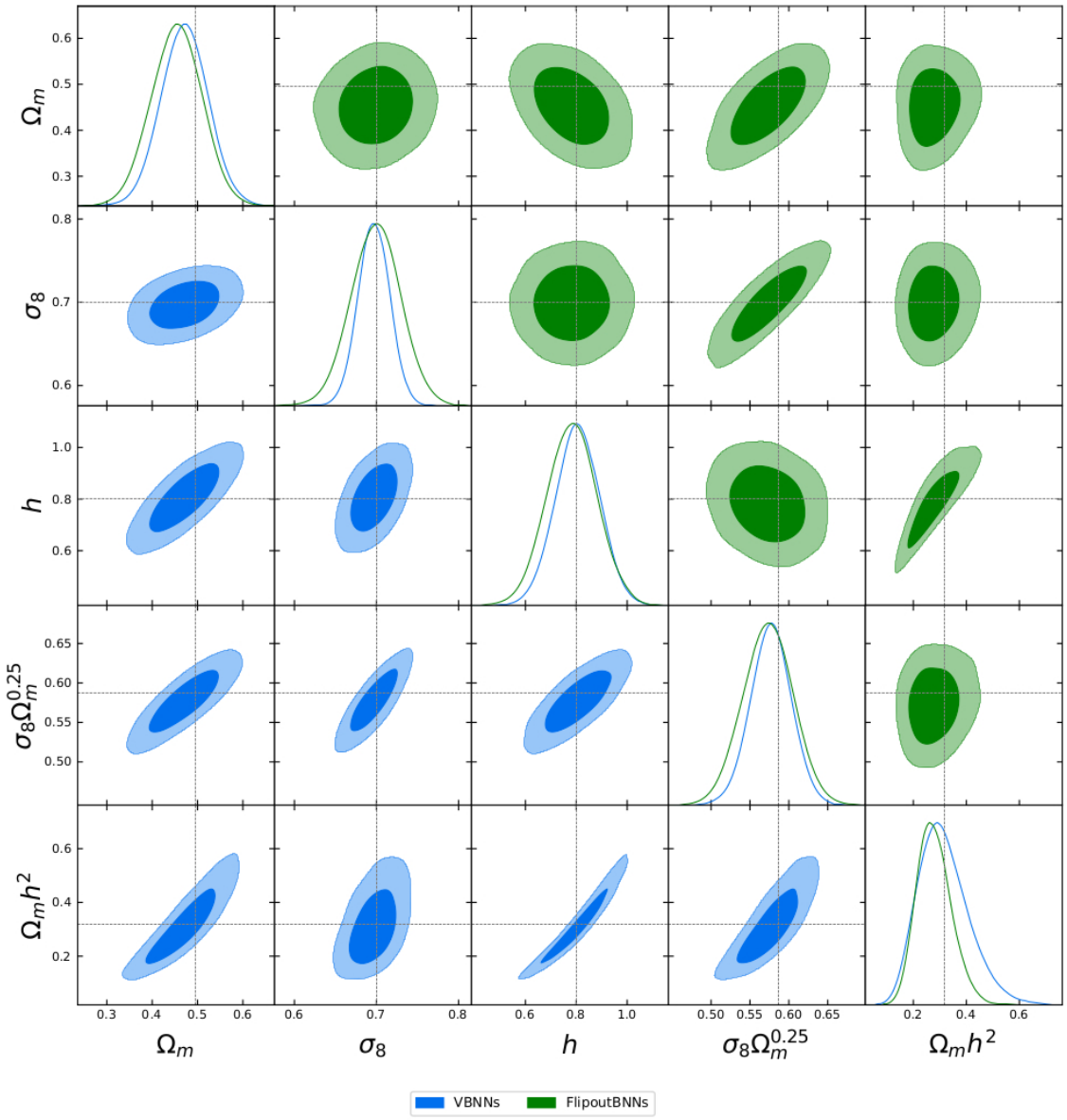


图 4. 使用变分贝叶斯神经网络和 Flipout 贝叶斯神经网络对 Quijote测试数据集中的一个样本得到的 68% 和 95% 参数约束轮廓。对角矩阵图显示边缘化后的参数约束，虚线表示真实值。此图使用 Getdist (42) 绘制。

同时报告将这些贝叶斯神经网络与乘性归一化流等技术结合，以提升变分后验复杂度后所带来的改进。我们发现，变分贝叶斯神经网络不仅能对所有宇宙学参数给出显著更严格且良好的约束（如图 4 所示），还能够提供校准良好的不确定性估计（如图 3 所示）。不过，本研究仍存在一些局限，包括简单的先验假设（平均场近似）、模拟的分辨率较低，以及缺少额外的校准技术。这些限制将在未来的论文中详细分析。

