arXiv:2303.06422v2 [stat.CO] 5 Jul 2023

一种用于多保真度分布估计的近似控制变量方法*

Ruijian Han†、Boris Kramer‡、Dongjin Lee‡、Akil Narayan§、和 Yiming Xu¶

摘要。 研究感兴趣量（QoI）分布的基于正向模拟的不确定性量化，是计算稳健工程设计与预测的重要组成部分。大量文献致力于准确评估感兴趣量的统计量；特别是，多层次或多保真度方法能够利用给定模型集合之间的成本—精度权衡，因此被证明是有效的。然而，能够估计感兴趣量完整分布的有效算法仍在积极发展中。本文提出了一个通用的多保真度框架，用于在预算约束下估计与高保真模型相关的向量值感兴趣量的累积分布函数（CDF）。给定由低保真度替代模型获得的一组适当控制变量，我们的框架首先识别最具成本效益的模型子集，然后利用该子集构建目标累积分布函数的近似控制变量估计量。我们通过使用中间线性近似器构造一组控制变量来实例化该框架，并对相应算法进行严格分析。分析表明，随着预算趋于无穷，所得累积分布函数估计量是一致相合且渐近最优的；这只需对感兴趣量联合分布作出较弱的矩和正则性假设。该方法提供了一种稳健的多保真度累积分布函数估计量，能够根据可用预算进行自适应，不要求 *a priori* 了解跨模型统计量或模型层次结构，并且适用于多个维度。我们通过参数偏微分方程和随机微分方程的测试示例展示了该方法的效率与稳健性，其中包括学术实例以及更具挑战性的工程问题。

关键词。 控制变量、分布估计、模型选择、多保真性、稳健性

AMS主题分类。 62J05, 62G30,62F12, 62-08

1. 引言。 物理系统通常通过计算机模拟或仿真器来建模，因此，理解这些构建出的近似中的误差至关重要。输出中的一个特定不确定性来源是这些模型中的输入不确定性，这种不确定性可能来自模型参数（其维度可以是有限维或无限维），也可能来自系统中建模的随机性，例如由白噪声过程驱动的系统。为了使最终模型值得信赖，量化感兴趣量（QoI）中由此产生的不确定性至关重要；也就是说，需要估计感兴趣量的分布或其某种统计摘要。实现这一目标的一种常用方法是蒙特卡罗方法（MC）模拟，它易于实现且结果稳健，但收敛速

*已提交给编辑。 **资助：** RH部分获得香港研究资助局（资助编号：14301821）和香港理工大学新聘人员启动基金的资助。AN部分获得美国国家科学基金会资助（DMS-1848508）以及美国空军科学研究办公室资助（FA9550-20-1-0338）。BK和DL通过国际合作研发项目（资助编号：P0019804，基于数字孪生的智能无人设施检测解决方案）获得韩国产业通商资源部和韩国技术振兴院的资助。本文所表达的观点不一定代表富国银行的观点。

†应用数学系，香港理工大学
‡加州大学圣迭戈分校机械与航空航天工程系§科学计算与成像研究所，以及犹他大学数学系

¶企业模型风险，富国银行（yiming2358@gmail.com）。

度较慢。典型的蒙特卡洛程序需要抽取大量样本或进行重复实验；随着计算机模拟日益复杂，这一过程成本高昂。

为解决这一问题，研究人员开发了基于多层次 [11, 12]和多保真度建模 [29, 28, 30, 19, 27, 16, 18, 32, 9, 35, 33, 17, 8, 10] 的方法，用于估计与（高保真）模型相关的感兴趣量的统计量。多层次／多保真方法的核心思想是利用精度和成本各不相同的模型来提高计算效率。然而，现有文献的一大局限是，主要关注感兴趣量统计均值（或分位数、条件期望等其他标量描述性统计量）的估计，因此只能部分揭示感兴趣量的不确定性。若要更全面地理解感兴趣量的不确定性，还需评估其高阶统计量，甚至完整分布。

在多层次和多保真设置下估计累积分布函数的方法已经取得了显著成功 [12, 23, 13, 21, 3, 36]。在 [12], 中，作者提出了一种计算由随机微分方程产生的一元随机变量的累积分布函数的多层次方法，并推导出了关于误差的成本上界。[12] 中的方法论随后在多项工作中得到进一步发展和应用 [23, 21, 13, 3]。具体而言，[23]设计了一种事后优化策略，用于校准平滑函数，并在油藏模拟中展示了其优于蒙特卡罗方法的性能； [21] 将 [12]中的思想推广到了更一般的参数化期望的近似，例如特征函数； [13]采用了一种用于参数选择的自适应方法，从而得到更优的成本界；[3] 提供了一种新颖且可计算的误差估计器，以改进算法调优。尽管这些方法作出了实质性贡献，但其中几乎所有方法都对模型层次结构作出了相对严格的假设（例如模型成本与准确度之间的权衡），因而无法直接扩展到一般的非层次化多保真度设置。对于这种更一般的累积分布函数多保真估计，据我们所知，唯一相关工作是用于分布学习的自适应探索后承诺算法（AETC-d） [36]。然而，AETC-d 在大预算下的性能受到其自身一组统计假设的限制，而这些假设在实践中往往过于严格，难以满足。此外，上述所有参考文献中的感兴趣量都被假设为标量。

本文的结构概述如下。本节其余部分列出我们的贡献，介绍总体记号，并总结主要理论与算法进展。第 2 节至第 4 节介绍我们的方法所需的数学与统计背景：第 2节简要概述控制变量法；第 3节介绍一种基于近似控制变量估计量的多保真累积分布函数估计框架。第 4 节通过线性近似量构造控制变量的计算方法。第 5 节构建我们新的元算法（cvMDL），该算法能够自主进行模型选择，并通过算法修正保持所得累积分布函数估计量的单调性。元算法 cvMDL 本身并未规定如何计算控制变量：采用第 4节中的线性近似进行专门化后，可得到一个计算上明确的算法；我们对其进行了详细研究，建立了一致一致性和预算渐近最优性。第 6 节包含详细的模拟研究，并展示了利用估计的累积分布函数计算概率风险度量的应用。

1.1. 贡献。 本文的主要目标是为缓解上述缺陷提供新颖的解决方案。我们在一般的非层次化多保真度近似设置下，针对计算预算约束，开发了一种用于估计累积分布函数的高效算法。所提出的方法满足以下标准：1）其输入既不需要跨模型统计量，也不需要模型层次结构；2）能够为向量值感兴趣量提供分布估计；3）在经验上具有稳健性，并享有理论保证。尽管我们的方法采用了与 [35, 36]（均借鉴了多臂老虎机学习中的先探索后承诺算法 [22]），但其中包含大量新要素，扩展了适用范围并提高了稳健性。更技术性地说，我们的贡献包括两个方面：

- 我们提出了一种基于控制变量的探索-利用策略，用于在预算约束下进行多保真累积分布函数估计。探索 阶段利用统计估计，从低保真度模型中选出一部分用于构造控制变量；随后，利用 阶段运用已学习的信息，为目标累积分布函数构造近似控制变量估计量。该过程在初始化时不需要任何关于模型间关系的先验 神谕信息[1]，这与若干需要将此类信息作为输入的方法不同。此外，我们提出的累积分布函数估计量既适用于标量值 QoI，也适用于向量值感兴趣量，这使其区别于仅适用于标量值 QoI 的现有方法。

- 通过考察平衡探索与利用中误差的加权 L^2 损失，我们设计了一种新的元算法，即控制变量多保真度分布学习算法（“cvMDL”，概述见 图1，详述见 算法 5.2），用于模型（子集）选择和累积分布函数估计。借助由线性近似器构造的控制变量，我们证明，随着预算趋于无穷，cvMDL 所产生估计量具有一致性和渐近最优性（定理 5.7）。我们的分析表明，所提出的过程显著缓解了 [36] 中严格的模型假设。

直接使用我们的方法所得到的累积分布函数估计量具备此前提及的理论保证，但不一定具有单调性，因此其自身可能并不是分布函数。为消除这一问题，我们采用经验算法修正，使估计的累积分布函数恢复单调性，同时也使其处理在计算上更加便利（例如提取分位数和条件期望）；参见 算法 5.1。我们观察到，在某些情况下，这种经验修正还能进一步减少误差。

1.2. 记号。 对于 $n \in \mathbb{N}$，令 $\{1\colon n\} := \{1, \dots, n\}$。我们分别用粗体大写字母和小写字母表示矩阵和向量。向量$\boldsymbol{v}$ 上的欧几里得（ℓ^2）范数记为 $\|\boldsymbol{v}\|_2$。对于矩阵 $\boldsymbol{A}$，$\boldsymbol{A}^\top$ 表示转置，$\boldsymbol{A}^\dagger$ 表示伪逆；当 $\boldsymbol{A}$ 可逆时，$\boldsymbol{A}^\dagger$与常规逆 $\boldsymbol{A}^{-1}$ 一致。$\boldsymbol{A}$ 的第 i列记为 $\boldsymbol{A}^{(i)}$。$\boldsymbol{A}$ 的弗罗贝尼乌斯范数记为 $\|\boldsymbol{A}\|_F = (\sum_i \|\boldsymbol{A}^{(i)}\|_2^2)^{1/2}$。我们用 $\otimes$ 表示张量积算子。对于集合 $\mathcal{T} \subseteq \mathbb{R}^d$，其内部记为 $\mathcal{T}^\circ$，并且

[1]在本文中，Oracle 信息是指我们视为精确的模型统计量。这些统计量可能可以精确计算，但更常见的是，它们是通过模拟确定的近似值；由于计算成本足够高，这些近似值会被视为真实值。

$\mathbf{1}_{\mathcal{T}}(\boldsymbol{x}) := \mathbf{1}_{\{\boldsymbol{x}\in\mathcal{T}\}}$ 表示集合 $\mathcal{T}$ 上的示性函数。对于两个向量 $\boldsymbol{x} = (\boldsymbol{x}^{(1)}、\ldots、\boldsymbol{x}^{(d)})^\top$ 和 $\boldsymbol{y} = (\boldsymbol{y}^{(1)}、\ldots、\boldsymbol{y}^{(d)})^\top$，我们用 $\vee$ 和 $\wedge$ 分别表示逐分量最大值 和最小值 算子，即:

$$\boldsymbol{x} \vee \boldsymbol{y} := \left(\max\{\boldsymbol{x}^{(1)}, \boldsymbol{y}^{(1)}\}, \ldots, \max\{\boldsymbol{x}^{(d)}, \boldsymbol{y}^{(d)}\}\right)^\top$$
$$\boldsymbol{x} \wedge \boldsymbol{y} := \left(\min\{\boldsymbol{x}^{(1)}, \boldsymbol{y}^{(1)}\}, \ldots, \min\{\boldsymbol{x}^{(d)}, \boldsymbol{y}^{(d)}\}\right)^\top.$$

此外，我们称 $\boldsymbol{x} \le \boldsymbol{y}$ 当且仅当 $\boldsymbol{x}^{(i)} \le \boldsymbol{y}^{(i)}$ 对所有 $i \in \{1$ 均有: $d\}$。我们将计算模型中的感兴趣量视为随机变量，它们共同存在于某个公共概率空间 $(\Omega\ \mathcal{F}, \mathbb{P})$ 中。对于随机向量 $X \in \mathbb{R}^d$，我们令 $F_X(\boldsymbol{x}) = \mathbb{P}(X^{(1)} \le \boldsymbol{x}^{(1)}, \ldots, X^{(d)} \le \boldsymbol{x}^{(d)})$ 表示其累积分布函数。对于两列随机变量 $\{a_m(\omega)\}$ 和 $\{b_m(\omega)\}$，其中 $\omega \in \Omega$ 是一个概率事件，如果几乎处处（a.s.）有 $a_m(\omega) \lesssim b_m(\omega)$，即对所有 $m \in \mathbb{N}$，$a_m(\omega) \le \eta(\omega) b_m(\omega)$，则我们记作该关系，其中常数 $\eta(\omega)$ 与 m 无关。为方便起见，我们令

$$Y = (Y^{(1)}, \ldots, Y^{(d)})^\top \in \mathbb{R}^d \qquad X_i = (X_i^{(1)}, \ldots, X_i^{(d_i)})^\top \in \mathbb{R}^{d_i} \qquad i \in \{1:n\}$$

分别表示高保真度和第 i 个低保真感兴趣量。这里，d，$d_i \in \mathbb{N}$ 分别是 Y 和 X_i的对应维度。总共有 n 个低保真模型。我们使用$\mathbb{E}[\cdot]$，$\mathrm{Var}[\cdot]$/$\mathrm{Cov}[\cdot]$和 $\mathrm{Corr}[\cdot]$ 分别表示期望、方差/协方差以及相关性算子。我们使用 $\perp\!\!\!\perp$ 表示概率独立性。

1.3. 模型假设。我们假设，对 Y和 X_1，……，X_n进行采样的成本分别由正数 c_0和 c_1，……，c_n表示，这些成本是确定且已知的。对于 $\mathcal{S} \subseteq \{1:\ n\}$，令$c_{\mathcal{S}} = \sum_{i\in\mathcal{S}} c_i$表示从子集$\mathcal{S}$中的所有（低保真）模型进行采样的成本。我们令 $B > 0$ 表示可用于模型采样的总预算（确定且已知）。此外，对于 $\mathcal{S} \subseteq \{1:\ n\}$，令 $X_{\mathcal{S}} = (X_i^\top)_{i\in\mathcal{S}}^\top \in \mathbb{R}^{d_{\mathcal{S}}}$，其中 $d_{\mathcal{S}} = \sum_{i\in\mathcal{S}} d_i$; 并令 $X_{\mathcal{S}^+} = (1, X_{\mathcal{S}}^\top)^\top$，后者用于考虑带截距/偏置项的线性模型近似。

本文其余部分的核心目标是开发一个多保真估计量，用于 $F_Y(\boldsymbol{x})$，方法是对（Y、$X_{\{1:n\}}$）以及 $X_{\mathcal{S}}$ 进行采样，其中某些 $\mathcal{S} \subseteq \{1$ 由自适应确定 $n\}$，并且采样成本不超过总预算约束 $B > 0$。我们不作其他高层次假设。换言之，我们仅假设 Y 是一个已知的高保真模型；我们不假设模型 $X_{\{1:n\}}$之间存在任何排序或层级关系，也不假设已知任意模型之间的统计量（例如相关性）。尽管这种一般性足以满足算法目的，但我们的理论保证还需要额外的技术假设，具体见第5.4小节。这些技术假设是温和的正则性条件，涉及随机变量的有限阶矩和累积分布函数泛函正则性。

我们引入的记号足以在下一节介绍完整的cvMDL算法。然而，要使该算法具备实用性，实际计算还需要更多技术细节，这些细节将在第2节至第5节中给出。

1.4. 算法总结。所提出的 cvMDL 元算法如图 1所示。总而言之，我们首先通过 m 收集（Y，$X_{\{1:n\}}$）的完整联合样本，经历一个探索阶段，以确定: （i）模型之间的关联方式；（ii）哪个模型子集 $\mathcal{S}$ 能够在成本与准确度之间实现最优平衡

，以及（iii）是否需要更多样本 m 来验证稳健的探索，或者 $\mathcal{S}$ 的选择在统计上是否足够稳健，可以继续进行利用。探索阶段之后是利用阶段，在此阶段我们用尽剩余的计算预算，对最优模型子集 $X_{\mathcal{S}}$进行采样。利用对应于针对 F_Y执行特定的近似控制变量估计量。具体描述如下：

探索阶段

– 最小探索：此步骤确保探索样本数 m 足够大，从而能够计算非退化的经验统计量。

– 分析低保真模型：我们的目标是估计与利用模型子集 $\mathcal{S}$的估计的累积分布函数相关联的最小损失。为此，此步骤为每个模型子集 $\mathcal{S}$ 确定估计的最优探索样本数 $\widehat{m}^*_{\mathcal{S}}$ 以及相应的损失函数最小值 $L(m \vee \widehat{m}^*_{\mathcal{S}}; m)$。值 $L(z; m)$ 是基于当前可用的 m 探索样本得到的估计量，用于衡量最终使用 z 探索样本时的估计损失。在评估最小损失时，我们要求输入 $z \leftarrow m \vee \widehat{m}^*_{\mathcal{S}}$，因为如果 $m > \widehat{m}^*_{\mathcal{S}}$，则探索样本数应为 m，而不是 $\widehat{m}^*_{\mathcal{S}}$（我们不能采用少于已经承诺的探索样本数，并且假设 $L(z; m)$ 是凸函数且具有唯一极小值点）。L和 $\widehat{m}^*_{\mathcal{S}}$的定义见（5.6）和（5.7）。 $\mathcal{S}^*$ – 选择最优模型：通过从上一步中选择损失最小的子集 $\mathcal{S}$，计算得到估计最优模型子集 $\mathcal{S}^*$。– 继续探索：如果当前探索样本数 m 小于最优子集 $\mathcal{S}^*$所需的估计最优样本数 $\widehat{m}^*_{\widehat{\mathcal{S}}^*}$，则继续探索；所需增加的确切探索样本数由函数 $Q(\cdot, \cdot)$ 确定，该函数定义于（5.12）中。如果 $m \geq \widehat{m}^*_{\widehat{\mathcal{S}}^*}$，则探索终止，并进入利用阶段。

利用阶段

– 消耗预算：探索结束并确定了一个“最优”模型子集后， $\mathcal{S}^*$ 我们将剩余的计算预算用于采样 $X_{\widehat{\mathcal{S}}^*}$。– 估计累积分布函数：利用收集到的样本，我们构造累积分布函数估计量 $F_{\widehat{\mathcal{S}}^*}$，用于估计 F_Y，其定义见（5.10）。

损失函数的具体计算方法以及累积分布函数估计量的构造方式，将在第5节中介绍；第2节至第4节则用于给出必要的数学和统计定义。

更详细的算法版本见算法 5.2，其中列出了必须执行的、更明确的计算步骤。接下来的章节将阐述图1中各量的理论构造；具体而言，第2节和第3节将给出损失函数的构造，而损失函数是探索决策的核心组成部分。

2. 背景：控制变量。我们首先介绍控制变量法，这是蒙特卡罗模拟中用于方差缩减的标准方法。对于具有有界方差 $\sigma_X^2 > 0$ 的随机变量 X，基于独立同分布数据 X_ℓ、 $\widehat{x} = \sum_{\ell \in \{1:\ m\}} X_\ell / m$ 的大小为 m 的蒙特卡罗估计量 $\mathbb{E}[X]$ 是无偏的，其方差为 σ_X^2/m。给定一个与 $\mathbb{E}[X]$ 位于同一概率空间中的随机向量 $Z =$

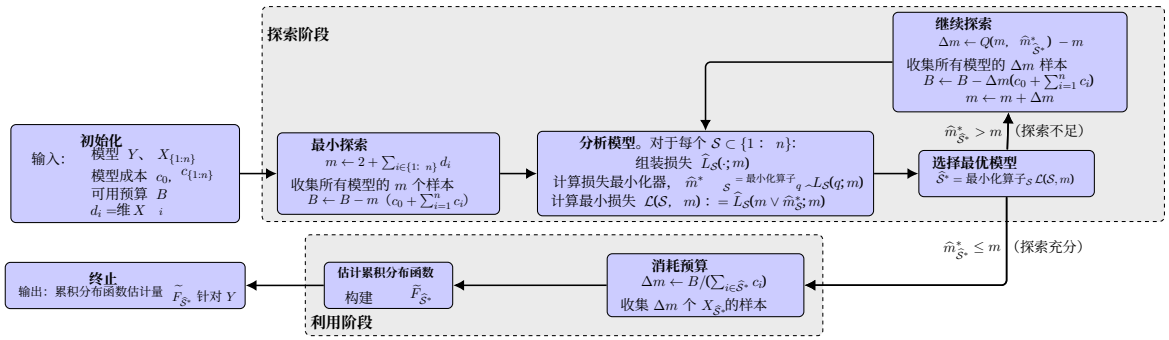


图1：cvMDL算法的流程图示意。有关各步骤的更多细节，参见第1.4小节。完整算法见算法 5.2。

$(Z^{(1)}, \ldots, Z^{(d)})^\top \in \mathbb{R}^d$，可以使用 (X, Z) 的联合独立同分布样本，即 $(X_\ell, Z_\ell^\top) = (X_\ell, Z_\ell^{(1)}, \ldots, Z_\ell^{(d)})$，其中 $\ell \in \{1:m\}$，以构造控制变量估计量 $\widehat{x}_{\mathrm{cv}}$，用于 $\mathbb{E}[X]$:

$$\widehat{x}_{\mathrm{cv}} = \frac{1}{m}\sum_{\ell\in\{1:m\}} X_\ell - \frac{1}{m}\sum_{\ell\in\{1:m\}} (Z_\ell^\top\beta - \mathbb{E}[Z]^\top\beta),$$

其中，$\beta \in \mathbb{R}^d$ 是适当选择的向量，$\mathbb{E}[Z]$ 假定为已知。估计量$\widehat{x}_{\mathrm{cv}}$ 同样是无偏的，其方差为

$$\sigma_{\mathrm{cv}}^2 = \mathrm{Var}[\widehat{x}_{\mathrm{cv}}] = \frac{\mathrm{Var}[X - Z^\top\beta]}{m}.$$

当 β 为中心化线性回归的最小二乘系数时，即将$(X - \mathbb{E}[X])$对$(Z - \mathbb{E}[Z])$进行回归时，该方差达到最小，

(2.1)

$$\beta = \mathrm{Cov}[Z]^{-1}\mathrm{Cov}[Z, X] \implies \sigma_*^2 = \min_{\beta\in\mathbb{R}^d}\sigma_{\mathrm{cv}}^2 = \frac{(1-\rho^2)\sigma_X^2}{m}, \quad \rho = \mathrm{Corr}(X, Z^\top\mathrm{Cov}[Z]^{-1}\mathrm{Cov}[Z, X]).$$

当 $|\rho| \approx 1$时，方差缩减十分显著，此时 $Z^\top\beta$ 解释了$\mathrm{Var}[X]$的大部分。

当 $\mathbb{E}[Z]$ 未知时，可以考虑以下近似控制变量估计量$\widetilde{}$，该估计量使用独立的大小为N 的蒙特卡罗方法估计量来代替 $\mathbb{E}[Z]$，并使用样本 Z_j:

(2.2)

$$\widehat{x}_{\mathrm{acv}} = \frac{1}{m}\sum_{\ell\in\{1:m\}} X_\ell - \frac{1}{m}\sum_{\ell\in\{1:m\}}\left(Z_\ell^\top\beta - \frac{1}{N}\sum_{j\in\{1:N\}}\widetilde{Z}_j^\top\beta\right) \quad (\ell, j) \in \{1:m\}\times\{1:N\},$$

$\widetilde{}$其中我们假设 $Z_\ell \perp\!\!\!\perp Z_j$。那么其方差为

(2.3)
$$\sigma_\star^2 = \sigma_*^2 + \frac{\mathrm{Var}[Z^\top\beta]}{N} = \frac{(1-\rho^2)\sigma_X^2}{m} + \frac{\rho^2\sigma_X^2}{N}.$$

此类近似控制变量估计量的构造，近期已在一阶统计量的多保真估计中得到研究 [16, 35]。$\mathsf{Cov}[Z]^{-1}$ 和 $\mathsf{Cov}[Z, X]$这两个项可以通过经验方式估计，但代价是在 m [14, 26] 中引入更高阶的轨迹级统计误差。

3. 用于累积分布函数估计的方差缩减。 控制变量还可以更一般地应用于随机变量的非线性函数的累积分布函数估计 [15, 20]。例如，在风险管理应用中 [15], , 作者考虑使用 *delta-gamma* 近似[2] （即损失函数的二阶泰勒展开） L 在给定头寸 $\boldsymbol{x}$ 沿随机市场变动方向 $\boldsymbol{\eta}$ 作为控制变量来计算其分位数。更准确地说，使用关于 $\boldsymbol{\eta}$ 的二次函数来近似在 $\boldsymbol{x}$处的损失：

$$-(L(\boldsymbol{x}+\boldsymbol{\eta})-L(\boldsymbol{x}))=:\ell(\boldsymbol{x})\approx\widehat{\ell}(\boldsymbol{x}):=-\nabla L(\boldsymbol{x})^{\top}\boldsymbol{\eta}-\frac{1}{2}\boldsymbol{\eta}^{\top}\nabla^{2}L(\boldsymbol{x})\boldsymbol{\eta}.$$

固定一个标量 C， $\mathbf{1}_{\{\widehat{\ell}(\boldsymbol{x})\leq C\}}$ 可用作 $\mathbf{1}_{\{\ell(\boldsymbol{x})\leq C\}}$ 的控制变量，以计算后者的期望；这尤其提供了一种计算累积分布函数的方法。 [20] 中介绍了更先进的近似技术，用于构造风险价值计算中的其他控制变量。

我们在此处提出的多保真设置中应用了类似的思想。在我们的设置中，生成特定的函数形式可能在计算上较为困难，而类泰勒近似在局部区域之外可能不够准确。我们的替代策略是利用一个基于 $X_{\{1:n\}}$线性组合的 Y 全局仿真器，当这些量之间的相关性较高时，这种策略可能十分有效。例如，在对参数偏微分方程进行建模时，通常就是这种情况。在本节的其余部分，我们将介绍一种一般的多保真方法，用于在预算约束下估计 $F_Y(x)$。

3.1. 用于多保真累积分布函数估计的控制变量。 在构建所提出的方法时，我们经常利用以下简单的观察结果：

$$F_Y(\boldsymbol{x})=\mathbb{E}[\mathbf{1}_{\{Y\leq\boldsymbol{x}\}}],\qquad\boldsymbol{x}\in\mathbb{R}^{d}.$$

如果固定 $\mathcal{S}\subseteq\{1:\ n\}$，则基于 $X_{\mathcal{S}}$ 且能使方差最小（因而是最优）的控制变量为 $\mathbb{E}[\mathbf{1}_{\{Y\leq\boldsymbol{x}\}}|X_{\mathcal{S}}]$ [31]。该量需要将 Y正交投影到由 $X_{\mathcal{S}}$生成的σ-代数上；若无特殊假设（例如联合正态性），这一过程在计算上难以处理。为了近似 $\mathbb{E}[\mathbf{1}_{\{Y\leq\boldsymbol{x}\}}|X_{\mathcal{S}}]$，我们使用 $h(X_{\mathcal{S}};\ \boldsymbol{x})$来表示一个一般的 $X_{\mathcal{S}}$-可测函数，该函数充当 $\mathbf{1}_{\{Y\leq\boldsymbol{x}\}}$的控制变量。我们在 h 中对第4节作出具体选择。

类似于（2.2），我们为 F_Y（$\boldsymbol{x}$）构造一个近似控制变量估计量，其中 m 和 N 在（2.2）中通过预算约束相关联（采样Y 和 $X_{\mathcal{S}}$的成本）。由于同时考虑不同的模型子集 $\mathcal{S}$ ，我们采用均匀探索策略，首先收集 m 个关于完整模型的独立同分布联合探索样本以降低方差，然后将剩余预算用于收集 $N_{\mathcal{S}}$ 个关于

[2]这里所说的是“完整”的Delta-Gamma近似。实践中更常用的Delta-Gamma近似并不考虑二阶交叉项。

选定的低保真度模型子集 $\mathcal{S}$，以计算控制变量均值。在均匀探索策略下，探索样本和利用样本表示为：

$$\text{Exploration samples: } \{(X_{\text{epr},\ell,1}^\top, \dots, X_{\text{epr},\ell,n}^\top, Y_{\text{epr},\ell}^\top)^\top\}_{\ell\in\{1:m\}} \subset \mathbb{R}^{d+\sum_{i=1}^n d_i} \tag{3.1}$$

$$\text{Exploitation samples: } \{X_{\text{ept},j,\mathcal{S}}\}_{j\in\{1:N_\mathcal{S}\}}, \tag{3.2}$$

其中，下标“epr”和“ept”分别表示样本使用的阶段。参数m 和 $N_\mathcal{S}$ 通过预算约束相关联：

$$N_\mathcal{S} = \frac{B - c_{\text{epr}} m}{c_\mathcal{S}} \qquad c_{\text{epr}} = \sum_{i=0}^{n} c_i, \qquad c_\mathcal{S} = \sum_{i\in\mathcal{S}} c_i, \tag{3.3}$$

其中，为简化讨论，我们忽略整数舍入效应。基于 $h(X_\mathcal{S}; \boldsymbol{x})$ 的 $F_Y(\boldsymbol{x})$ 的控制变量估计量为

$$\widehat{F}_\mathcal{S}(\boldsymbol{x}) = \frac{1}{m}\sum_{\ell\in\{1:m\}} \mathbf{1}_{\{Y_{\text{epr},\ell}\le\boldsymbol{x}\}} - \frac{1}{m}\sum_{\ell\in\{1:m\}} \alpha(\boldsymbol{x}) \left(h(X_{\text{epr},\ell,\mathcal{S}};\boldsymbol{x}) - \frac{1}{N_\mathcal{S}}\sum_{j\in\{1:N_\mathcal{S}\}} h(X_{\text{ept},j,\mathcal{S}};\boldsymbol{x}) \right), \tag{3.4}$$

其中 $\alpha(\boldsymbol{x})$ 是（2.1）中所述的最优缩放系数：

$$\alpha(\boldsymbol{x}) = \text{Cov}[h(X_\mathcal{S};\boldsymbol{x})]^{-1}\text{Cov}[\mathbf{1}_{\{Y\le\boldsymbol{x}\}}, h(X_\mathcal{S};\boldsymbol{x})]. \tag{3.5}$$

注意，如果 $\text{Cov}[h(X_\mathcal{S}; \boldsymbol{x})] = 0$，则 $\alpha(\boldsymbol{x})$ 未定义。在这种情况下，估计量 $F_\mathcal{S}(\boldsymbol{x})$ 的值不依赖于 $\alpha(\boldsymbol{x})$，为方便起见，我们将 $\alpha(\boldsymbol{x})$ 设为 0。量 $F_\mathcal{S}(\boldsymbol{x})$ 是 $F_Y(\boldsymbol{x})$ 的无偏估计量，其方差为

$$\text{Var}[\widehat{F}_\mathcal{S}(\boldsymbol{x})] = \frac{(1-\rho_\mathcal{S}^2(\boldsymbol{x}))F_Y(\boldsymbol{x})(1-F_Y(\boldsymbol{x}))}{m} + \frac{\rho_\mathcal{S}^2(\boldsymbol{x})F_Y(\boldsymbol{x})(1-F_Y(\boldsymbol{x}))}{N_\mathcal{S}},$$

其中

$$\rho_\mathcal{S}(\boldsymbol{x}) = \text{Corr}[\mathbf{1}_{\{Y\le\boldsymbol{x}\}}, h(X_\mathcal{S};\boldsymbol{x})]. \tag{3.6}$$

3.2. 控制变量损失函数。为了测量 $F_\mathcal{S}$（$\boldsymbol{x}$）的整体准确度，我们引入损失 $L_\mathcal{S}$，其定义为 ω（$\boldsymbol{x}$）加权的 L^2-范数平方的平均值，其中 $\widehat{F}_\mathcal{S}$（$\boldsymbol{x}$）$-F_Y$（$\boldsymbol{x}$）：

$$L_\mathcal{S} \coloneqq \mathbb{E}\left[\int_{\mathbb{R}^d} \omega(\boldsymbol{x})|\widehat{F}_\mathcal{S}(\boldsymbol{x}) - F_Y(\boldsymbol{x})|^2 \mathrm{d}\boldsymbol{x}\right], \tag{3.7}$$

其中，$\omega(\boldsymbol{x})$：$\mathbb{R}^d \to \mathbb{R}_{\ge 0}$ 是一个权重函数。$\omega(\boldsymbol{x})$-加权的 L^2-范数平方与分布上的其他更广泛使用的度量有关，例如，当 $\omega(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = \mathrm{d}F_Y(\boldsymbol{x})$时，它可化简为Cramér–von Mises距离。为估计 $L_\mathcal{S}$，注意根据Tonelli定理，我们有：

$$L_\mathcal{S} = \int_{\mathbb{R}^d} \omega(\boldsymbol{x})\text{Var}[\widehat{F}_\mathcal{S}(\boldsymbol{x})]\mathrm{d}\boldsymbol{x} = \frac{k_1(\mathcal{S})}{m} + \frac{k_2(\mathcal{S})}{B - c_{\text{epr}}m}, \tag{3.8}$$

其中

$$k_1(\mathcal{S}) = \int_{\mathbb{R}^d} \omega(\boldsymbol{x})(1-\rho^2_{\mathcal{S}}(\boldsymbol{x}))F_Y(\boldsymbol{x})(1-F_Y(\boldsymbol{x}))\mathrm{d}\boldsymbol{x}$$

$$k_2(\mathcal{S}) = c_{\mathcal{S}} \int_{\mathbb{R}^d} \omega(\boldsymbol{x})\rho^2_{\mathcal{S}}(\boldsymbol{x})F_Y(\boldsymbol{x})(1-F_Y(\boldsymbol{x}))\mathrm{d}\boldsymbol{x}. \tag{3.9}$$

由于 $k_1(\mathcal{S})$ 和 $k_2(\mathcal{S})$ 均为非负数，因此 $k_1(\mathcal{S})$ 和$k_2(\mathcal{S})$ 有明确定义（即为有限值）的充要条件是

$$k_1(\mathcal{S}) + c_{\mathcal{S}}^{-1}k_2(\mathcal{S}) = \int_{\mathbb{R}^d} \omega(\boldsymbol{x})F_Y(\boldsymbol{x})(1-F_Y(\boldsymbol{x}))\mathrm{d}\boldsymbol{x} < \infty, \tag{3.10}$$

但这并不一定对任意选择的 ω成立。例如，当 $\omega(\boldsymbol{x}) \equiv 1$时，(3.10) 在 $d=1$ 若 $\mathbb{E}[|Y|^{1+\delta}] < \infty$ 对某个 $\delta > 0$成立。然而，当 $d \geq 2$时，由于 Y的分布支集是无界的，(3.10) 通常并不成立，因为 $F_Y^{-1}([\varepsilon, 1-\varepsilon])$ 在 $\mathbb{R}^d$ 上可能具有无穷勒贝格测度，且 $\varepsilon > 0$对某个 $\omega(\boldsymbol{x})$ 可积可确保 (3.10)，即

$$\int_{\mathbb{R}^d} \omega(\boldsymbol{x})F_Y(\boldsymbol{x})(1-F_Y(\boldsymbol{x}))\mathrm{d}\boldsymbol{x} < \int_{\mathbb{R}^d} \omega(\boldsymbol{x})\mathrm{d}\boldsymbol{x} < \infty.$$

一些典型的可积 $\omega(\boldsymbol{x})$ 选择包括 $\omega(\boldsymbol{x})$，其中 $= \mathbf{1}_{\mathcal{T}}$ 是一个感兴趣的有界域，或者 $\mathcal{T} \subset \mathbb{R}^d$ ($\omega(\boldsymbol{x})$ 具有相当快的衰减尾部，因为 $\|\boldsymbol{x}\|_2 \to \infty$。在接下来的讨论中，我们假设 (3.10) 成立（之后将其编纂为假设5.5）。在 ω 第6节的数值结果中，我们对作出不同选择。

3.3. 探索-利用权衡。 公式（3.8）与最初在 [35], 其中 k_1 和 k_2 分别衡量探索和利用所造成的误差。注意， $L_{\mathcal{S}}$ 对于有效的探索率 m（即 $0 < m < B/c_{\mathrm{epr}}$），是一个严格凸函数，并在 $m^*_{\mathcal{S}}$ 处取得唯一最小值，对应的最小损失为 $L^*_{\mathcal{S}}$:

$$m^*_{\mathcal{S}} = \frac{B}{c_{\mathrm{epr}} + \sqrt{\frac{c_{\mathrm{epr}}k_2(\mathcal{S})}{k_1(\mathcal{S})}}}, \quad L^*_{\mathcal{S}} \coloneqq \min_{0<m<\frac{B}{c_{\mathrm{epr}}}} L_{\mathcal{S}}(m) = \frac{(\sqrt{c_{\mathrm{epr}}k_1(\mathcal{S})} + \sqrt{k_2(\mathcal{S})})^2}{B} =: \frac{\gamma_{\mathcal{S}}}{B}. \tag{3.11}$$

最优子集 $\mathcal{S}$ 是使经 m 优化后的损失值最小的子集，

$$\mathcal{S}^* = \arg\min_{\mathcal{S}\subseteq\{1:n\}} \gamma_{\mathcal{S}}. \tag{3.12}$$

如果均匀探索策略在探索阶段收集 $m^*_{\mathcal{S}^*}$ 个联合样本，并在利用阶段使用模型 $\mathcal{S}^*$ ，则称该策略为最优策略。实际上，这是一个模型选择过程，因为最优探索策略会通过以最优方式平衡探索与利用之间的权衡，选择能使误差最小的模型子集。在接下来的讨论中，我们假设 $\mathcal{S}^*$ 是唯一的。

作为对上述过程（具有神谕信息）的基准，可以考虑一种经验（ECDF）过程：将全部预算用于对高保真模型Y进行采样，而忽略低保真模型。以下结果描述了这两种方法之间的误差关系。

引理 3.1。 设 $L^*_{\mathcal{S}^*} = \gamma_{\mathcal{S}^*}/B$ 为上述均匀探索策略所达到的最小误差，并且 $c_{epr}\int_{\mathbb{R}^d}\omega(\boldsymbol{x})F_Y(\boldsymbol{x})(1-F_Y(\boldsymbol{x}))d\boldsymbol{x}/B$为经验累积分布函数估计量对于 F_Y所达到的期望误差，则

$$\frac{c_{epr}\int_{\mathbb{R}^d}\omega(\boldsymbol{x})F_Y(\boldsymbol{x})(1-F_Y(\boldsymbol{x}))d\boldsymbol{x}/B}{\gamma_{\mathcal{S}^*}/B} \geq \frac{1}{2\left(\frac{c_{\mathcal{S}}}{c_{epr}}+\mathbb{E}_Z[(1-\rho^2_{\mathcal{S}^*}(Z))]\right)} \geq \frac{1}{4}$$

其中， Z 是一个具有（未归一化）密度 ω（$\boldsymbol{z}$）F_Y（$\boldsymbol{z}$）（$1-F_Y$（$\boldsymbol{z}$））的随机变量。

证明。 我们有

$$\begin{aligned}\frac{c_{\mathrm{epr}}\int_{\mathbb{R}^d}\omega(\boldsymbol{x})F_Y(\boldsymbol{x})(1-F_Y(\boldsymbol{x}))\mathrm{d}\boldsymbol{x}/B}{\gamma_{\mathcal{S}^*}/B} &= \frac{c_{\mathrm{epr}}\int_{\mathbb{R}^d}\omega(\boldsymbol{x})F_Y(\boldsymbol{x})(1-F_Y(\boldsymbol{x}))\mathrm{d}\boldsymbol{x}}{\left(\sqrt{c_{\mathrm{epr}}k_1(\mathcal{S}^*)}+\sqrt{k_2(\mathcal{S}^*)}\right)^2}\\ &\geq \frac{c_{\mathrm{epr}}\int_{\mathbb{R}^d}\omega(\boldsymbol{x})F_Y(\boldsymbol{x})(1-F_Y(\boldsymbol{x}))\mathrm{d}\boldsymbol{x}}{2\left(c_{\mathrm{epr}}k_1(\mathcal{S}^*)+k_2(\mathcal{S}^*)\right)}\\ &\geq \frac{1}{2\left(\frac{c_{\mathcal{S}}}{c_{\mathrm{epr}}}+\mathbb{E}_{\boldsymbol{x}}[(1-\rho^2_{\mathcal{S}^*}(\boldsymbol{x}))]\right)} \geq \frac{1}{4},\end{aligned}$$

其中，期望 $\mathbb{E}_{\boldsymbol{x}}[\cdot]$ 是相对于以下对象取得的

$$\boldsymbol{x} \sim \frac{\omega(\boldsymbol{x})F_Y(\boldsymbol{x})(1-F_Y(\boldsymbol{x}))\mathrm{d}\boldsymbol{x}}{\int_{\mathbb{R}^d}\omega(\boldsymbol{z})F_Y(\boldsymbol{z})(1-F_Y(\boldsymbol{z}))\mathrm{d}\boldsymbol{z}},$$