参数	95% 限值 变分贝叶斯神经网络	95% 限值Flipout贝叶斯神经网络	真实值
Ω_m	$0.47^{+0.10}_{-0.10}$	$0.45^{+0.11}_{-0.11}$	0.495
σ_8	$0.697^{+0.038}_{-0.038}$	$0.699^{+0.059}_{-0.060}$	0.699
h	$0.81^{+0.17}_{-0.17}$	$0.78^{+0.20}_{-0.19}$	0.800
$\sigma_8\Omega_m^{0.25}$	$0.577^{+0.051}_{-0.052}$	$0.573^{+0.063}_{-0.064}$	0.587
$\Omega_m h^2$	$0.31^{+0.19}_{-0.18}$	$0.573^{+0.063}_{-0.064}$	0.317

表3. 使用变分贝叶斯神经网络和Flipout贝叶斯神经网络，对Quijote测试数据集中的一个样本所得参数约束轮廓（图 4）计算出的参数 95% 区间。

致谢

本文的研究工作得到了 Google Cloud研究积分计划的支持，获奖编号为 GCP19980904。

Leonardo Casta˜neda 获得了 patrimonio aut´onomo fondo Nacional de financiamiento para la ciencia y la tecnolog´ıa y la innovacion Francisco Jos´e de Caldas（哥伦比亚科学部）项目资助，资助编号为 110685269447 RC-80740-465-2020 projects69723。H. J. Hort´ua 感谢 cr´editos educaci´on de doctorados nacionales y en el exterior- colciencias 的支持，以及 Google Cloud研究积分计划提供的资助。

参考文献

1.Stefano B, Kravtsov A. 星系团的宇宙学模拟。《*Advanced Science Letters*》**4**(2011) 204–227. doi:10.1166/asl.2011.1209.**2** .Dodelson S. 《现代宇宙学》(Academic Press, Elsevier Science) (2003).**3** .Planck C, Aghanim N, Akrami Y, Ashdown M, Aumont J, Baccigalupi C, et al. 普朗克2018年结果。《天文学与天体物理学》**641** (2020) A6. doi:10.1051/0004-6361/201833910.**4**.Tinker JL, Sheldon ES, Wechsler RH, Becker MR, Rozo E, Zu Y, et al. 基于星系聚类和星系团质量数目比的宇宙学限制。《天体物理学杂志》**745** (2011) 16. doi:10.1088/0004-637x/745/1/16.**5**.Yusofi E, Ramzanpour MA. 空洞主导宇宙学中的宇宙学常数问题与 h_0 张力（2022）。doi:10.48550/ARXIV.2204.12180.**6**.Mesinger A, Furlanetto S, Cen R. 21cmfast：一种快速的高红移21厘米信号半数值模拟。《皇家天文学会月刊》**411** (2011) 955–972. doi:10.1111/j. 1365-2966.2010.17731.x.**7**.Hamann J, Hannestad S, Lesgourgues J, Rampf C, Wong YY. 来自大尺度结构的宇宙学参数——几何信息与形状信息。《宇宙学与天体粒子物理学杂志》**2010**(2010) 022–022. doi:10.1088/1475-7516/2010/07/022.**8** .Abdalla E, Abell'an GF, Aboubrahim A, Agnello A, Akarsu O, Akrami Y, et al. 交织的宇宙学：关于与宇宙学张力和异常相关的粒子物理学、天体物理学和宇宙学的综述。《高能天体物理学杂志》(2022).**9**.Gillet N, Mesinger A, Greig B, Liu A, Ucci G. 从宇宙黎明和再电离的21厘米断层成像中进行深度学习。《皇家天文学会月刊》**484** (2019) 282–293. doi:10.1093/mnras/stz010.**10**.Dvorkin C, Mishra-Sharma S, Nord B, Villar VA, Avestruz C, Bechtol K, et al. 机器学习与宇宙学（2022）。doi:10.48550/ARXIV.2203.08056.