最后一个不等式由注意到 $c_{\mathcal{S}} \leq c_{\mathrm{epr}}$ 和 $0 \leq \mathbb{E}_{\boldsymbol{x}}[(1-\rho^2_{\mathcal{S}^*}(\boldsymbol{x}))] \leq 1$而得。 ■

因此，与经验累积分布函数估计量相比，均匀探索策略的相对效率无条件地以下界 $1/4$为界，从而均匀探索策略最差也只会实现朴素经验累积分布函数过程4倍的损失值。另一方面，当 $c_{\mathcal{S}}/c_{\mathrm{epr}}$ 和 $\mathbb{E}_{\boldsymbol{x}}[(1-\rho^2_{\mathcal{S}^*}(\boldsymbol{x}))]$ 都较小时，相对效率为 $\gg 1$。例如，当 $X_{\mathcal{S}^*}$的采样成本远低于 Y，且 $h(X_{\mathcal{S}};\boldsymbol{x})$对于 $\mathbf{1}_{\{Y\leq\boldsymbol{x}\}}$ 在所有 $\boldsymbol{x}\in\mathbb{R}^d$下均为“良好”的控制变量时，就会出现这种情况；在多保真应用中，这两种情况都很现实。

4. 从线性近似中选择控制变量。我们提出了一种选择控制变量的程序 h，其归结为构造 $\mathbb{E}[\mathbf{1}_{\{Y\leq\boldsymbol{x}\}}|X_{\mathcal{S}}]$的近似，这些近似既能与 Y 保持较高的相关性，又符合预算要求。虽然在特定情况下可以为近似生成特殊形式，但我们的目标是采用一种简单且通用的选择，适用于许多实际应用。

回顾一下， $X_{\mathcal{S}^+} = (\mathbf{1}, X_i^\top)^\top_{i\in\mathcal{S}} \in \mathbb{R}^{d_{\mathcal{S}}+1}$。对于 $i\in\{1:\ d\}$，令 $\boldsymbol{\beta}^{(i)}_{\mathcal{S}^+}$ 表示用于使用 $X_{\mathcal{S}^+}$估计 i的第 Y 个分量的最优线性投影系数：

(4.1) $\boldsymbol{\beta}^{(i)}_{\mathcal{S}^+} = (\mathbb{E}[X_{\mathcal{S}^+}X_{\mathcal{S}^+}^\top])^{-1}\mathrm{Cov}[X_{\mathcal{S}^+}, Y^{(i)}] \in \mathbb{R}^{d_{\mathcal{S}}+1}$， $\boldsymbol{B}_{\mathcal{S}^+} = [\boldsymbol{\beta}^{(1)}_{\mathcal{S}^+},\cdots,\boldsymbol{\beta}^{(d)}_{\mathcal{S}^+}] \in \mathbb{R}^{(d_{\mathcal{S}}+1)\times d}$，

1 使用 $X_{\mathcal{S}}$ 和 $\mathbf{1}$ 对 Y 进行的最小二乘近似为

$$H_{\mathcal{S}}(X_{\mathcal{S}}) := (X_{\mathcal{S}^+}^\top\boldsymbol{B}_{\mathcal{S}^+})^\top = \begin{bmatrix} X_{\mathcal{S}^+}^\top\boldsymbol{\beta}^{(1)}_{\mathcal{S}^+} \\ \vdots \\ X_{\mathcal{S}^+}^\top\boldsymbol{\beta}^{(d)}_{\mathcal{S}^+}\end{bmatrix} \in \mathbb{R}^d.$$

当所有量都是标量，即 $d = d_1 = \cdots = d_n = 1$ 时，可以直接对 $H_{\mathcal{S}}$ 进行操作，以估计 Y [35, 36] 的统计量。这种方法易于实现，并且对一阶统计量具有一定的稳健性；但由于线性近似的限制，当需要学习 Y 的整个分布时，更容易受到模型设定错误的影响（例如线性模型的表达能力、噪声假设等） [36]。

为了解决这一问题，我们在 $H_{\mathcal{S}}$ 之外再执行一个额外的非线性步骤。具体而言，我们考虑以下一族对估计量 $H_{\mathcal{S}}$ 进行切片的控制变量：

$$h(X_{\mathcal{S}}; \boldsymbol{x}) = \mathbf{1}_{\{H_{\mathcal{S}} \le \boldsymbol{x}\}}. \tag{4.2}$$

直观地说，当 $\mathbb{E}[\|Y - H_{\mathcal{S}}\|_2^2]$ 较小时，我们可能会认为 $\mathbf{1}_{\{Y \le \boldsymbol{x}\}}$ 和 $h(X_{\mathcal{S}}; \boldsymbol{x})$ 具有相关性。然而，当 $\boldsymbol{x}$ 接近 Y 的尾部时，情况可能并非如此。例如，假设 $d = 1$，且令 (X, Y) 为相关性为 ρ 的标准联合高斯随机向量，

$$\lim_{x \to -\infty} \operatorname{Corr}(\mathbf{1}_{\{Y \le x\}}, \mathbf{1}_{\{X \le x\}}) = \lim_{x \to 0} \frac{C_{X,Y}(x, x)}{x} = \begin{cases} 1, & |\rho| = 1, \\ 0, & |\rho| < 1 \end{cases}$$

其中，$C_{X,Y}(x, y) = \mathbb{P}\left(\Phi^{-1}(X) \le x, \ \Phi^{-1}(Y) \le y\right)$ 是高斯 copula，Φ^{-1} 是标准正态分布的分位数；参见 [24]。因此，当 $|x| \to \infty$ 时，除非 $|\rho| = 1$（即仅当 $X \propto Y$ 时），$\mathbf{1}_{\{X \le x\}}$ 并不是 $\mathbf{1}_{\{Y \le x\}}$ 的良好控制变量。尽管如此，我们在第6节中的实验表明，在多保真度模拟中，对于许多场景，$h(X_{\mathcal{S}}; \boldsymbol{x})$ 实际上是一个合理的控制变量选择，这说明上述情形在我们所关注的应用中并不常见。我们在第5.1节中讨论使用 (4.2) 作为控制变量的计算方面。

选取 h 如（4.2）所示，系数 $\alpha(\boldsymbol{x})$ 在（3.5）中可显式计算为

$$\alpha(\boldsymbol{x}) = \begin{cases} \frac{F_{Y \vee H_{\mathcal{S}}}(\boldsymbol{x}) - F_Y(\boldsymbol{x}) F_{H_{\mathcal{S}}}(\boldsymbol{x})}{F_{H_{\mathcal{S}}}(\boldsymbol{x})(1 - F_{H_{\mathcal{S}}}(\boldsymbol{x}))} & \boldsymbol{x} \in \operatorname{supp}(F_{H_{\mathcal{S}}}(\boldsymbol{x}))^\circ \\ 0 & \text{otherwise} \end{cases}. \tag{4.3}$$

一个有用的技术结果是，$\alpha(\boldsymbol{x})$ 有界。

引理4.1. 设 $\alpha(\boldsymbol{x})$ 如（4.3）所示，则 $|\alpha(\boldsymbol{x})| \le 1$。

证明。 只需检查以下条件是否同时成立：$\boldsymbol{x} \in \operatorname{supp}(F_{H_{\mathcal{S}}}(\boldsymbol{x}))^\circ$，$\alpha(\boldsymbol{x}) \le 1$ 以及 $-\alpha(\boldsymbol{x}) \le 1$：

$$\begin{aligned} \frac{F_{Y \vee H_{\mathcal{S}}}(\boldsymbol{x}) - F_Y(\boldsymbol{x}) F_{H_{\mathcal{S}}}(\boldsymbol{x})}{F_{H_{\mathcal{S}}}(\boldsymbol{x})(1 - F_{H_{\mathcal{S}}}(\boldsymbol{x}))} &\le \frac{F_Y(\boldsymbol{x}) \wedge F_{H_{\mathcal{S}}}(\boldsymbol{x}) - F_Y(\boldsymbol{x}) F_{H_{\mathcal{S}}}(\boldsymbol{x})}{F_{H_{\mathcal{S}}}(\boldsymbol{x})(1 - F_{H_{\mathcal{S}}}(\boldsymbol{x}))} \\ &= \frac{F_Y(\boldsymbol{x})}{F_{H_{\mathcal{S}}}(\boldsymbol{x})} \wedge \frac{1 - F_Y(\boldsymbol{x})}{1 - F_{H_{\mathcal{S}}}(\boldsymbol{x})} \le 1, \end{aligned}$$

and

$$\begin{aligned} \frac{F_Y(\boldsymbol{x}) F_{H_{\mathcal{S}}}(\boldsymbol{x}) - F_{Y \vee H_{\mathcal{S}}}(\boldsymbol{x})}{F_{H_{\mathcal{S}}}(\boldsymbol{x})(1 - F_{H_{\mathcal{S}}}(\boldsymbol{x}))} &\le \frac{F_Y(\boldsymbol{x}) F_{H_{\mathcal{S}}}(\boldsymbol{x}) - F_Y(\boldsymbol{x}) + 1 - F_{H_{\mathcal{S}}}(\boldsymbol{x})}{F_{H_{\mathcal{S}}}(\boldsymbol{x})(1 - F_{H_{\mathcal{S}}}(\boldsymbol{x}))} \wedge \frac{F_Y(\boldsymbol{x}) F_{H_{\mathcal{S}}}(\boldsymbol{x})}{F_{H_{\mathcal{S}}}(\boldsymbol{x})(1 - F_{H_{\mathcal{S}}}(\boldsymbol{x}))} \\ &\le \frac{1 - F_Y(\boldsymbol{x})}{F_{H_{\mathcal{S}}}(\boldsymbol{x})} \wedge \frac{F_Y(\boldsymbol{x})}{1 - F_{H_{\mathcal{S}}}(\boldsymbol{x})} \le 1. \end{aligned}$$

■

5. 算法。 我们重新审视图1中的cvMDL算法图1：损失函数 $L_{\mathcal{S}}$ （3.8）是我们希望优化的目标损失函数，但它需要预言机统计量（即 $k_1(\mathcal{S})$ 和 $k_2(\mathcal{S})$ ）。因此，我们用本节描述的近似 $L_{\mathcal{S}}$ 来替代它。此外，“分析模型”步骤中的计算现在更加清晰：预言机计算由（3.12）和（3.11）给出。在实际的算法场景中，我们用一致的近似计算替代这些计算，这正是本节讨论的主题。

使用近似量计算 $L_{\mathcal{S}}$时，式（3.8）中的显式探索-利用损失分解可能不再成立。然而，如果我们估计的量足够准确，则这种分解预计仍近似有效。因此，在设计实际算法时，我们使用 Oracle损失形式（3.8）（采用估计系数）而不是（3.7）作为模型选择的标准。我们在数值实验部分提供了一些经验证据，表明这种替换对模型选择的影响很小。

由于所提出的估计量会在收集新的探索样本时发生变化，因此必须明确体现其对探索样本数量的依赖。对于 $\mathcal{S} \subseteq \{1: n\}$，我们令 $L_{\mathcal{S}}(\cdot; t)$ 表示在收集了 t 探索样本后得到的估计损失函数 $L_{\mathcal{S}}$。然后令 $\widehat{m}^*_{\mathcal{S}}$ 表示最优探索样本量 $m^*_{\mathcal{S}}$所对应的估计量。总结来说，cvMDL算法背后的直觉是：我们利用当前收集的探索数据（t 样本）来寻找估计最优模型（$\mathcal{S}^*$）及相应的探索率（$\widehat{m}^*_{\widehat{\mathcal{S}}^*}$）。根据 $\widehat{m}^*_{\widehat{\mathcal{S}}^*}$ 相对于 t的值，我们决定继续探索，还是转而进行利用。

5.1. 预言机量的估计量。 本节讨论在使用第 $L_{\mathcal{S}}$、 $m^*_{\mathcal{S}}$和 $\alpha(\boldsymbol{x})$ 中介绍的线性近似器实例化 cvMDL 时，如何根据探索数据估计 4。控制变量 $h(X_{\mathcal{S}}; \boldsymbol{x}) = \mathbf{1}_{\{H_{\mathcal{S}}(X_{\mathcal{S}}) \le \boldsymbol{x}\}}$ 属于由 $\boldsymbol{\beta}^{(i)}_{\mathcal{S}^+}$、 $i \in \{1$ 刻画的参数族：根据（ $d\}$ 4.1 ），该参数可由探索数据估计。

回顾（3.1）， $\mathcal{S}$ 中所有低保真模型的第 ℓ个探索样本记为 $X_{\text{epr},\ell,\mathcal{S}}$。类似地，我们定义 $X_{\text{epr},\ell,\mathcal{S}^+} := (1, X^\top_{\text{epr},\ell,\mathcal{S}})^\top$。为估计 $\boldsymbol{\beta}^{(i)}_{\mathcal{S}^+}$，我们使用最小二乘估计量：

$$\widehat{\boldsymbol{\beta}}^{(i)}_{\mathcal{S}^+} = \boldsymbol{Z}^\dagger_{\mathcal{S}} \boldsymbol{Y}^{(i)} \qquad \boldsymbol{Z}_{\mathcal{S}} = \begin{bmatrix} X^\top_{\text{epr},1,\mathcal{S}^+} \\ \vdots \\ X^\top_{\text{epr},m,\mathcal{S}^+} \end{bmatrix} \in \mathbb{R}^{m \times (d_{\mathcal{S}}+1)} \quad \boldsymbol{Y}^{(i)} = \begin{bmatrix} Y^{(i)}_{\text{epr},1} \\ \vdots \\ Y^{(i)}_{\text{epr},m} \end{bmatrix} \in \mathbb{R}^m, \tag{5.1}$$

其中 $(X^\top_{\text{epr},\ell,1}, \ldots\ldots, Y^\top_{\text{epr}\,\ell})^\top_{\ell \in \{1:m\}}$ 为联合探索样本，并假设设计矩阵 $\boldsymbol{Z}_{\mathcal{S}}$具有列满秩[3]。

对于 $\boldsymbol{x} \in \mathbb{R}^d$， $h(X_{\mathcal{S}}; \boldsymbol{x})$ 可估计为

$$\widehat{h}(X_{\mathcal{S}}; \boldsymbol{x}) = \mathbf{1}_{\{\widehat{H}_{\mathcal{S}}(X_{\mathcal{S}}) \le \boldsymbol{x}\}} \qquad \widehat{H}_{\mathcal{S}}(X_{\mathcal{S}}) := \widehat{\boldsymbol{B}}^\top_{\mathcal{S}^+} X_{\mathcal{S}^+} =: \begin{bmatrix} X^\top_{\mathcal{S}^+} \widehat{\boldsymbol{\beta}}^{(1)}_{\mathcal{S}^+} \\ \vdots \\ X^\top_{\mathcal{S}^+} \widehat{\boldsymbol{\beta}}^{(d)}_{\mathcal{S}^+} \end{bmatrix}, \tag{5.2}$$

[3]这促使我们在算法1中引入最小探索规模条件，这是此处满秩的必要条件。

为便于记号表示，我们将 $H_{\mathcal{S}}(X_{\mathcal{S}})$ 和 $h_{\mathcal{S}}(X_{\mathcal{S}};\boldsymbol{x})$ 分别记为 $H_{\mathcal{S}}$和 $h_{\mathcal{S}}$，其中 $X_{\mathcal{S}^+}$ 是一般性的，不一定与探索数据和利用数据相关。对于涉及估计系数和使用探索数据得到的经验累积分布函数的量，我们引入一些额外的记号：

$$\widehat{F}_Y(\boldsymbol{x}) = \frac{1}{m}\sum_{\ell\in\{1:m\}} \mathbf{1}_{\{Y_{\mathrm{epr},\ell}\le\boldsymbol{x}\}}$$
$$\widehat{F}_{\widehat{H}_{\mathcal{S}}}(\boldsymbol{x}) = \frac{1}{m}\sum_{\ell\in\{1:m\}} \widehat{h}(X_{\mathrm{epr},\ell,\mathcal{S}};\boldsymbol{x}) = \frac{1}{m}\sum_{\ell\in\{1:m\}} \mathbf{1}_{\{\widehat{H}_{\mathcal{S}}(X_{\mathrm{epr},\ell,\mathcal{S}})\le\boldsymbol{x}\}}$$
$$\widehat{F}_{Y\vee\widehat{H}_{\mathcal{S}}}(\boldsymbol{x}) = \frac{1}{m}\sum_{\ell\in\{1:m\}} \mathbf{1}_{\{Y_{\mathrm{epr},\ell}\vee\widehat{H}_{\mathcal{S}}(X_{\mathrm{epr},\ell,\mathcal{S}})\le\boldsymbol{x}\}}.$$

为计算损失函数近似，我们构造 k_1 和 k_2 （3.9）中的近似，这要求我们计算 $\rho^2_{\mathcal{S}}$ （3.6）。为此，请注意

$$(1-\rho^2_{\mathcal{S}}(\boldsymbol{x}))F_Y(\boldsymbol{x})(1-F_Y(\boldsymbol{x})) = \mathbb{E}[(\mathbf{1}_{\{Y\le\boldsymbol{x}\}} - F_Y(\boldsymbol{x})) - \alpha(\boldsymbol{x})(\mathbf{1}_{\{H_{\mathcal{S}}\le\boldsymbol{x}\}} - F_{H_{\mathcal{S}}}(\boldsymbol{x}))]^2,$$

其中，$\alpha(\boldsymbol{x})$ 定义于（4.3）中。期望中的量是两个中心化伯努利随机变量 $\mathbf{1}_{\{Y\le\boldsymbol{x}\}}$和 $\mathbf{1}_{\{H_{\mathcal{S}}\le\boldsymbol{x}\}}$之间的均方回归残差。因此，对于（$1-\rho^2_{\mathcal{S}}(\boldsymbol{x})$）$F_Y(\boldsymbol{x})(1-F_Y(\boldsymbol{x}))$，一种自然估计量是计算 $\mathbf{1}_{\{Y\le x\}}$与具有协变量 $\mathbf{1}_{\{\widehat{H}_{\mathcal{S}}\le x\}}$的回归器之间的经验均方差，这要求有关$Y$的数据。由于我们在探索样本$Y_{\mathrm{epr}}$上拥有 Y的（未中心化）数据，对于 $j\in\{1:\ m\}$、，因此我们可以在探索数据位置上，使用协变量 $H_{\mathcal{S}}$ 以及截距项，对 Y的回归器进行评估。由此得到以下估计量 $\mathcal{K}_1$，用于（$1-\rho^2_{\mathcal{S}}(\boldsymbol{x})$） $F_Y(\boldsymbol{x})(1-F_Y(\boldsymbol{x}))$

$$\mathcal{K}_1(\boldsymbol{x}) = \frac{1}{m}\sum_{\ell\in\{1:m\}} (\mathbf{1}_{\{Y_{\mathrm{epr},\ell}\le\boldsymbol{x}\}} - r_\ell(\boldsymbol{x}))^2 \qquad \begin{bmatrix} r_1(\boldsymbol{x}) \\ \vdots \\ r_m(\boldsymbol{x}) \end{bmatrix} = \boldsymbol{W}_{\mathcal{S}}\boldsymbol{W}_{\mathcal{S}}^{\dagger} \begin{bmatrix} \mathbf{1}_{\{Y_{\mathrm{epr},1}\le\boldsymbol{x}\}} \\ \vdots \\ \mathbf{1}_{\{Y_{\mathrm{epr},m}\le\boldsymbol{x}\}} \end{bmatrix}, \tag{5.3}$$

其中

$$\boldsymbol{W}_{\mathcal{S}} = \begin{bmatrix} 1 & \widehat{h}_{\mathcal{S}}(X_{\mathrm{epr},1,\mathcal{S}};\boldsymbol{x}) \\ \vdots & \vdots \\ 1 & \widehat{h}_{\mathcal{S}}(X_{\mathrm{epr},m,\mathcal{S}};\boldsymbol{x}) \end{bmatrix}.$$

这使我们能够将 $\rho^2_{\mathcal{S}}(\boldsymbol{x})F_Y(\boldsymbol{x})(1-F_Y(\boldsymbol{x}))$ 估计为

$$\mathcal{K}_2(\boldsymbol{x}) = \widehat{F}_Y(\boldsymbol{x})(1-\widehat{F}_Y(\boldsymbol{x})) - \mathcal{K}_1(\boldsymbol{x}). \tag{5.4}$$

因此，我们可以将 $k_1(\mathcal{S})$ 和 $k_2(\mathcal{S})$ 估计为

$$\widehat{k}_1(\mathcal{S}) = \int_{\mathbb{R}^d} \omega(\boldsymbol{x})\mathcal{K}_1(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \qquad \widehat{k}_2(\mathcal{S}) = c_{\mathcal{S}}\int_{\mathbb{R}^d} \omega(\boldsymbol{x})\mathcal{K}_2(\boldsymbol{x})\mathrm{d}\boldsymbol{x}. \tag{5.5}$$

上述 $k_1(\mathcal{S})$ 和 $k_2(\mathcal{S})$ 的估计量为正，并且在这些量有定义时，与其经验估计量一致（见附录 A.2.3），这对于我们后续的一致性结果是一个至关重要的认识。将上述估计代入（3.8）和（3.11）即可得到 $L_{\mathcal{S}}$ 和 $m^*_{\mathcal{S}}$ 的估计值：

$$\widehat{L}_{\mathcal{S}}(z;m) = \frac{\widehat{k}_1(\mathcal{S})}{z} + \frac{\widehat{k}_2(\mathcal{S})}{B - c_{\mathrm{epr}} z} \qquad \widehat{m}^*_{\mathcal{S}} = \frac{B}{c_{\mathrm{epr}} + \sqrt{\frac{c_{\mathrm{epr}}\widehat{k}_2(\mathcal{S})}{\widehat{k}_1(\mathcal{S})}}}. \tag{5.6}$$

请注意，$\widehat{L}_{\mathcal{S}}(z;m)$ 有一个参数 m，表示用于计算 $\widehat{k}_1(\mathcal{S})$ 和 $\widehat{k}_2(\mathcal{S})$ 的探索样本数量；还有一个变量 z，表示用于评估 $\widehat{L}_{\mathcal{S}}$ 的探索率。我们将 $\mathcal{S}^*$ 定义为该估计量选出的最优模型，

$$\widehat{\mathcal{S}}^* = \operatorname*{arg\,min}_{\mathcal{S}\subseteq\{1:n\}} \widehat{L}_{\mathcal{S}}(\widehat{m}^*_{\mathcal{S}};m), \tag{5.7}$$

这与预言机选择（3.12）相对应。我们已描述了完成图1中探索阶段图1所需的所有量。接下来需要说明如何在 $F_{\widehat{\mathcal{S}}^*}$图1的利用阶段图1生成累积分布函数估计量。

我们的利用目标是为（3.4）生成一个估计量，因此我们需要估计$\alpha(\boldsymbol{x})$：

$$\widehat{\alpha}(\boldsymbol{x}) = \frac{\widehat{F}_{Y\vee\widehat{H}_{\mathcal{S}}}(\boldsymbol{x}) - \widehat{F}_Y(\boldsymbol{x})\widehat{F}_{\widehat{H}_{\mathcal{S}}}(\boldsymbol{x})}{\widehat{F}_{\widehat{H}_{\mathcal{S}}}(\boldsymbol{x})(1-\widehat{F}_{\widehat{H}_{\mathcal{S}}}(\boldsymbol{x}))} \qquad \boldsymbol{x}\in \operatorname{supp}(\widehat{F}_{\widehat{H}_{\mathcal{S}}}(\boldsymbol{x}))^\circ, \tag{5.8}$$

以及 $\widehat{\alpha}(\boldsymbol{x}) = 0$ 在其他情况下为零。根据与引理4.1类似的推理，可得

$$|\widehat{\alpha}(\boldsymbol{x})| \le 1. \tag{5.9}$$

最后，基于估计参数的利用 估计量 $F_{\mathcal{S}}(\boldsymbol{x})$ 用于 $F_Y(\boldsymbol{x})$，利用$N_{\mathcal{S}}$ 利用样本（即耗尽剩余预算 B），其定义为：

$$\widetilde{F}_{\mathcal{S}}(\boldsymbol{x}) \coloneqq \widehat{F}_Y(\boldsymbol{x}) - \frac{1}{m}\sum_{\ell\in\{1:m\}}\left(\widehat{\alpha}(\boldsymbol{x})\widehat{h}_{\mathcal{S}}(X_{\mathrm{epr},\ell,\mathcal{S}};\boldsymbol{x}) - \frac{1}{N_{\mathcal{S}}}\sum_{j\in\{1:N_{\mathcal{S}}\}}\widehat{\alpha}(\boldsymbol{x})\widehat{h}_{\mathcal{S}}(X_{\mathrm{ept},\ell,\mathcal{S}};\boldsymbol{x})\right), \tag{5.10}$$

其中，$\mathcal{S} = \mathcal{S}^*$ 是基于 $k_1(\mathcal{S})$ 和 $k_2(\mathcal{S})$ 选定的模型。通过观察，我们发现 $\widetilde{F}_{\mathcal{S}}(\boldsymbol{x})$ 是对 F_Y 的分段仿射校正，其中校正基于控制变量 $h_{\mathcal{S}}$。

注5.1。 估计量 $\widehat{\alpha}(\boldsymbol{x})$ 未定义，并且在 $\boldsymbol{x}$ 位于的支集之外时手动设为零 $F_{\widehat{H}_{\mathcal{S}}}$，因为在这种情况下分母为零。或者，也可以定义 $\widehat{\alpha}(\boldsymbol{x})$，当 $\boldsymbol{x}$ 位于 $F_{\widehat{H}_{\mathcal{S}}}$ 的支集之外时，将其定义为 $\widehat{\alpha}(\boldsymbol{x}')$，其中 $\boldsymbol{x}'$ 位于 $F_{\widehat{H}_{\mathcal{S}}}$ 的支集内，能够被准确估计，但仍接近于 $\boldsymbol{x}$。作为说明，考虑 $d = 1$。假设

$\sim\alpha(x)$ 是 x 在 $supp(F_{H_S})$ 中以及 $supp(F_{\widehat{H}_S}(x)) = [x_{\min}, x_{\max}]$ 中关于某个$\sim x_{\min} < x_{\max}$的连续函数，对于 $x \in supp(F_{\widehat{H}_S}(x))^c$，我们可以估计 $\alpha(x)$ 在 $[x_{\min}, x_{\max}]$ 之外的值，如下所示：

$$\widehat{\alpha}(x) = \begin{cases} \frac{\widehat{F}_{Y\vee\widehat{H}_S}(x(\tau)) - \widehat{F}_Y(x(\tau))\tau}{\tau(1-\tau)} & x \le x_{\min} \\ \frac{\widehat{F}_{Y\vee\widehat{H}_S}(x(1-\tau)) - \widehat{F}_Y(x(1-\tau))(1-\tau)}{\tau(1-\tau)} & x \ge x_{\max}, \end{cases} \tag{5.11}$$

$\widehat{}$其中，$x(\tau)$ 和 $x(1-\tau)$ 分别是 $F_{\widehat{H}_S}$的 τ分位数和 $1-\tau$分位数，其中某个较小的 $\tau \in (0、1)$：

$$x(\tau) = \widehat{F}^{-1}_{\widehat{H}_S}(\tau) \qquad\qquad x(1-\tau) = \widehat{F}^{-1}_{\widehat{H}_S}(1-\tau).$$

这使我们能够在 $[x_{\min}, x_{\max}]$之外，即尾部区域内，获得 F_Y 的非平凡估计。当 $d \ge 2$时，可以通过将尾部区域中的点投影到 $\mathbb{R}^d$ 中的某个有界集合来推广上述思想，该集合包含区域中的大部分测度。

$\widetilde{}$5.2. 利用阶段 CDF 估计器的单调性。 </style}根据构造，() 是 的某个 维矩形划分上的分段常数函数，但由于构造中所使用的估计量存在波动，它不一定在每个方向上都是单调不减函数。为解决这一问题，我们引入一个对 值域中的值进行逐维递归排序的后处理过程，以恢复所需的单调性，并确保我们计算的是一个实际的分布函数。为将 () 表示为一个 维张量，我们引入指标集 (，1，……，$z_{i,\ M_i}$)，$-\infty = z_{i,\ 1} \le \cdots \le z_{i,\ M_i} = +\infty$，$\widetilde{}$其中 $z_{i,\ j}$ 是与 $F_S(\boldsymbol{x})$ 相关的投影划分点中的第 j个次序统计量，$\widetilde{}$而 M_i 是此类投影点的总数。利用这一记号，我们将 $F_S(\boldsymbol{x})$ 表示为张量 $\boldsymbol{T}$，其中

$$\widetilde{F}_S(\boldsymbol{x}) = \boldsymbol{T}_{z_{1,s_1},\cdots,z_{d,s_d}} \qquad\qquad \boldsymbol{x} \in \prod_{i\in\{1:d\}} [z_{i,s_i}, z_{i,s_i+1}).$$

通过对 $\boldsymbol{T}$ 中的条目交替进行逐维排序，可以恢复各维度上所需的单调性，直到收敛。具体步骤见算法 5.1。

当 $d = 2$ 给定时，一个示例如下：

$$\begin{bmatrix} 0.7 & 0.4 & 0 \\ 0.3 & 0.5 & 0.2 \\ 1 & 0.8 & 0.6 \end{bmatrix} \xrightarrow{\text{sort rows}} \begin{bmatrix} 0 & 0.4 & 0.7 \\ 0.2 & 0.3 & 0.5 \\ 0.6 & 0.8 & 1 \end{bmatrix} \xrightarrow{\text{sort columns}} \begin{bmatrix} 0 & 0.3 & 0.5 \\ 0.2 & 0.4 & 0.7 \\ 0.6 & 0.8 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 0.7 & 0.4 & 0 \\ 0.3 & 0.5 & 0.2 \\ 1 & 0.8 & 0.6 \end{bmatrix} \xrightarrow{\text{sort columns}} \begin{bmatrix} 0.3 & 0.4 & 0 \\ 0.7 & 0.5 & 0.2 \\ 1 & 0.8 & 0.6 \end{bmatrix} \xrightarrow{\text{sort rows}} \begin{bmatrix} 0 & 0.3 & 0.4 \\ 0.2 & 0.5 & 0.7 \\ 0.6 & 0.8 & 1 \end{bmatrix}$$

如上所示，排序经过有限步后最终会到达某个具有所需单调性的驻点（参见定理 5.2），但不同的排序顺序可能导致不同的$\widetilde{}$排序后的 CDF表示，当 $d \ge 2$时。然而，在我们的情形中，$F_S(\boldsymbol{x})$ 本身就是 Y的 CDF 的一种扰动，因此排序过程通常有助于稳定算法。第 6节对此进行了更为详细的实证研究。所述排序过程会在有限次迭代内收敛（即使 $\boldsymbol{T}$的取值达到单调性）。

算法 5.1 交替排序。

输入：表示估计的累积分布函数的张量 $\boldsymbol{T}$ $\widetilde{F}_{\mathcal{S}}(\boldsymbol{x})$

输出：具有所需单调性的排序后的张量 sorted($\boldsymbol{T}$)

将 sorted($\boldsymbol{T}$) 初始化为一个与 $\boldsymbol{T}$ 大小相同的全零张量
当 $\text{sorted}(\boldsymbol{T}) \neq \boldsymbol{T}$ **时**
  $\text{sorted}(\boldsymbol{T}) \leftarrow \boldsymbol{T}$
  对于 $i \in \{1:\ d\}$，**执行**
    对于 $\boldsymbol{z} := (z_{1,\ s_1},\ \ldots\ldots,\ z_{i-1,\ s_{i-1}},\ z_{i+1,\ s_{i+1}},\ \ldots\ldots,\ z_{d,\ s_d}) \in \otimes_{j\in\{1:d\}\setminus\{i\}} (z_{j,\ 1},\ \ldots\ldots,\ z_{j,\ M_i})$ **do**
      $\boldsymbol{T}[\boldsymbol{z}, :] \leftarrow \text{sort}(\boldsymbol{T}[\boldsymbol{z}, :])$，其中 $\boldsymbol{T}[\boldsymbol{z}, :] := (\boldsymbol{T}_{z_{1,\ s_1},\ldots,z_{i,\ j},\ldots,z_{d,\ s_d}})_{j\in\{1:M_i\}}$
    结束循环
  结束循环
结束循环
返回排序后的$\boldsymbol{T}$

定理 5.2。 假设 $\boldsymbol{T}$ 中的所有条目均不相同。文中 算法 5.1 所述的交替排序算法将在有限次迭代内收敛到具有所需单调性的驻点。

证明。 请参见附录 A.1。 ■

5.3. 探索采样。接下来，我们将描述决定继续探索时所采取的具体操作。特别地，我们需要定义图1中的函数 $Q(m, \widehat{m}_{\widehat{\mathcal{S}}^*})$。当当前的探索样本数 m 小于估计的最优样本数 $\widehat{m}^*_{\widehat{\mathcal{S}}^*}$时，函数 Q 决定如何增加 m。一种自然的选择是令 Q 为$Q(m, \widehat{m}^*_{\widehat{\mathcal{S}}^*}) = m+1$，即简单地增加一个额外的探索样本。在实践中，我们观察到，当 B 较大时，这种行为可能过于保守且耗时。作为替代方案，可以采用更激进策略，例如 $Q(m, \widehat{m}^*_{\widehat{\mathcal{S}}^*}) = \frac{1}{2}\left(m + \widehat{m}^*_{\widehat{\mathcal{S}}^*}\right)$，从而更快地缩小 m 与 $\widehat{m}_{\widehat{\mathcal{S}}^*}$之间的差距。然而，在某些情况下，这样做又过于激进。例如，如果 m 较小（如在初始化阶段），那么估计量可能是较差的近似，并且在某些情况下，$\widehat{m}_{\widehat{\mathcal{S}}^*}$ 会被严重高估，因此将 m 增加到 $\frac{1}{2}\left(m + \widehat{m}^*_{\widehat{\mathcal{S}}^*}\right)$ 实际上可能导致大幅超过 $m^*_{\mathcal{S}^*}$的预言机值。此类事件发生的概率通常为正，并且不会随着B 增加而趋于零。

作为在这两种保守与激进的行为之间的折中方案，我们选择以下形式：

$$Q(m, \widehat{m}^*_{\widehat{\mathcal{S}}^*}) = \begin{cases} 2m, & 1 \le m < \frac{\widehat{m}^*_{\widehat{\mathcal{S}}^*}}{2} \\ \frac{1}{2}\left(m + \widehat{m}^*_{\widehat{\mathcal{S}}^*}\right), & \frac{\widehat{m}^*_{\widehat{\mathcal{S}}^*}}{2} \le m < \widehat{m}^*_{\widehat{\mathcal{S}}^*} \end{cases}. \tag{5.12}$$

由于 $\widehat{m}^*_{\widehat{\mathcal{S}}^*}$ 与 B成正比，上述选择确保算法有足够的时间进行指数探索，其增长方式与 $\widehat{m}^*_{\widehat{\mathcal{S}}^*}$ 的取值无关，从而保证算法的效率和准确度。我们注意到，指数速率为二，以及在（5.12）中取

特殊的，并且在进行适当修改的前提下，可以分别替换为其他大于一的速率和非均匀平均操作。后续各节中的理论结论和数值模拟均假设 Q 采用上述形式，但对 Q 作出其他合理选择并不会改变理论结论。

我们已经完成了图1的所有技术描述。更为完整的伪代码版本见算法 5.2，其中对每个步骤都提供了更多细节。接下来，我们将证明，相对于 Oracle 生成的模型选择和预算分配策略，所提出的算法具有最优性保证。

算法 5.2 详细的 cvMDL算法。

输入： B：总预算，模型成本 $c_0,\ c_1,\ \ldots,\ c_n$
输出：针对 $F_Y(\boldsymbol{x})$ 的估计量

初始化探索 =TRUE
初始化 $m = \sum_{i\in\{1:\ n\}} d_i + 2$
生成（$Y,\ X_{\{1:n\}}$）的探索样本
当 探索 = TRUE**时**
  对于 $\mathcal{S} \subseteq \{1:\ n\}$ **执行**
    计算回归系数 $\widehat{\boldsymbol{\beta}}^{(i)}_{\mathcal{S}+},\ i \in \{1:\ d\}$ 来自（5.1）
    构建 $\widehat{H}_{\mathcal{S}}(X_{\mathcal{S}})$和 $\widehat{h}\ (X_{\mathcal{S}};\ \boldsymbol{x})$ 来自（5.2）
    计算回归系数$r_j\ (\boldsymbol{x}),\ j \in \{1:\ m\}$ 来自（5.3）
    构造 $\mathcal{K}_1$ 和 $\mathcal{K}_2$ 分别来自（5.3）和（5.4）
    评估 $\widehat{k_1}(\mathcal{S})$和 $\widehat{k_2}\ (\mathcal{S})$ 使用（5.5）以及对 $\mathbb{R}^d$ 应用求积规则
    计算 $\widehat{m}^*_{\mathcal{S}}$并 $\widehat{L_{\mathcal{S}}}\ (\cdot;\ m)$ 由（5.6）
    计算最小期望损失 $\widehat{L}_{\mathcal{S}}\ (m \vee \widehat{m}^*_{\mathcal{S}};\ m)$ 由（5.6）
  结束循环
  选择 $\widehat{\mathcal{S}}^* = \text{参数最小值}_{\mathcal{S}\subseteq\{1:n\}}\ \widehat{L}_{\mathcal{S}}(m \vee \widehat{m}^*_{\mathcal{S}};\ m)$;
  如果 $m < \widehat{m}^*_{\widehat{\mathcal{S}}^*}$ **then**
    生成 $Q(m,\ \widehat{m}^*_{\widehat{\mathcal{S}}^*}) - m$ 的额外样本（$Y,\ X_{\{1:n\}}$），其中 Q 在（5.12）中给出
    增加 m：$m \leftarrow Q(m,\ \widehat{m}^*_{\widehat{\mathcal{S}}^*})$
  else
    探索 =FALSE
  结束 if
结束 while
生成 $N_{\widehat{\mathcal{S}}^*}$ 样本 $X_{\widehat{\mathcal{S}}^*}$，其中 $N_{\mathcal{S}}$ 给定于（3.3）
构造 $\widehat{\alpha}\ (\boldsymbol{x})$ 用于 $\mathcal{S} \leftarrow \widehat{\mathcal{S}}^*$ 使用（5.8）
生成 $\widehat{\mathcal{S}}^*$ 利用估计量 $\widetilde{F}_{\widehat{\mathcal{S}}^*}$ 使用（5.10）。

5.4. 模型一致性与最优性。 现在，我们为 算法 5.2 （对应于 图1中的流程图）提供理论保证。总而言之，我们将证明，随着 $\widehat{}$预算 B 趋于无穷，所选择的模型子集 $\mathcal{S}^*$ 以及在 算法 5.2 中获取的探索样本数量 m 都将收敛于预言机最优模型 $\mathcal{S}^*$ 以及

最优探索样本数量分别为 $m^*_{\mathcal{S}^*}$。

为了继续推导我们的结果，我们需要一些技术假设。由于我们估计的是二阶矩，因此要求二阶矩存在。我们还要求不存在完全相关的低保真QoI对。这些要求具体表述为以下两个假设。

假设 5.3。 模型 $X_{\{1:n\}}$ 和 Y 具有有界二阶矩：

$$\mathbb{E}[\|X_{\{1:n\}}\|_2^2 + \mathbb{E}\|Y\|_2^2] < \infty. \tag{5.13}$$

假设5.4。 非中心二阶矩矩阵 $\mathbb{E}[X_{\{1:n\}+}X_{\{1:n\}+}^\top]$ 是可逆的，其中 $X_{\{1:n\}+} = (1, X_{\{1:n\}}^\top)^\top$。