11 .Guo C, Pleiss G, Sun Y, Weinberger KQ. 现代神经网络的校准。第34届国际机器学习会议论文集——第70卷（JMLR.org）（2017），ICML 17，1321–1330。 12 .Chang DT. 贝叶斯神经网络：基础知识（2021）。doi:10.48550/ARXIV.2106.13594. 13 .Ravanbakhsh S, Oliva J, Fromenteau S, Price LC, Ho S, Schneider J, et al. 根据暗物质分布估计宇宙学参数（2017）。doi:10.48550/ARXIV.1711.02033. 14 .Lazanu A. 使用机器学习技术从 N 体模拟中提取宇宙学参数。《宇宙学与天体粒子物理学杂志》2021（2021）039。doi:10.1088/1475-7516/2021/09/039. 15 .Wang BY, Pisani A, Villaescusa-Navarro F, Wandelt BD. 根据空洞性质进行机器学习宇宙学（2022）。doi:10.48550/ARXIV.2212.06860. 16 .Hort´ua HJ, Volpi R, Marinelli D, Malag`o L. 使用贝叶斯神经网络估计宇宙微波背景参数。《物理评论D》102（2020）。doi:10.1103/physrevd.102.103509. 17 .Hort´ua HJ, Malag`o L, Volpi R. 使用贝叶斯归一化流限制再电离历史。《机器学习：科学与技术》1（2020）035014。doi:10.1088/2632-2153/aba6f1. 18 .Hortua HJ. 使用贝叶斯神经网络从 N 体模拟中限制宇宙学参数（2021）。doi:10.48550/ARXIV.2112.11865. 19 .Mancarella M, Kennedy J, Bose B, Lombriser L. 利用贝叶斯神经网络在宇宙学中寻找新物理：暗能量与修正引力。《物理评论D》105（2022）023531。doi:10.1103/ PhysRevD.105.023531. 20 .List F, Rodd NL, Lewis GF, Bhat I. 以全新视角看银心超额：利用贝叶斯图卷积神经网络分离伽马射线天空。《物理评论快报》125（2020）。doi:10.1103/physrevlett.125.241102. 21 .Wagner-Carena S, Park JW, Birrer S, Marshall PJ, Roodman A, Wechsler RH. 使用贝叶斯神经网络进行层次推断：强引力透镜应用。《天体物理学杂志》909（2021）187。doi:10.3847/1538-4357/abdf59. 22 . Graves A，编者。《神经网络实用变分推断》，第24卷（Curran Associates, Inc.）（2011）。 23 .Charnock T, Perreault-Levasseur L, Lanusse F. 贝叶斯神经网络（ 2020 ）。doi:10.48550/ARXIV. 2006.01490. 24 .Villaescusa-Navarro F, Hahn C, Massara E, Banerjee A, Delgado AM, Ramanah DK, et al. QUIJOTE模拟。《天体物理学杂志增刊系列》250（2020）2。doi:10.3847/1538-4365/ ab9d82. 25 .Abdar M, Pourpanah F, Hussain S, Rezazadegan D, Liu L, Ghavamzadeh M, et al. 深度学习中不确定性量化综述：技术、应用与挑战。《信息融合》76（2021）243–297。doi:10.1016/j.inffus.2021.05.008. 26 .Gal Y. 深度学习中的不确定性。博士论文，剑桥大学（2016）。 27 .Wen Y, Vicol P, Ba J, Tran D, Grosse R. Flipout：在小批次上进行高效的伪独立权重扰动（2018）。doi:10.48550/ARXIV.1803.04386. 28 .Kiureghian AD, Ditlevsen O. 偶然不确定性还是认知不确定性？这重要吗？《结构安全》31（2009）105 – 112。doi:https://doi.org/10.1016/j.strusafe.2008.06.020. 风险接受与风险沟通。 29 .Kendall A, Gal Y. 计算机视觉的贝叶斯深度学习需要哪些不确定性？（2017）。 30 .Kwon Y, Won JH, Joon Kim B, Paik M. 国际医学影像深度学习会议（2018）13。

31 .Laves MH、Ihler S、Fast JF、Kahrs LA、Ortmaier T. 利用深度学习实现医学影像中校准良好的回归不确定性。医学影像深度学习(2020).**32**.Louizos C、Welling M. 用于变分贝叶斯神经网络的乘性归一化流。第*34*届国际机器学习会议论文集——第*70*卷(JMLR.org) (2017), I CML'17, 2218–2227.**33** .Ranganath R、Tran D、Blei DM. 层次变分模型。第*33*届国际机器学习会议（国际机器学习会议）论文集——第*48*卷(JMLR.org) (2016), ICML'16, 2568–2577.**34**.Touati A、Satija H、Romoff J、Pineau J、Vincent P. 通过乘性归一化流实现随机化价值函数 (2018)。doi:10.48550/ARXIV.1806.02315.**35**.Dinh L、Sohl-Dickstein J、Bengio S. 使用 real NVP 进行密度估计。国际学习表征会议(2017).**36** .Springel V. 宇宙学模拟代码 gadget-2。《皇家天文学会月刊》**364**(2005) 1105–1134。doi:10.1111/j.1365-2966.2005.09655.x.**37**.Scoccimarro R. 初始条件引发的瞬态：微扰分析。《皇家天文学会月刊》**299** (1998) 1097–1118。doi:10.1046/j.1365-8711.1998.01845.x.**38** . Abadi M、Agarwal A、Barham P、Brevdo E、Chen Z、Citro C 等。TensorFlow：异构系统上的大规模机器学习 (2015)。软件可从 tensorflow.org 获取。**39**.Solovyev R、Kalinin AA、Gabruseva T. 用于脑毛细血管停滞检测的三维卷积神经网络。《生物学与医学中的计算机》**141** (2022) 105089。doi:10.1016/j.compbiomed.2021. 105089.**40** .Kingma DP、Ba J. Adam：一种随机优化方法 (2014)。doi:10.48550/ARXIV.1412. 6980.**41**.Sønderby CK、Raiko T、Maaløe L、Sønderby SK、Winther O. 阶梯式变分自编码器。第*30*届神经信息处理系统国际会议论文集(美国纽约州雷德胡克：Curran Associates 公司) (2016), NIPS'16, 3745–3753.**42** . Lewis A. GetDist：用于分析蒙特卡洛样本的 Python 软件包 (2019)。