假设 5.3 是我们为使预言量定义良好而对感兴趣量所要求的最小矩条件。违反假设5.4 的随机变量表现出完全多重共线性，在实践中相对少见。 假设5.4 被违反并不会导致我们的方法在概念上失效；唯一的后果是，所有线性回归过程都会遭遇最优协变量的不可辨识性。尽管有许多解决多重共线性的标准方法，例如删除协变量或正则化，但在我们的实验中并未出现违反该假设的情况，因此我们不采用这些补救措施。

模型选择过程要求估计平均的 ω-加权 L^2 范数。这要求我们对 ω 作出某些假设。

假设5.5。 权重 $\omega(\boldsymbol{x})$ 的选取应满足以下条件之一：

(a) $\|\omega\|_{L^\infty(\mathbb{R}^d)} < \infty$ （例如 $\omega(\boldsymbol{x}) \equiv 1$）且 $d = 1$；

或 *(b)* $\|\omega\|_{L^1(\mathbb{R}^d)} < \infty$。

我们所需的最后一个、更具技术性的假设涉及分布函数的某些正则性。具体而言，我们证明估计量 $\widehat{\alpha}(\boldsymbol{x})$ 在 $\boldsymbol{x}$ 中逐点收敛于预言机参数 $\alpha(\boldsymbol{x})$；为实现这一点，我们需要对模型选择过程中构造的 $F_{H_\mathcal{S}}$ 和 $F_{H_\mathcal{S}\vee Y}$ 的局部变差给出界。更技术性地说，一个充分的假设是：涉及 $X_{\{1:n\}}$ 和 Y 的某些 d 维投影的累积分布函数满足有界局部变差条件。

假设5.6。 定义

$$V(\boldsymbol{A}) := (X_{\mathcal{S}^+}^\top \boldsymbol{A})^\top \in \mathbb{R}^d \qquad\qquad \boldsymbol{A} := [\boldsymbol{A}^{(1)}, \cdots, \boldsymbol{A}^{(d)}] \in \mathbb{R}^{d_\mathcal{S} \times d},$$

并回顾式 $\boldsymbol{B}_{\mathcal{S}^+}$ 中的最优系数矩阵（4.1）。我们假设， V（$\boldsymbol{A}$）和V（$\boldsymbol{A}$）$\vee Y$的累积分布函数，记为 $F_{V(\boldsymbol{A})}$ 和 $F_{V(\boldsymbol{A})\vee Y}$，在 $\boldsymbol{B}_{\mathcal{S}^+}$ 附近对所有 $\mathcal{S}$均为全局利普希茨的。特别地，我们假设存在 $\varepsilon > 0$ 使得

$$\max_{\mathcal{S}\subseteq\{1:n\}} \sup_{\boldsymbol{A}:\|\boldsymbol{A}-\boldsymbol{B}_{\mathcal{S}^+}\|_F \le \varepsilon} \left\{\|F_{V(\boldsymbol{A})}\|_{Lip} + \|F_{V(\boldsymbol{A})\vee Y}\|_{Lip}\right\} = C < \infty,$$

其中， $\|\cdot\|_{Lip}$ 是定义如下的利普希茨常数：

$$\|f\|_{Lip} = \sup_{\boldsymbol{x}\neq\boldsymbol{x}'} \frac{|f(\boldsymbol{x}) - f(\boldsymbol{x}')|}{\|\boldsymbol{x} - \boldsymbol{x}'\|_2} \qquad\qquad f : \mathbb{R}^d \to \mathbb{R}.$$

这一最终假设不像前面的假设那样直观。要确保假设5.6 成立，一个不必要地更强的充分条件是：假设 Y 以及 $X_{\{1:n\}}$ 的所有单位线性组合都具有一致有界密度，并且每个高保真协变量 $Y^{(i)}$ 都与每个低保真度协变量 $X_j^{(r)}$ 相关，即 $\min_{i,\,j,\,r} |\mathrm{Corr}(Y^{(i)}, X_j^{(r)})| > 0$。或者，也可以假设相同的有界密度条件，并假设一个相当合理的条件：预言回归系数 $\boldsymbol{B}_{\mathcal{S}^+}$ 对每个 $\mathcal{S}$ 至少选择一个非确定性协变量。

现在，我们可以给出关于将 cvMDL算法应用于 $h(X_{\mathcal{S}};\ \boldsymbol{x})$ 的主要结果，其中该算法使用线性近似构造，并根据 (5.5) 和 (5.6) 估计相应的损失函数参数。特别地，我们有，自适应探索率 $m(B)$ 渐近地匹配第 3.3 节中定义的最优（Oracle）探索率 $m^*_{\mathcal{S}^*}$，并且所选模型 $\mathcal{S}(B)$ 收敛于最优（Oracle）模型 $\mathcal{S}^*$，当 $B \to \infty$：

定理 5.7（算法 5.2 中 cvMDL 的一致一致性与渐近最优性）。
设 $h(X_{\mathcal{S}};\ \boldsymbol{x})$ 定义于 (4.2)，即我们使用第 4 节中的线性近似估计量，并假设模型参数通过 (5.5) 和 (5.6) 进行估计。然后，考虑输入预算为 B 的算法 5.2，并令

- $m(B) = \widehat{m}_{\widehat{\mathcal{S}}^*}$ 表示由算法 5.2选择的探索样本数量，
- $\mathcal{S}(B) = \mathcal{S}^*$ 表示为利用阶段选择的模型，
- $F(\boldsymbol{x};\ B) = F_{\widehat{\mathcal{S}}^*}(\boldsymbol{x})$ 表示 F_Y的累积分布函数估计量。

在假设5.3、5.4、5.5和5.6下，几乎必然有，

$$\lim_{B\to\infty} \frac{m(B)}{m^*_{\mathcal{S}^*}} = 1, \tag{5.14a}$$

$$\lim_{B\to\infty} \mathcal{S}(B) = \mathcal{S}^*, \tag{5.14b}$$

$$\lim_{B\to\infty} \sup_{\boldsymbol{x}\in\mathbb{R}^d} |\widetilde{F}(\boldsymbol{x}; B) - F_Y(\boldsymbol{x})| = 0, \tag{5.14c}$$

其中，$\mathcal{S}^*$ 和 $m^*_{\mathcal{S}^*}$ 分别是在第3.3节中定义的唯一最优（*Oracle*）模型选择和探索样本数量。

证明见附录 A.4。结果（5.14c）不应令人意外，因为经验 CDF 估计器通常具有一致一致性。因此，虽然（5.14a）和（5.14b）表明，算法 5.2中的 cvMDL 在探索样本的选择以及跨模型的样本分配方面实现了最优性（相对于 Oracle），但（5.14c）并不能证明多保真估计量 $F(\boldsymbol{x};\ B)$ 优于仅使用高保真样本的经验 CDF 估计器，尽管它证实了 $F(\boldsymbol{x};\ B)$ 的表现符合预期。将 $F(\boldsymbol{x};\ B)$ 与标准经验 CDF 估计器区分开来的主要差异，是一个源于均值 ω加权 L^2 收敛速率的常数项；参见第 3.3 小节末尾附近的讨论。

定理中的陈述 5.7 和 [36, 定理 5.2] 相似，但前者所需的假设 要 弱得多，且保证更强。事实上，要使[36, 定理 5.2] 成立，必须假设 $\mathbb{E}[Y|X_{\mathcal{S}}]$ 是 $X_{\mathcal{S}}$ 的线性函数（$Y - \mathbb{E}[Y|X_{\mathcal{S}}]$ $\perp\!\!\!\perp$ $X_{\mathcal{S}}$ 对所有 $\mathcal{S} \subseteq \{1:\ n\}$均成立。然而，定理 5.7不需要这些假设。此外，定理 5.7 保证对一个多元分布

函数，而非 [36,定理 5.2]中的单变量收敛陈述。

5.5. 证明 定理 5.7的简要介绍。 虽然我们将证明定理 5.7的技术细节留给附录 A.4，但我们可以总结使证明得以成立的关键中间结果。我们所需的主要结果围绕各种估计量的一致性展开，其中 m 和/或 $N_{\mathcal{S}}$ 趋于无穷大。以下两组结果利用这些假设，得出算法中间计算的一致性。

第一组结果表明，在探索阶段计算的各量的有限样本估计量是一致的，即当探索样本数 m 趋于无穷大时成立。

引理 5.8（探索估计量的渐近一致性）。 对于所有 $\mathcal{S} \subseteq \{1$，我们有以下技术估计和一致性结果 $n\}$:

(*i*) 在假设 *5.3* 和 *5.6*下，以概率 *1*，

$$\sup_{\|\boldsymbol{A}-\boldsymbol{B}_{\mathcal{S}^+}\|_F<\varepsilon} \sup_{\boldsymbol{x}\in\mathbb{R}^d} |F_{V(\boldsymbol{A})}(\boldsymbol{x}) - F_{V(\boldsymbol{B}_{\mathcal{S}^+})}(\boldsymbol{x})| \lesssim \|\boldsymbol{A}-\boldsymbol{B}_{\mathcal{S}^+}\|_F^{2/3} \lesssim \varepsilon^{2/3} \tag{5.15a}$$

$$\sup_{\|\boldsymbol{A}-\boldsymbol{B}_{\mathcal{S}^+}\|_F<\varepsilon} \sup_{\boldsymbol{x}\in\mathbb{R}^d} |F_{V(\boldsymbol{A})\vee Y}(\boldsymbol{x}) - F_{V(\boldsymbol{B}_{\mathcal{S}^+})\vee Y}(\boldsymbol{x})| \lesssim \|\boldsymbol{A}-\boldsymbol{B}_{\mathcal{S}^+}\|_F^{2/3} \lesssim \varepsilon^{2/3}. \tag{5.15b}$$

(*ii*) 在假设 *5.3* 和 *5.4*下，以概率 *1*，

$$\lim_{m\to\infty} \widehat{\boldsymbol{B}}_{\mathcal{S}^+} = \boldsymbol{B}_{\mathcal{S}^+}.$$

(*iii*) 在假设*5.3*、 *5.4*和 *5.6*下，以概率 *1*，

$$\lim_{m\to\infty} \sup_{\boldsymbol{x}\in\mathbb{R}^d} |\widehat{F}_{\widehat{H}_{\mathcal{S}}}(\boldsymbol{x}) - F_{H_{\mathcal{S}}}(\boldsymbol{x})| = 0 \qquad \lim_{m\to\infty} \sup_{\boldsymbol{x}\in\mathbb{R}^d} |\widehat{F}_{Y\vee\widehat{H}_{\mathcal{S}}}(\boldsymbol{x}) - F_{Y\vee H_{\mathcal{S}}}(\boldsymbol{x})| = 0.$$

(*iv*) 在假设 *5.3*、 *5.4*和 *5.6*下，当 $m \to \infty$ 几乎处处有

$$\mathcal{K}_1(\boldsymbol{x}) \to (1-\rho_{\mathcal{S}}^2(\boldsymbol{x}))F_Y(\boldsymbol{x})(1-F_Y(\boldsymbol{x})) \quad \mathcal{K}_2(\boldsymbol{x}) \to \rho_{\mathcal{S}}^2(\boldsymbol{x})F_Y(\boldsymbol{x})(1-F_Y(\boldsymbol{x}))$$

对所有 $\boldsymbol{x} \in \mathbb{R}^d$.^(*v*)，在假设 *5.3*、*5.4*、*5.5* 和 *5.6* 下，以概率 *1*，$\lim_{m\to\infty} k_1(\mathcal{S})$ ^ $= k_1(\mathcal{S})$ 且 $\lim_{m\to\infty} k_2(\mathcal{S}) = k_2(\mathcal{S})$。 (*vi*) 在假设 *5.3*、*5.4* 和 *5.6* 下，对于 $\boldsymbol{x} \in (\mathrm{supp}(F_{H_{\mathcal{S}}}))^\circ$，$\widehat{\alpha}(\boldsymbol{x})$ 几乎处处是 $\alpha(\boldsymbol{x})$ 的一致估计量，即对于每个 $\boldsymbol{x} \in (\mathrm{supp}(F_{H_{\mathcal{S}}}))^\circ$，$\lim_{m\to\infty} \widehat{\alpha}(\boldsymbol{x}) = \alpha(\boldsymbol{x})$。

证明见附录 A.2。

注记 5.9。 请注意，$\widehat{\alpha}(\boldsymbol{x})$ 在$(supp(F_{H_{\mathcal{S}}}))^\circ$之外可能不一致，其中 $\alpha(\boldsymbol{x})$ 的值在定义中为方便起见设为零；见(4.3)。然而，这不会影响利用估计量的准确度，因为 $\mathbf{1}_{\{H_{\mathcal{S}}\le\boldsymbol{x}\}}$ 在该区域内为常数。

我们的第二个中间结果表明，Y的累积分布函数的利用估计量随着探索样本数 m 和利用样本数 $N_{\mathcal{S}}$的增加而渐近一致。

引理 5.10（利用阶段 CDF 估计器的一致渐近一致性）。在假设 5.3、5.4和 5.6下，几乎必然有 $\sup_{\boldsymbol{x}\in\mathbb{R}^d} |F_{\mathcal{S}}(\boldsymbol{x}) - F_Y(\boldsymbol{x})| \to 0$ 当$m,\ N_{\mathcal{S}} \to \infty$。

参见 附录 A.3中的证明。主要结果 定理 5.7的证明见附录 A.4，其中利用了 引理 5.8 和引理 5.10中的结果。证明 定理 5.7 所需的另一个高层次步骤，是证明当预算 B 渐近增大时，m 和 $N_{\mathcal{S}}$ 均趋于无穷大。这是附录 A.4中所给证明的第一部分。

6. 数值模拟。 本节将在三个正向不确定性量化场景中，将 cvMDL 及其变体与其他算法进行比较。在第 6.1节中，我们考察一个标量参数线性弹性偏微分方程问题；随后在第 6.2节中，研究一个关于有限区间上几何布朗运动极值的向量值随机微分方程问题。最后，在第 6.3 节中，我们在一个纤维增强基体中脆性断裂的标量实际工程问题上评估 cvMDL 方法。我们将所考虑的算法标记如下：

（ECDF）仅使用 F_Y 高保真样本的经验 CDF 估计器；（AETC-d）来自 [36]的 AETC-d 算法；（cvMDL）算法 5.2中的 cvMDL；（cvMDL-sorted）使用算法 5.1进行利用 CDF 单调性修正的 cvMDL；（cvMDL*）在尾部使用（ $\widehat{\alpha}(\boldsymbol{x})$ 通过（5.11）估计 $\tau = 0.05$ ，当$d = 1$时；（cvMDL*-sorted）使用算法 5.1进行 CDF 单调性修正的 cvMDL*。

对于 cvMDL 算法及其变体中的权重函数，当 Y 为标量值时，我们对所有 $x \in \mathbb{R}$ 选择 $\omega(x) \equiv 1$；而当 Y 为向量值时，则根据具体情况选择。由于 cvMDL 类型和 AETC-d 算法生成的估计量是随机的（取决于探索数据），因此，对于每个预算值 B，我们重复实验 100 次，并报告均值 ω 加权 L^2 误差的平均值，以及相应的 5%-95% 分位数，以衡量不确定性。

6.1. 线性弹性。 我们考虑一组具有不同保真度的模型，这些模型与一个参数椭圆方程相关，其中低保真模型通过网格粗化确定。该设置取自 [35, 第 7 节1]。椭圆型偏微分方程描述了方形空间域 $\mathcal{D} = [0,1]^2$上的线性弹性位移，其中 $\boldsymbol{x} = (x_1, x_2)^\top$；参见 图 2。该问题的参数化方程旨在求得满足以下偏微分方程的位移场 $\boldsymbol{u} = (u, v)^\top$：

$$-\nabla \cdot (\kappa(\boldsymbol{p}, \boldsymbol{x})\, \boldsymbol{\sigma}(\boldsymbol{x})) = \boldsymbol{F}(\boldsymbol{x}), \quad \forall (\boldsymbol{p}, \boldsymbol{x}) \in \mathcal{P} \times \mathcal{D}$$

其中，$\boldsymbol{p}$ 是用于参数化标量 κ的随机输入向量。此外，$\boldsymbol{\sigma}$ 是柯西应力张量，表示为

$$\boldsymbol{\sigma}(\boldsymbol{x}) = \begin{bmatrix} \sigma_1(\boldsymbol{x}) & \sigma_{12}(\boldsymbol{x}) \\ \sigma_{12}(\boldsymbol{x}) & \sigma_2(\boldsymbol{x}) \end{bmatrix}, \qquad \begin{bmatrix} \sigma_1(\boldsymbol{x}) \\ \sigma_2(\boldsymbol{x}) \\ \sigma_{12}(\boldsymbol{x}) \end{bmatrix} = \frac{1}{1-\nu^2} \begin{bmatrix} \frac{\partial u(\boldsymbol{x})}{\partial x_1} + \frac{\partial v(\boldsymbol{x})}{\partial x_2} \\ \frac{\partial v(\boldsymbol{x})}{\partial x_2} + \nu \frac{\partial u(\boldsymbol{x})}{\partial x_1} \\ \frac{1-\nu}{2}\left(\frac{\partial u(\boldsymbol{x})}{\partial x_1} + \frac{\partial v(\boldsymbol{x})}{\partial x_2}\right) \end{bmatrix}$$

其中，我们将泊松比设为 $\nu = 0.3$。这里， $\kappa(\boldsymbol{p}, \boldsymbol{x})$是一个采用具有四个模态的 Karhunen- Lo` eve 展开进行建模的标量，表示为

$$\kappa(\boldsymbol{p}, \boldsymbol{x}) = 1 + 0.5 \sum_{i=1}^{4} \sqrt{\lambda_i} \phi_i(\boldsymbol{x}) p_i,$$

其中，(λ_i, ϕ_i) 是指数协方差核 K 在 $\mathcal{D}$上的有序特征对，即：

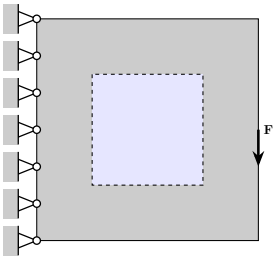
$$K(\boldsymbol{x}, \boldsymbol{y}) = \exp(-\|\boldsymbol{x} - \boldsymbol{y}\|_1 / \eta),$$

其中， $\|\cdot\|_1$ 是向量的 ℓ^1范数，我们取 $\eta = 0.7$。因此， $\boldsymbol{p} \in \mathbb{R}^4$ 是一个随机输入向量，其独立分量在 $[-1, 1]$上服从均匀分布。

位移 $\boldsymbol{u}$ 用于计算标量感兴趣量 QoI，即解的结构柔度 或能量范数，它表示结构因载荷作用而吸收的弹性能测度：

$$E(\boldsymbol{u}; \boldsymbol{p}) \coloneqq \int_D (\boldsymbol{u}(\boldsymbol{x}) \cdot \boldsymbol{F}(\boldsymbol{x}))\, \mathrm{d}\boldsymbol{x}.$$

我们针对每个固定的 $\boldsymbol{p}$ ，通过有限元方法求解上述系统，采用矩形网格 [2]上的标准双线性方形各向同性有限元。



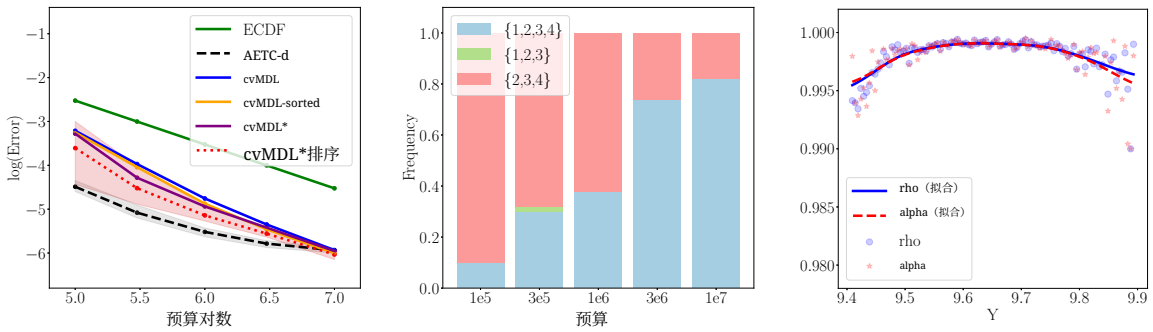
模型 $\mathcal{S}$	$\{1\}$	$\{2\}$	$\{3\}$	$\{4\}$	$\{1,2\}$	$\{1,3\}$	$\{1,4\}$	$\{2,3\}$
$\gamma_{\mathcal{S}}$	123.8	149.3	203.9	304.8	25.2	48.6	93.7	62.2
$m^*_{\mathcal{S}}$ 当 $B = 10^7$	1998	2231	2337	2390	1253	1657	1909	2054

模型 $\mathcal{S}$	$\{2,4\}$	$\{3,4\}$	$\{1,2,3\}$	$\{1,2,4\}$	$\{1,3,4\}$	$\{2,3,4\}$	**{1,2,3,4}**
$\gamma_{\mathcal{S}}$	107.6	129.7	11.8	11.7	14.3	11.9	**11.5**
$m^*_{\mathcal{S}}$ 当 $B = 10^7$	2175	2292	669	734	976	1540	**638**

图 2： 线性弹性。左：线性弹性结构的几何形状和边界条件。右：不同低保真模型索引子集的选择对应的预言机缩放损失 $\gamma_{\mathcal{S}}$ （3.11）和最优探索样本数 $m^*_{\mathcal{S}}$ （3.11） $\mathcal{S}$。最优模型子集 $\mathcal{S}$ 以粗体显示。预言机统计量使用 50,000 个样本计算。

在此示例中，我们通过使用网格参数 h进行网格粗化来构建多保真层次结构。柔度 E 是每个模型的标量值 QoI，即，对于 $i = 1, \ldots, 4$， $d = d_i = 1$。网格参数 $h = 2^{-7}$ 对应于高保真模型 Y，而$h = 2^{-4}$、 2^{-3}、 2^{-2}、 2^{-1} 分别对应于低保真模型 $X_1, \ldots, X_4$。

每个模型的成本是计算时间，我们将其视为与网格尺寸的平方成反比，即 h^2。这相当于使用具有最优线性复杂度的线性求解器。我们将成本归一化，使保真度最低的模型成本为单位成本，即 $(c_0$、 c_1、 c_2、 c_3、 $c_4) = (4096$、64、16、4、$1)$。 Y与X_1、 X_2、 X_3、 X_4 的感兴趣量之间的相关性分别为 0。976、0。940、0。841 和 -0.146。总预算 B 取值于区间 $[10^5, 10^7]$。



(a) (b) (c)

图3：线性弹性。（a）在1%-99%分位数范围内，使用ECDF、AETC-d、cvMDL及其变体得到的估计累积分布函数与 F_Y 之间的均值 $\omega(x)$加权 L^2 误差，以及5%-95%分位数（为便于可视化，我们仅绘制AETC-d和cvMDL*排序的分位数），以衡量不确定性。（b）cvMDL选取的不同模型的频数。（c）使用50,000个独立同分布样本时估计的 $\alpha(x)$和 $\rho(x)$的散点图，其中 $\mathcal{S} = \{1, 2, 3, 4\}$ Y的1%-99%分位数范围。对两组数据均应用高斯核平滑，并使用五折交叉验证选择带宽bd $= 0.0358$ 。

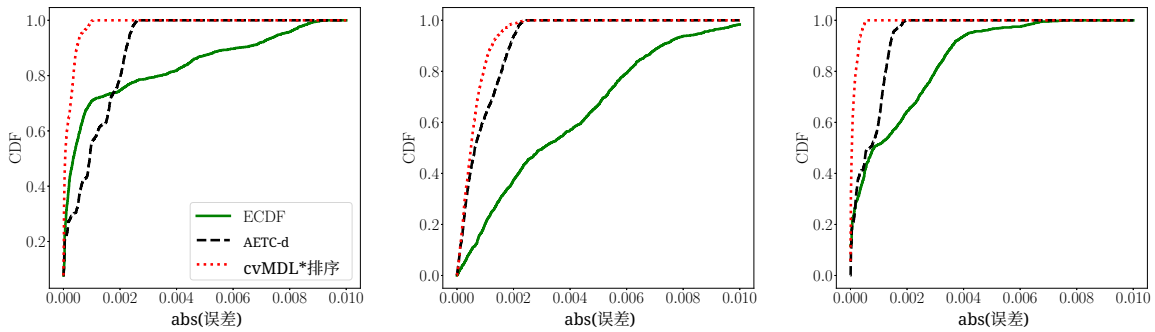
6.1.1. 用于估计累积分布函数的结果。 图 3 展示了不同多保真估计量给出的仿真结果，以及与 cvMDL 相关算法更细致的统计量。在 图 3（a）中，我们看到，AETC-d 在较小预算下的误差最小，但其渐近收敛受到模型设定错误效应的限制（这与关于 AETC-d 适用性的理论假设有关），即当 B 超过 10^6时，误差曲线开始趋于平台。尽管可以通过加入额外的非线性项（例如多项式项）作为额外协变量来缓解这一问题，但目前对于这种方法仍缺乏可靠的实践指导。另一方面，cvMDL 及其变体的表现均优于 ECDF，其中 cvMDL*排序 的结果具有与 AETC-d 相当的竞争力，且没有出现平台效应。

在图 3（b）中，我们注意到，随着预算增加，模型 $\mathcal{S}^*$ 被 cvMDL 选定并收敛到 $\{1, 2, 3, 4\}$，这与图2 （右）中根据 Oracle 统计量计算出的最优模型相同。我们注意到，次优模型 $\mathcal{S} = \{2, 3, 4\}$ 经常被 cvMDL 选中，但其他模型并不会，即使其他模型的 $\gamma_{\mathcal{S}}$ 与 $\{1, 2, 3, 4\}$接近。我们认为，这种情况是由于 cvMDL 所采取的激进探索步骤造成的，尤其是在我们将探索样本数量加倍时，会使 $\mathcal{S}$ 具有较大值的次优模型 $m^*_{\mathcal{S}}$ （例如，$\mathcal{S} = \{2, 3, 4\}$）成为首选模型。

cvMDL 所实现的显著误差降低，可由接近 1 的数值体现$\rho_{\mathcal{S}}(x)$ $=$相关系数$[\mathbf{1}_{\{Y \le x\}}, h(X_{\mathcal{S}};\ x)]$其中 $\mathcal{S} = \{1, 2, 3, 4\}$，如图 3(c) 所示。对于 cvMDL 的变体，通过（cvMDL*）估计尾部区域中的 $\alpha(x)$，具体见（5.8），或对 CDF 值进行排序以确保单调性（cvMDL/cvMDL*-排序），均有助于进一步降低误差。前一种方法在预算较小的区域尤其有帮助，因为此时探索数据不足以

估计感兴趣量的完整支集。

权重函数 $\omega(x)$ 在此情形下在 $\mathbb{R}$ 上为常数，因此预计 cvMDL 型估计器产生的估计结果能够捕捉到 F_Y（例如主体区域和尾部）的全局结构。为检查这一点，我们固定 $B = 10^7$，并分别绘制尾部和主体区域中的估计的累积分布函数。三个区域中逐点误差（在 1000 个离散化点处）的累积分布函数如图4所示。可以看出，cvMDL*排序在三个区域中的误差均最小。



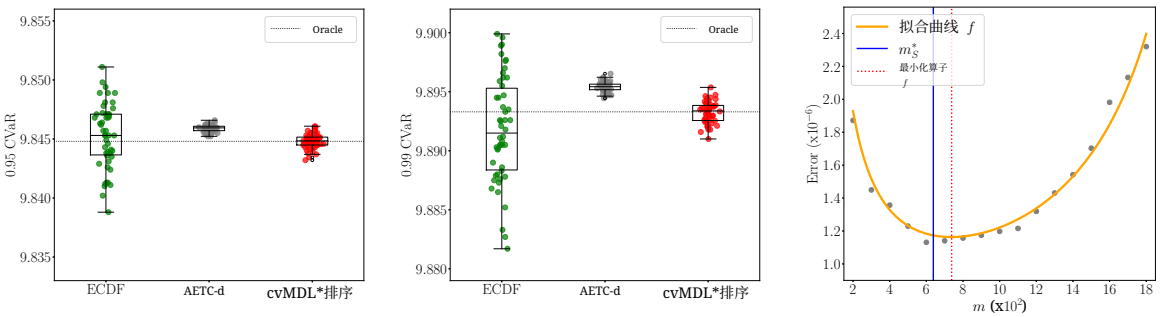
(a) (b) (c)

图 4：线性弹性。对于预算 $B = 10^7$，由 cvMDL*排序、AETC-d 和 ECDF 计算得到的逐点累积分布函数误差分布的一次实现。我们绘制了三个不同区域中误差的累积分布函数：（a）由 $0 - 0.05$ 分位数区域定义的 Y 下尾部；（b）由 $0.05 - 0.95$ 分位数区域定义的主体区域；（c）由 $0.95 - 1.00$ 分位数区域定义的上尾部。

6.1.2. 风险度量。 现在，我们比较估计的累积分布函数的一些风险度量。例如，一个常用的风险度量是条件风险价值（CVaR），也称为预期短缺，它被定义为 Y 在尾部区域中的条件期望（此处，Y 表示较大）：

$$\mathsf{CVaR}_a(Y) := \mathbb{E}[Y | F_Y(Y) \geq a] = \frac{1}{1-a} \int_a^1 F_Y^{-1}(x) \mathrm{d}x \qquad 0 < a < 1,$$

其中 a 是分位数水平。假设 F_Y 已知，条件风险价值可以通过对 F_Y^{-1} 进行线性插值来数值计算。固定 $B = 10^7$ 与之前相同，我们使用 ECDF、AETC-d 和 cvMDL*排序得到的估计的累积分布函数，计算 Y 在 $a = 0.95$ 和 0.99 分位数水平下的条件风险价值。实验重复 50 次，相应的统计量在图5 （a）-（b）中使用箱线图进行了汇总。对于 a的两种取值，cvMDL*排序均以明显优势优于其他方法。值得注意的是，尽管在测试预算下，AETC-d 和 cvMDL*排序的总体误差相近（图3 （a）），模型误设效应导致前者的估计向上偏置。另一方面，cvMDL*排序的估计保持无偏。



(a) (b) (c)

图5：线性弹性。（a）：使用 ECDF、AETC-d 和 cvMDL*排序 给出的估计的累积分布函数计算的 $\mathrm{CVaR}_{0.95}(Y)$ 的箱线图，其中 $B = 10^7$，共进行 50 次实验。（b）：$\mathrm{CVaR}_{0.99}(Y)$ 的结果相同。（c）：检查估计的损失 $L_{\mathcal{S}}$ 如何随 m 的变化逼近理想损失曲线。使用函数 $f(m; a, b)$ 对离散数据进行拟合，其形式为 $\frac{a}{m} + \frac{b}{B - c_{\mathrm{epr}} m}$，其中 a 和 b 的拟合值分别由 2.64×10^{-4} 和 5.57 给出。

6.1.3. Oracle 与估计损失。 我们研究 cvMDL 中使用的模型选择标准。注意，对于每个 $\mathcal{S} \subseteq \{1 : n\}$，精确的损失函数 $\widehat{\ }$ 与使用经验数据构造的估计量 $L_{\mathcal{S}}$ 之间存在差异。我们考察这种近似是否合理。为了数值确定探索-利用权衡是否最优，我们固定 $B = 10^7$ 和 $\mathcal{S} = \{1, 2, 3, 4\}$。对于给定的 m值，我们首先取得 m 探索样本，用于 $\widetilde{\ }$ 估计控制变量参数，然后使用这些参数构造估计量 $F_{\mathcal{S}}$，用于 F_Y，如（5.10）所示。随后，我们计算与该 L^2 值相关的（精确）均值加权 m损失。实验重复 10 次，并计算平均损失值。我们汇总了上述实验中 m 取值范围为 200 至 1800 时的结果。结果如图5（c）所示。可以看到，Oracle 损失下的最优探索率 $L_{\mathcal{S}}$为 638（见图 2右），与经验确定的最优探索率 736 几乎一致。这一小幅差距可归因于有限样本估计参数时对探索误差的低估。

6.2. 几何布朗运动的极值。 几何布朗运动是一种连续时间随机过程，广泛应用于金融建模。在简单情形下，几何布朗运动 S_t 是一个具有恒定初始状态 $s_0 > 0$ 的随机过程，其演化由随机微分方程描述

$$dS_t = \mu S_t dt + \sigma S_t dW_t \qquad t \geq 0, \qquad S_0 = s_0,$$

其中 μ 和 $\sigma > 0$ 为常数，W_t 是标准布朗运动过程。S_t 存在唯一的显式解，可写为

$$S_t = s_0 \exp\left(\left(\mu - \frac{\sigma^2}{2}\right) t + \sigma W_t\right).$$

设 $\mu = 0.05$, $\sigma = 0.2$, $s_0 = 1$。我们感兴趣的是估计 S_t 在时间区间 $[0,1]$上的极值的联合分布:

$$(S_{\min}, S_{\max})^\top \in \mathbb{R}^2 \qquad S_{\min} := \min_{0\le t\le 1} S_t, \quad S_{\max} := \max_{0\le t\le 1} S_t.$$

因此，我们将随机向量（$S_{\min}$, $S_{\max}$）$^\top$作为感兴趣量。我们通过使用 Euler–Maruyama格式以时间步长 Δt 对随机微分方程进行时间离散化，并计算离散极值，来评估这些量。相应过程的计算复杂度（成本）与用于离散化的网格点数量成正比。

我们基于时间离散化为该问题构建了一个多保真模型。具体而言，我们考虑了四种不同的时间尺度 $\Delta t \in \{2^{-4}, 2^{-6}, 2^{-8}, 2^{-14}\}$，其中高保真模型 Y对应于 $\Delta t = 2^{-14}$，而 X_1、X_2、X_3分别对应于 $\Delta t = 2^{-8}$、2^{-6}、2^{-4}。由此得到（归一化的）模型成本（c_0、c_1、c_2、c_3）=（1024、16、4、1）。总预算 B 取自 $[10^4, 10^6]$。为了生成联合样本，W_t 的随机性使用高保真模型中所采用的同一随机实现进行模拟。高保真模型的基准累积分布函数使用包含 10^5 个样本的蒙特卡罗方法计算。表1给出了高保真模型与低保真度模型的感兴趣量之间的先知相关性。

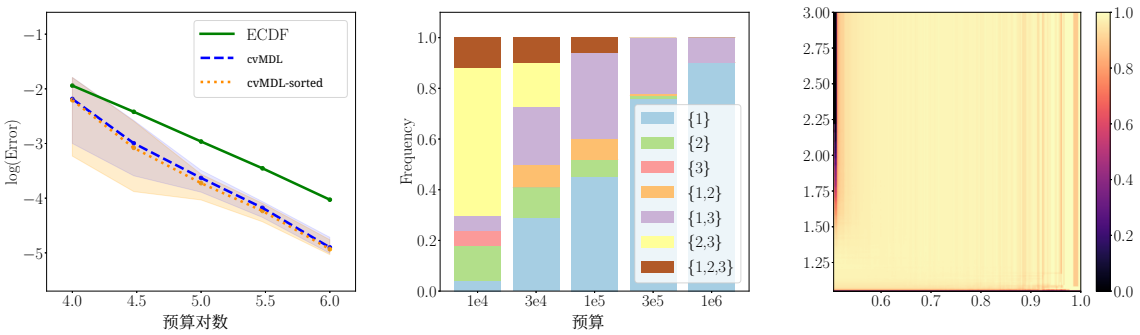
表1：几何布朗运动。使用50,000个样本计算得到的高保真与低保真模型感兴趣量之间的先知相关性。

模型感兴趣量	$S_{\min}(2^{-8})$	$S_{\max}(2^{-8})$	$S_{\min}(2^{-6})$	$S_{\max}(2^{-6})$	$S_{\min}(2^{-4})$	$S_{\max}(2^{-4})$
$S_{\min}(2^{-14})$	0.999	0.682	0.997	0.682	0.984	0.680
$S_{\max}(2^{-14})$	0.681	0.999	0.681	0.998	0.674	0.988

在本例中，所有模型的感兴趣量都是二维随机向量，因此无法直接应用 AETC-d。对于 cvMDL 及其变体，设置 $\omega(\boldsymbol{x}) \equiv 1$ 违反了假设5.5。相反，由于 $S_{\min} \le s_0 = 1 \le S_{\max}$，我们选择 $\omega(\boldsymbol{x}) = \mathbf{1}_{\mathcal{T}}(\boldsymbol{x})$ 作为二维有界区域上的示性函数 $\mathcal{T} \subset \mathbb{R}^2$，最可能的结果位于该区域内。例如，这里我们取 $\mathcal{T} = [0.5, 1] \times [1, 3]$。估计误差统计量以及 cvMDL 选出的模型如图6（a）、（b）所示。面板（c）显示，相关系数$\rho_{\mathcal{S}}(\boldsymbol{x})$在整个区域内都接近于 1，这表明我们所选的控制变量（3.6）是一个不错的选择。

图6 显示，cvMDL 在区域 $\mathcal{T}$上一致，其对应的估计误差平均而言远低于 ECDF。随着预算趋于无穷，cvMDL 选出的模型收敛于单个低保真模型 $\{1\}$，这与使用 Oracle 统计量计算得到的最优模型一致。通过进一步排序来稳定算法，cvMDL-sorted 使 cvMDL 的误差略有进一步降低，这与 1d 情况下的观察结果一致。在预算较小的前渐近阶段，cvMDL 选出的模型存在相对较大的波动，但随着预算增加，这些波动会趋于稳定。该实验的更多结果见附录B.1。

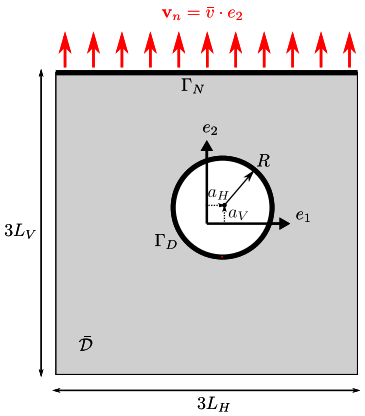
6.3. 纤维增强基体的脆性断裂行为。我们研究一种二维纤维增强基体，这是断裂力学领域中常见的研究对象。我们的



(a) (b) (c)

图6：几何布朗运动。（a）：均值 $\omega(\boldsymbol{x})$加权的 L^2 误差，介于 F_Y 和由经验累积分布函数、cvMDL 及 cvMDL-sorted 给出的估计的累积分布函数之间，并使用 5%-50%-95% 分位数来测量不确定性。（b）：由 cvMDL 选择的不同模型的频数；最优模型损失见图 11。（c）由 $\rho_{\mathcal{S}}(\boldsymbol{x})$根据3.6)估计的 $\mathcal{S} = \{1\}$，其中 $\boldsymbol{x} \in \mathcal{T}$使用了 50,000 个独立同分布样本。

感兴趣量是使纤维相邻基体区域发生脆性断裂的最大载荷。为获得感兴趣量，我们求解一个准静态二维有限元问题。

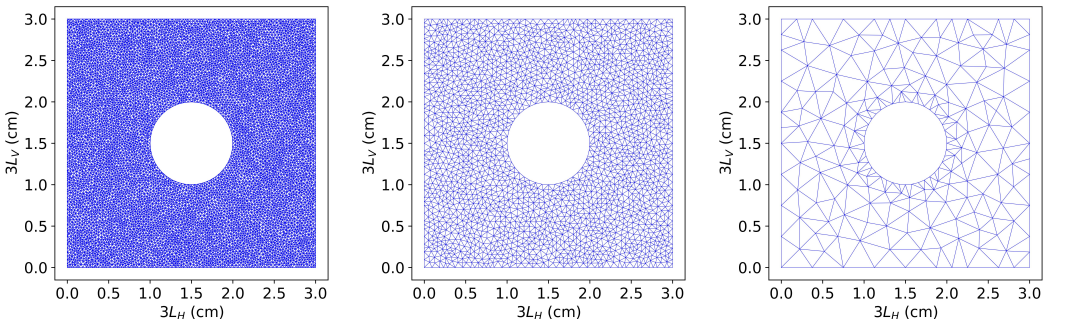


随机变量	属性	Mean	变异系数$_i$ (%)[a]	下限边界	上限边界	概率分布
p_1	κ (GPa)	55	5	45	60	截断正态分布
p_2	ν	0.25	5	0.20	0.30	截断正态分布
p_3	G_c (GPa)	1	10	0	∞	对数正态分布
p_4	L_H(cm)	1	–	0.9	1.1	均匀
p_5	L_V (cm)	1	–	0.9	1.1	均匀
p_6	R (厘米)	0.5	–	0.4	0.6	均匀
p_7	a_H (厘米)	0	–	−0.05	0.05	均匀
p_8	a_V (厘米)	0	–	−0.05	0.05	均匀

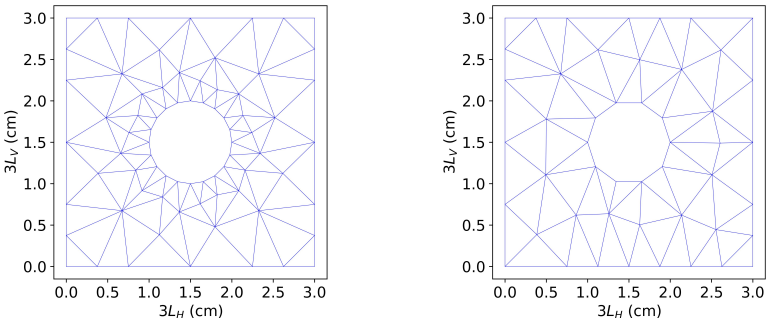
a. 变异系数 $\text{COV}_i = 100 \times \sigma_i / \mathbb{E}[p_i]$，对于 $i = 1$、.....、8。

图 7：纤维增强基体。左：几何形状、载荷和边界条件。我们考虑区域 $\mathcal{D} = [-1.5L_H, 1.5L_H] \times [-1.5L_V, 1.5L_V] \in \mathbb{R}^2$，其中包含一个半径为 R、位于（a_H，a_V）处的圆孔，该圆孔位于中心线的 e_1和 e_2方向。右：纤维增强基体中八个随机输入的属性。其中，κ为杨氏模量，ν 为泊松比，详见附录 B.2。

图 7 （左）展示了一块长度为 $3L_H$ 、沿 e_1 方向以及沿 $3L_V$ 方向的方板，其中包含一个半径为 R 的圆形夹杂。在区域 $\mathcal{D} = [-1.5L_H, 1.5L_H] \times [-1.5L_V, 1.5L_V] \in \mathbb{R}^2$ 中，载荷由施加的法向位移 $\mathbf{v}_n = \bar{v} \cdot e_2$



(a) Fine mesh (53,766 DOFs) (b) Coarse mesh (13,416 DOFs) (c) Coarse mesh (975 DOFs)



(d) Coarse mesh (324 DOFs) (e) Coarse mesh (195 DOFs)

图8：纤维增强基体。(a) 中的精细有限元网格用于生成高保真感兴趣量 Y，而 (b)–(e) 中的粗网格分别用于生成低保真感兴趣量 X_1，X_2，X_3，$X4$。

在边界 Γ_N上给定。其他边界记为 Γ_D，为自由边界，对应于在 Γ_D上法向位移为零。区域的闭包为 $\mathcal{D} \equiv \mathcal{D} \cup \Gamma$，其中有 $\Gamma = \Gamma_D \cup \Gamma_N$。未知量为弹性体区域 $\mathcal{D}$ 中的位移场 $\mathbf{u} = (u, v)^\top \in \mathbb{R}^2$ 以及标量损伤变量 $\phi_d \in \mathbb{R}$ 。该设置对纤维增强基体的拉伸实验 [1, 6]，进行建模，相应的边值问题如 [25]所述。偏微分方程表述为：求 $\mathbf{u}(\mathbf{x})$ 和 $\phi_d(\mathbf{x})$，其中 $\mathbf{x} = (x_1, x_2)^\top \in \bar{\mathcal{D}}$，使得

$$[(1 - \phi_d(\mathbf{x}))^2 + q]\nabla \cdot \boldsymbol{\sigma}(\mathbf{x}) = \mathbf{0}, \tag{6.1}$$

$$-G_c \ell_0 \nabla^2 \phi_d(\mathbf{x}) + \left[\frac{G_c}{\ell_0} + 2H(\mathbf{x})\right] \phi_d(\mathbf{x}) = 2H(\mathbf{x}), \tag{6.2}$$

并在 Γ_N 和 Γ_D 上施加相应的边界条件。完整模型的详细信息，包括 ℓ_0、$\boldsymbol{\sigma}$、q和 H 的定义，见 附录 B.2。我们考虑八个输入随机变量 p_1、$\ldots p_8$，这些变量源自材料性质和几何形状，见图 7(右)。

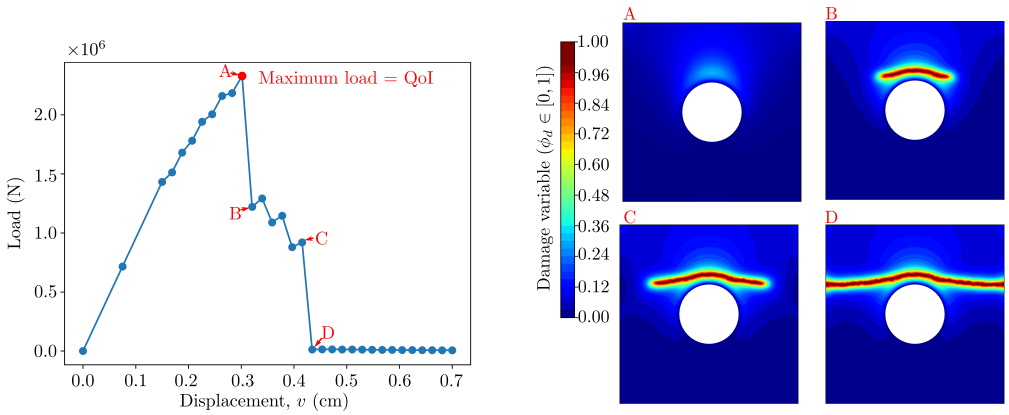
6.3.1. 高保真模型和低保真模型。用于脆性断裂分析的模型（6.1）和（6.2）采用迭代求解器求解，其中使用位移场（ϕ_d）求解标量损伤变量（$\mathbf{u}$）。随后，使用更新后的损伤变量

来求解位移场，并重复这一过程，直到当前迭代值与前一迭代值之差小于用户定义的容差 $\delta \ll 1$。我们为$\delta = 5 \times 10^{-3}$高保真模型设置该值，而为低保真模型设置 $\delta = 5 \times 10^{-2}$，0。2、0。4。图8 分别展示了高保真模型和低保真模型所用的细网格和若干粗网格。高保真模型和低保真模型的详细信息，包括实施有限元分析所需的 CPU 时间，列于表2中，表中还报告了（归一化的）模型成本。高保真模型的感兴趣量（Y）与其低保真感兴趣量 X_1、 X_2、 X_3、X_4 之间的先知相关性分别为 0。96、0。93、0。87 和 0。74。

表2：纤维增强基体。用于计算感兴趣量的高保真和四种不同低保真有限元模型的比较。

模型类型	容差（δ）	DOFs	CPU 时间（秒）[a]	归一化成本[b]
高保真， Y	5×10^{-3}	53，766	250.51	108.9
低保真度 1， X_1	5×10^{-2}	13，416	20.95	9.1
低保真度 2， X_2	0.2	975	2.97	1.3
低保真度 3， X_3	0.2	324	2.50	1.1
低保真度 4， X_4	0.4	195	2.30	1

a. CPU 时间取 5 次试验的平均值。b. 成本经过归一化处理，使采样 X_4 的成本为单位成本。



(a) 载荷-位移曲线 (b) 损伤变量等值线

图 9：纤维增强基体。有限元分析结果：(a) 将载荷-位移曲线中的极限拉伸载荷记录为感兴趣量。(b) 损伤变量等值线显示了载荷-位移曲线中“A”–“D”阶段发生的损伤程度（$0 < \phi_d \leq 1$），表明在完全断裂发生于“D”阶段之前，圆孔顶部在“A”至“B”阶段发生了脆性断裂。

We 从载荷-位移曲线中测量最大拉伸载荷，作为感兴趣量。图 -

示9a展示了纤维增强基体顶部的结果载荷与施加位移之间的关系。当顶部的施加位移 $\bar{v}$ 从 0 cm 增加至 7.5×10^{-2} cm 时，结果载荷几乎呈线性增加，直到结构开始出现损伤。在达到峰值载荷后，结果载荷相对于位移的变化率显著降低。从图9a中的阶段“A”到“B”可以观察到这一行为。这些阶段是由于基体发生部分断裂而出现的，如图 $\phi_d = 1$ 中的损伤变量值 $\phi_d = 1$ 9b所示。从阶段“C”到“D”，载荷还会出现另一次显著下降，表明整个基体区域发生了完全断裂。“A”处的最大拉伸载荷表示为一个标量；因此，对于高保真QoI，我们有 $d = 1$，而对于低保真QoIs，我们有 $d_1 = d_2 = d_3 = d_4 = 1$。

表3：纤维增强基体。在 $\beta = 0.99$ 处，比较 cvMDL*排序和经验累积分布函数（ECDF）在估计感兴趣量（QoI，即极限拉伸载荷）的累积分布函数、均值、标准差和条件风险价值（CVaR）方面的准确度。我们给出了这些估计值相对于 Oracle 估计值的平均误差；Oracle 估计值由 5500 个独立同分布的高保真样本（$B = 577,170$）计算得出，统计基于 100 次试验。

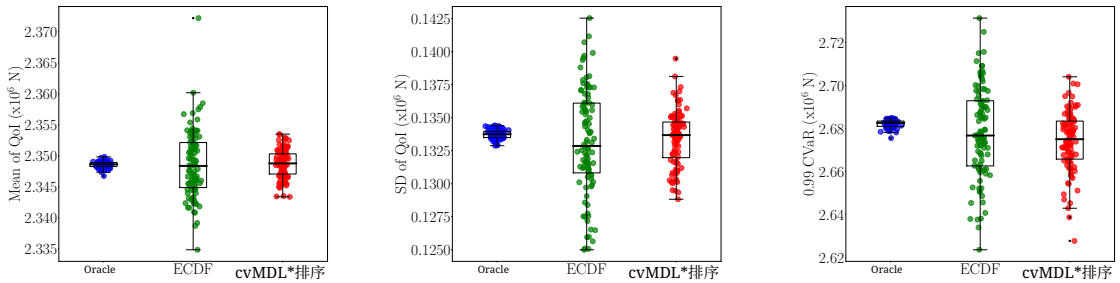
方法	累积分布函数（误差）[a]	均值（-）[b]	标准差（-）[b]	条件风险价值$_{0.99}$（-）[b]
		预算 $B = 20{,}000$		
cvMDL*排序	2.192×10^{-4}	1.559×10^{-3}	2.291×10^{-2}	6.698×10^{-3}
ECDF	3.916×10^{-4}	2.991×10^{-3}	3.837×10^{-2}	9.435×10^{-3}
		预算 $B = 35{,}000$		
cvMDL*排序	1.194×10^{-4}	1.168×10^{-3}	1.575×10^{-2}	5.531×10^{-3}
ECDF	2.269×10^{-4}	2.311×10^{-3}	2.955×10^{-2}	7.577×10^{-3}
		预算 $B = 50{,}000$		
cvMDL*排序	8.441×10^{-5}	8.957×10^{-4}	1.340×10^{-2}	4.846×10^{-3}
ECDF	1.627×10^{-4}	1.983×10^{-3}	2.670×10^{-2}	7.102×10^{-3}

a. 我们确定 cvMDL*排序和 ECDF 给出的估计的累积分布函数之间的均值 $\omega(x)$ 加权 L^2 误差。均值 $\omega(x)$ 加权 L^2 误差取独立的 100 次试验的平均值。b. 在比较先验最优解估计值时，我们使用独立的 100 次试验中估计值的均值相对误差。

6.3.2. 累积分布函数、均值、标准差和条件风险价值估计的结果。这里的高保真模拟成本足够高，因此我们必须使用有限的样本近似先验最优解：生成 6000 次高保真模拟，并随机选择其中 5500 次来估计某个量。我们生成 100 个这样的实例组成的集合，并将其均值作为先验最优解。对于多保真过程，使用 6000 个高保真与低保真联合样本作为样本池，从中抽取模型样本。我们研究了三个预算值，如表3所示。在相应的 100 次试验中，cvMDL*排序主要选择模型子集 $\mathcal{S} = \{2, 4\}$（对于 3 个预算值，分别选择了 95、96 和 98 次），而较少选择模型子集 $\mathcal{S} = \{4\}$（分别选择了 5、4 和 2 次），这两个模型子集均来自模型集合 $\{1, 2, 3, 4\}$。该过程得到的最优探索样本数量均值分别为 $m^*_{\mathcal{S}} = 140$、245 和 350，对应于各个预算值。

In T能够3，就 CDF、均值 ,

标准差和条件风险价值方面， $a = 0.99$ 对于感兴趣量。表3 报告了加权均值 L^2 误差，针对 F_Y在 $Y \in [1.5 \times 10^6, 3 \times 10^6]$ （N）上的结果。该表的最后四列表明，与ECDF方法相比，所提出的cvMDL方法的准确度几乎提高了一倍，而且这种准确度的提升也延伸到了估计统计量和风险度量。



（a）均值 （b）标准差 (c) 条件风险价值 0.99

图10：纤维增强基体。（a）：当 $B = 50,000$次实验时，通过经验累积分布函数和cvMDL*排序方法计算的感兴趣量均值的箱线图。（b）：感兴趣量标准差的相同箱线图。（c）：条件风险价值的相同箱线图（a=0.99）。

图10a–10c 通过箱线图展示了在 $a = 0.99$ 处的统计均值、标准差和条件风险价值结果。与神谕结果相比，cvMDL*排序方法的准确度明显高于经验累积分布函数。箱线图表明，cvMDL*排序方法得到的统计估计值相比经验累积分布函数具有更小的离散范围，说明在此示例中，cvMDL*排序估计值的方差更小。

7. 结论。 我们开发了一个通用框架，用于在预算约束下高效估计感兴趣量的累积分布函数。为实现这一框架，我们基于线性代理模型构建了一组二元控制变量，并将其用于估计累积分布函数的自适应元算法（cvMDL）中。我们证明了，当预算趋于无穷大时，相应算法同时具备一致一致性和权衡最优性。

尽管所提出的框架建立在现有的多臂老虎机学习范式之上，但我们对探索与利用的处理方式不同于以往工作。具体而言，我们框架中采用的新方法产生了创新性的估计量，大幅降低了对底层模型假设的依赖。此外，该方法还支持处理不同类型的感兴趣量，包括向量值和标量值感兴趣量。据我们所知，我们的框架首次在预算约束下提供了一种稳健的多保真度累积分布函数估计量，能够同时处理异构模型集合和多值感兴趣量，同时无需预先获取跨模型统计量。

附录A。主要结果的证明。

A.1. 定理 5.2的证明。 将 $\boldsymbol{T}$中的各项按递增顺序排列：$z^{(1)} < \cdots < z^{(M)}$，其中 M是 M_i的每个维度 $\boldsymbol{T}$的乘积：$M = M_1 \cdots M_d$。在算法开始时，$z^{(1)}$ 的索引在每个方向上都严格递减。因此，$z^{(1)}$经过有限次迭代后到达 $\boldsymbol{T}$中索引为 (1，...，1) 的项，之后在后续迭代中保持不变。事实上，对于每个 $s < M$，假设$z^{(1)}$、...、$z^{(s)}$ 已到达其最终位置且不再发生变化，若算法尚未收敛，则 $z^{(s+1)}$的索引在每个方向上都在递减。注意到 M 是有限的，且驻点必然具有所需的单调性，结论即得。

A.2. 引理 5.8的证明。本节包含以下陈述的证明：（i）至（vi），它们位于引理 5.8中。陈述（ii）（即 $\boldsymbol{B}_{\mathcal{S}^+}$的渐近一致性）的证明直接由大数强律得到，因此我们省略该证明。

A.2.1. 陈述（i）的证明。我们只证明（5.15b），因为（5.15a）的证明类似。记 $\boldsymbol{e}_i$ i为 $\mathbb{R}^d$ 中的第 i个单位向量，即 $\boldsymbol{e}_d^{(j)} = \delta_{ij}$，$j \in \{1:\ d\}$，其中 δ 为克罗内克记号，而 $\boldsymbol{e}:=\sum_{i\in\{1:d\}} \boldsymbol{e}_i$ 为 $\mathbb{R}^d$中的全 1 向量。对于固定的 $\boldsymbol{x} \in \mathbb{R}^d$，不失一般性地，假设 $F_{V(\boldsymbol{A})\vee Y}(\boldsymbol{x}) \le F_{V(\boldsymbol{B}_{\mathcal{S}^+})\vee Y}(\boldsymbol{x})$，因为另一种情况通过反转 $F_{V(\boldsymbol{A})\vee Y}(\boldsymbol{x})$ 和 $F_{V(\boldsymbol{B}_{\mathcal{S}^+})\vee Y}(\boldsymbol{x})$ 即可类似处理。同时，

$$V(\boldsymbol{B}_{\mathcal{S}^+}) \vee Y \le \boldsymbol{x} \Rightarrow V(\boldsymbol{A}) \vee Y \le \boldsymbol{x} + \Delta\boldsymbol{x} \Rightarrow V(\boldsymbol{A}) \vee Y - \|\Delta\boldsymbol{x}\|_\infty \boldsymbol{e} \le \boldsymbol{x},$$

其中

$$\Delta\boldsymbol{x} = \sum_{i\in\{1:d\}} |X_{\mathcal{S}^+}^\top (\boldsymbol{A}^{(i)} - \boldsymbol{B}_{\mathcal{S}^+}^{(i)})| \boldsymbol{e}_i.$$

因此，在假设 5.3 和 5.6下，对于 $t > 0$,

$$\begin{aligned}
|F_{V(\boldsymbol{A})\vee Y}(\boldsymbol{x}) - F_{V(\boldsymbol{B}_{\mathcal{S}^+})\vee Y}(\boldsymbol{x})| &= F_{V(\boldsymbol{B}_{\mathcal{S}^+})\vee Y}(\boldsymbol{x}) - F_{V(\boldsymbol{A})\vee Y}(\boldsymbol{x}) \\
&\le F_{V(\boldsymbol{A})\vee Y - \|\Delta\boldsymbol{x}\|_\infty \boldsymbol{e}}(\boldsymbol{x}) - F_{V(\boldsymbol{A})\vee Y}(\boldsymbol{x}) \\
&\le F_{V(\boldsymbol{A})\vee Y - t\boldsymbol{e}}(\boldsymbol{x}) - F_{V(\boldsymbol{A})\vee Y}(\boldsymbol{x}) + \mathbb{P}\left(\|\Delta\boldsymbol{x}\|_\infty \ge t\right) \\
&= F_{V(\boldsymbol{A})\vee Y}(\boldsymbol{x} + t\boldsymbol{e}) - F_{V(\boldsymbol{A})\vee Y}(\boldsymbol{x}) + \mathbb{P}\left(\|\Delta\boldsymbol{x}\|_\infty \ge t\right) \\
&\lesssim C\sqrt{d}t + \sum_{i\in\{1:d\}} \mathbb{P}\left(\|X_{\mathcal{S}^+}^\top (\boldsymbol{A}^{(i)} - \boldsymbol{B}_{\mathcal{S}^+}^{(i)})\|_2 \ge t\right) \\
&\overset{(5.13)}{\lesssim} C\sqrt{d}t + \frac{1}{t^2} \sum_{i\in\{1:d\}} \|\boldsymbol{A}^{(i)} - \boldsymbol{B}_{\mathcal{S}^+}^{(i)}\|_2^2,
\end{aligned}$$

其中倒数第二个不等式由 $F_{V(\boldsymbol{A})\vee Y}$上的 Lipschitz 条件和并集界得出，最后一个不等式由马尔可夫不等式得出。取 $t = \|\boldsymbol{A} - \boldsymbol{B}_{\mathcal{S}^+}\|_F^{2/3}$ 即可得到所需结果。

A.2.2. 命题证明 (iii)。 我们只证明第一个命题；第二个命题可用类似方法证明。回忆

$$\widehat{F}_{\widehat{H}_{\mathcal{S}}}(\boldsymbol{x}) = G(\widehat{\boldsymbol{B}}_{\mathcal{S}^+}; \boldsymbol{x}) \qquad G(\boldsymbol{A}; \boldsymbol{x}) := \frac{1}{m} \sum_{\ell\in\{1:m\}} \mathbf{1}_{\{(X_{\mathrm{epr},\ell,\mathcal{S}^+}^\top \boldsymbol{A})^\top \le \boldsymbol{x}\}} \quad \boldsymbol{A} \in \mathbb{R}^{(d_{\mathcal{S}}+1)\times d},$$

其中 $X_{\mathrm{epr},\ \ell,\ \mathcal{S}^+}$ 表示 $X_{\mathcal{S}}$ 的第 ℓ个带截距的探索样本。直接计算可得

$$
\begin{aligned}
&\sup_{\boldsymbol{x}\in\mathbb{R}^d}|\widehat{F}_{\widehat{H}_{\mathcal{S}}}(\boldsymbol{x})-F_{H_{\mathcal{S}}}(\boldsymbol{x})| \\
\leq\ &\sup_{\boldsymbol{x}\in\mathbb{R}^d}\left|G(\widehat{\boldsymbol{B}}_{\mathcal{S}^+};\boldsymbol{x})-\mathbb{P}((X_{\mathcal{S}^+}^\top\widehat{\boldsymbol{B}}_{\mathcal{S}^+})^\top\leq\boldsymbol{x})\right|+\sup_{\boldsymbol{x}\in\mathbb{R}^d}\left|\mathbb{P}((X_{\mathcal{S}^+}^\top\widehat{\boldsymbol{B}}_{\mathcal{S}^+})^\top\leq\boldsymbol{x})-\mathbb{P}((X_{\mathcal{S}^+}^\top\boldsymbol{B}_{\mathcal{S}^+})^\top\leq\boldsymbol{x})\right| \\
\leq\ &\sup_{\boldsymbol{A}\in\mathbb{R}^{(d_{\mathcal{S}}+1)\times d}}\sup_{\boldsymbol{x}\in\mathbb{R}^d}\left|G(\boldsymbol{A};\boldsymbol{x})-\mathbb{P}((X_{\mathcal{S}^+}^\top\boldsymbol{A})^\top\leq\boldsymbol{x})\right|+\sup_{\boldsymbol{x}\in\mathbb{R}^d}\left|\mathbb{P}((X_{\mathcal{S}^+}^\top\widehat{\boldsymbol{B}}_{\mathcal{S}^+})^\top\leq\boldsymbol{x})-\mathbb{P}((X_{\mathcal{S}^+}^\top\boldsymbol{B}_{\mathcal{S}^+})^\top\leq\boldsymbol{x})\right| \\
=\ &\underbrace{\sup_{\boldsymbol{A}\in\mathbb{R}^{(d_{\mathcal{S}}+1)\times d}}\left|G(\boldsymbol{A};\boldsymbol{0})-\mathbb{E}[G(\boldsymbol{A};\boldsymbol{0})]\right|}_{\Lambda_{m,1}}+\underbrace{\sup_{\boldsymbol{x}\in\mathbb{R}^d}\left|\mathbb{P}((X_{\mathcal{S}^+}^\top\widehat{\boldsymbol{B}}_{\mathcal{S}^+})^\top\leq\boldsymbol{x})-\mathbb{P}((X_{\mathcal{S}^+}^\top\boldsymbol{B}_{\mathcal{S}^+})^\top\leq\boldsymbol{x})\right|}_{\Lambda_{m,2}}
\end{aligned}
$$

其中，$X_{\mathcal{S}^+}$ 是一个独立于 $\boldsymbol{B}_{\mathcal{S}^+}$ 的一般记号，而$\boldsymbol{0}$ 是全零向量。注意，$\Lambda_{m,\ 1}$ 在 $\boldsymbol{x}$ 上没有上确界，因为可以改变 $\boldsymbol{A}$ 中的截距系数，从而得到 $\boldsymbol{x}\in\mathbb{R}^d$ 的不同取值，同时不改变 $X_{\mathcal{S}}$ 的系数。下文将证明，$\Lambda_{m,\ 1}$ 和$\Lambda_{m,\ 2}$ 几乎处处收敛于 0。

为了界定 $\Lambda_{m,\ 1}$，我们借助经验过程理论。对于任意 $\boldsymbol{A}\in\mathbb{R}^{(d_{\mathcal{S}}+1)\times d}$，示性函数 $\mathbf{1}_{\{(X_{\ell,\mathcal{S}^+}^\top\boldsymbol{A})^\top\leq\boldsymbol{0}\}}\leq 1$. 根据 Massart集中不等式 [7,定理 14。2], 对于任意 $t>0$ 满足

$$\mathbb{P}(\Lambda_{m,1}>\mathbb{E}[\Lambda_{m,1}]+t)\leq\exp\bigl(-mt^2/8\bigr).$$

取 $t=4\sqrt{\log m/m}$,

$$\mathbb{P}\left(\Lambda_{m,1}>\mathbb{E}[\Lambda_{m,1}]+4\sqrt{\frac{\log m}{m}}\right)\leq m^{-2}. \tag{A.1}$$

由于 $\sum_{m=1}^{\infty}m^{-2}<\infty$，根据Borel-Cantelli引理，我们得出

$$\Lambda_{m,1}\leq\mathbb{E}[\Lambda_{m,1}]+4\sqrt{\frac{\log m}{m}} \tag{A.2}$$

对所有充分大的 m，几乎处处成立。为界定 $\mathbb{E}[\Lambda_{m,1}]$，注意到 $\mathbb{E}[\Lambda_{m,1}]$ 中的上确界取遍了定义在 d 个超平面的交集上的所有示性函数，其中 $\mathbb{R}^{d_{\mathcal{S}}}$ （常数维数只会带来平移），这类函数的 Vapnik–Chervonenkis（VC）维数为 $d_{\mathcal{S}}d\log\ d$ [4]阶。根据 [34, 定理 8.3，存在一个通用常数 C' 使得 $\mathbb{E}[\Lambda_{m,1}]\leq C'\sqrt{d_{\mathcal{S}}d}\log\ d/m$。结合（A.2）可知，$\Lambda_{m,\ 1}\to 0$ 几乎处处成立。

为界定 $\Lambda_{m,\ 2}$，注意到根据（ii）引理 5.8中的陈述，几乎处处地，对所有充分大的m，$\|\boldsymbol{B}_{\mathcal{S}^+}-\boldsymbol{B}_{\mathcal{S}^+}\|_F<\varepsilon$，其中 ε 与假设5.6中的相同。由于 $X_{\mathcal{S}^+}$ 与 $\boldsymbol{B}_{\mathcal{S}^+}$独立于$\widehat{\ }$，在 $\|\boldsymbol{B}_{\mathcal{S}^+}-\boldsymbol{B}_{\mathcal{S}^+}\|_F<\varepsilon$ 条件下并应用引理 5.8的陈述（i），$\Lambda_{m,\ 2}\lesssim\|\widehat{\boldsymbol{B}}_{\mathcal{S}^+}-\boldsymbol{B}_{\mathcal{S}^+}\|_F^{2/3}$。现在令 $m\to\infty$，可得$\Lambda_{m,\ 2}\to 0$ 几乎处处成立。

A.2.3. 陈述（iv）的证明。注意 $\mathcal{K}_1(\boldsymbol{x}) + \mathcal{K}_2(\boldsymbol{x}) = F_Y(\boldsymbol{x})(1 - F_Y(\boldsymbol{x}))$，根据大数强律，对于所有 $\boldsymbol{x} \in \mathbb{R}^d$ 几乎处处成立，它是 $F_Y(\boldsymbol{x})(1 - F_Y(\boldsymbol{x}))$ 的一个一致估计量。因此，只需证明 $\mathcal{K}_2(\boldsymbol{x})$ 本身的一致性。

注 $\mathcal{K}_2(\boldsymbol{x})$ 在 (5.3) 中可改写为

$$\mathcal{K}_2(\boldsymbol{x}) = \widehat{\rho}^2_{\mathcal{S}}(\boldsymbol{x})\widehat{F}_Y(\boldsymbol{x})(1 - \widehat{F}_Y(\boldsymbol{x})) = \begin{cases} \frac{(\widehat{F}_{Y\vee\widehat{H}_{\mathcal{S}}}(\boldsymbol{x}) - \widehat{F}_Y(\boldsymbol{x})\widehat{F}_{\widehat{H}_{\mathcal{S}}}(\boldsymbol{x}))^2}{\widehat{F}_{\widehat{H}_{\mathcal{S}}}(\boldsymbol{x})(1-\widehat{F}_{\widehat{H}_{\mathcal{S}}}(\boldsymbol{x}))} & \boldsymbol{x} \in (\mathrm{supp}(\widehat{F}_{\widehat{H}_{\mathcal{S}}}))^\circ \\ 0 & \text{otherwise} \end{cases} \tag{A.3}$$

其中

$$\widehat{\rho}^2_{\mathcal{S}}(\boldsymbol{x}) = \begin{cases} \frac{(\widehat{F}_{Y\vee\widehat{H}_{\mathcal{S}}}(\boldsymbol{x}) - \widehat{F}_Y(\boldsymbol{x})\widehat{F}_{\widehat{H}_{\mathcal{S}}}(\boldsymbol{x}))^2}{\widehat{F}_{\widehat{H}_{\mathcal{S}}}(\boldsymbol{x})(1-\widehat{F}_{\widehat{H}_{\mathcal{S}}}(\boldsymbol{x}))\widehat{F}_Y(\boldsymbol{x})(1-\widehat{F}_Y(\boldsymbol{x}))} & \boldsymbol{x} \in (\mathrm{supp}(\widehat{F}_Y))^\circ \cap (\mathrm{supp}(\widehat{F}_{\widehat{H}_{\mathcal{S}}}))^\circ \\ 0 & \text{otherwise} \end{cases}$$

是 $\rho^2_{\mathcal{S}}(\boldsymbol{x})$ 的经验估计量。另一方面，

$$\rho^2_{\mathcal{S}}(\boldsymbol{x})F_Y(\boldsymbol{x})(1 - F_Y(\boldsymbol{x})) = \begin{cases} \frac{(F_{Y\vee H_{\mathcal{S}}}(\boldsymbol{x}) - F_Y(\boldsymbol{x})F_{H_{\mathcal{S}}}(\boldsymbol{x}))^2}{F_{H_{\mathcal{S}}}(\boldsymbol{x})(1-F_{H_{\mathcal{S}}}(\boldsymbol{x}))} & \boldsymbol{x} \in (\mathrm{supp}(F_{H_{\mathcal{S}}}))^\circ \\ 0 & \text{otherwise} \end{cases} \tag{A.4}$$

比较 (A.3) 和 (A.4)，所需结果由引理 5.8中的陈述 (iii)得出。

A.2.4. 陈述 (v) 的证明。我们证明 $k_2(\mathcal{S})$ 的一致性；$k_1(\mathcal{S})$ 的一致性可类似证明。根据引理 5.8中的陈述 (iv)，$\mathcal{K}_2(\boldsymbol{x})$ 收敛于$\rho^2_{\mathcal{S}}(\boldsymbol{x})F_Y(\boldsymbol{x})(1 - F_Y(\boldsymbol{x}))$，对所有 $\boldsymbol{x} \in \mathbb{R}^d$，当 $m \to \infty$ a.s.

我们首先证明 $d = 1$ 和 $\|\omega\|_{L^\infty(\mathbb{R})} = C < \infty$的第一种情形，并将记号 $\boldsymbol{x}$ 改为小写形式 x。根据假设 5.3中的矩条件，依据[5,定理 2。13],

$$W_1\left(F_Y, \widehat{F}_Y\right) = \int_{\mathbb{R}} |\widehat{F}_Y(x) - F_Y(x)|\mathrm{d}x \to 0 \qquad\qquad m \to \infty,$$

其中，W_1 是 Wasserstein-1度量。固定样本空间中的任意一条轨迹，使得 $\mathcal{K}_2(x) \to \rho^2_{\mathcal{S}}(x)F_Y(x)(1 - F_Y(x))$，且 $\int_{\mathbb{R}} |\widehat{F}_Y(x) - F_Y(x)|\mathrm{d}x \to 0$。下文中，我们将 $\mathcal{K}_2(x)$ 视为一个确定性序列。

为证明{v1}的一致性，$\widehat{k}_2(\mathcal{S})$，还需说明交换取极限与积分次序的合理性：

$$\begin{aligned} \lim_{m\to\infty} \widehat{k}_2(\mathcal{S}) = \lim_{m\to\infty} c_{\mathcal{S}} \int_{\mathbb{R}} \omega(x)\mathcal{K}_2(x)\mathrm{d}x &= c_{\mathcal{S}} \int_{\mathbb{R}} \lim_{m\to\infty} \omega(x)\mathcal{K}_2(x)\mathrm{d}x \\ &= c_{\mathcal{S}} \int_{\mathbb{R}} \omega(x)\rho^2_{\mathcal{S}}(x)F_Y(x)(1 - F_Y(x))\mathrm{d}x = k_2(\mathcal{S}), \end{aligned}$$

为此，我们应用Vitali收敛定理。为了应用Vitali收敛定理，需要验证序列 $\omega(x)\mathcal{K}_2(x)$ 一致可积，并且具有

绝对连续积分。为此，回顾表示式 $\mathcal{K}_2(x)$（见 A.3）。由于经验相关性估计量的平方几乎处处有界于 1，

$$\omega(x)\mathcal{K}_2(x) \le \omega(x)\widehat{F}_Y(x)(1-\widehat{F}_Y(x)) \le \frac{C}{4} < C.$$

绝对连续积分部分直接由一致有界性得出。对于一致可积性，我们首先注意到

$$\begin{aligned}\int_{\mathbb{R}} |\omega(x)F_Y(x)(1-F_Y(x)) - \omega(x)\widehat{F}_Y(x)(1-\widehat{F}_Y(x))|\mathrm{d}x &\le \int_{\mathbb{R}} \omega(x)|\widehat{F}_Y(x) - F_Y(x)|\mathrm{d}x \\ &\le C\int_{\mathbb{R}} |\widehat{F}_Y(x) - F_Y(x)|\mathrm{d}x \to 0.\end{aligned}$$

因此，

$$\begin{aligned}\int_{|x|>M} \omega(x)\mathcal{K}_2(x)\mathrm{d}x &\le \int_{|x|>M} \omega(x)\widehat{F}_Y(x)(1-\widehat{F}_Y(x))\mathrm{d}x \\ &\le \int_{|x|>M} \omega(x)F_Y(x)(1-F_Y(x))\mathrm{d}x + C\int_{|x|>M} |\widehat{F}_Y(x) - F_Y(x)|\mathrm{d}x \\ &\lesssim \int_{|x|>M} \frac{C}{x^2}\mathrm{d}x + C\int_{\mathbb{R}} |\widehat{F}_Y(x) - F_Y(x)|\mathrm{d}x,\end{aligned}$$

其中最后一步由假设 5.3 和切比雪夫不等式得出。对于每个 $\varepsilon > 0$，我们可以选择 m 和 M 足够大，使得右侧小于 ε。通过增大 M 以涵盖前 m 项，即可得到一致可积性。

(b) 的证明类似。只需验证序列$\omega(\boldsymbol{x})\mathcal{K}_2(\boldsymbol{x})$的换序即可。由于 $\omega(\boldsymbol{x})\mathcal{K}_2(\boldsymbol{x}) \le \omega(\boldsymbol{x})$，且后者可积并且与 m 独立，控制收敛便可完成其余证明。

A.2.5. 命题 (vi) 的证明。 对于 $\boldsymbol{x} \in (\mathrm{supp}(F_{H_{\mathcal{S}}}))^\circ$，通过反证法很容易证明 $\boldsymbol{x} \in \mathrm{supp}(F_{\widehat{H}_{\mathcal{S}}})$，对所有充分大的 m 几乎处处成立。根据 (iii) 中的 引理 5.8，$F_{Y\vee\widehat{H}_{\mathcal{S}}}$ 和 $F_{\widehat{H}_{\mathcal{S}}}$ 是相合估计量。同时，由于大数强律，$F_Y(\boldsymbol{x})$ 是 $F_Y(\boldsymbol{x})$ 的一致估计量。因此，我们得到

$$\widehat{\alpha}(\boldsymbol{x}) = \frac{\widehat{F}_{Y\vee\widehat{H}_{\mathcal{S}}}(\boldsymbol{x}) - \widehat{F}_Y(\boldsymbol{x})\widehat{F}_{\widehat{H}_{\mathcal{S}}}(\boldsymbol{x})}{\widehat{F}_{\widehat{H}_{\mathcal{S}}}(\boldsymbol{x})(1-\widehat{F}_{\widehat{H}_{\mathcal{S}}}(\boldsymbol{x}))} \to \alpha(\boldsymbol{x}) = \frac{F_{Y\vee H_{\mathcal{S}}}(\boldsymbol{x}) - F_Y(\boldsymbol{x})F_{H_{\mathcal{S}}}(\boldsymbol{x})}{F_{H_{\mathcal{S}}}(\boldsymbol{x})(1-F_{\boldsymbol{H}_{\mathcal{S}}}(\boldsymbol{x}))}$$

以至于 $m \to \infty$ 几乎处处。

A.3. 引理 5.10的证明。 回顾（5.10）可知

$$\begin{aligned}\widetilde{F}_{\mathcal{S}}(\boldsymbol{x}) &= \widehat{F}_Y(\boldsymbol{x}) - \frac{1}{m}\sum_{\ell\in\{1:m\}}\left(\widehat{\alpha}(\boldsymbol{x})\widehat{h}_{\mathcal{S}}(X_{\mathrm{epr},\ell,\mathcal{S}};\boldsymbol{x}) - \frac{1}{N_{\mathcal{S}}}\sum_{j\in\{1:N_{\mathcal{S}}\}}\widehat{\alpha}(\boldsymbol{x})\widehat{h}_{\mathcal{S}}(X_{\mathrm{ept},\ell,\mathcal{S}};\boldsymbol{x})\right) \\ &= \widehat{F}_Y(\boldsymbol{x}) - \widehat{\alpha}(\boldsymbol{x})\widehat{F}_{\widehat{H}_{\mathcal{S}}}(\boldsymbol{x}) + \widehat{\alpha}(\boldsymbol{x})\left(\frac{1}{N_{\mathcal{S}}}\sum_{j\in\{1:N_{\mathcal{S}}\}}\mathbf{1}_{\{(X_{\mathrm{ept},j,\mathcal{S}^+}^\top\widehat{\boldsymbol{B}}_{\mathcal{S}^+})^\top\le\boldsymbol{x}\}}\right).\end{aligned}$$

因此，

$$\begin{aligned}\sup_{\boldsymbol{x}\in\mathbb{R}^d}|\widetilde{F}_{\mathcal{S}}(\boldsymbol{x})-F_Y(\boldsymbol{x})| &\le \sup_{\boldsymbol{x}\in\mathbb{R}^d}|\widehat{F}_Y(\boldsymbol{x})-F_Y(\boldsymbol{x})|+\sup_{\boldsymbol{x}\in\mathbb{R}^d}|\widehat{\alpha}(\boldsymbol{x})(\widehat{F}_{\widehat{H}_{\mathcal{S}}}(\boldsymbol{x})-F_{H_{\mathcal{S}}}(\boldsymbol{x}))| \\ &\quad+\sup_{\boldsymbol{x}\in\mathbb{R}^d}\left|\widehat{\alpha}(\boldsymbol{x})\left(\frac{1}{N_{\mathcal{S}}}\sum_{j\in\{1:N_{\mathcal{S}}\}}\mathbf{1}_{\{(X^\top_{\mathrm{ept},j,\mathcal{S}+}\widehat{\boldsymbol{B}}_{\mathcal{S}+})^\top\le\boldsymbol{x}\}}-F_{H_{\mathcal{S}}}(\boldsymbol{x})\right)\right| \\ &\overset{(5.9)}{\le} \underbrace{\sup_{\boldsymbol{x}\in\mathbb{R}^d}|\widehat{F}_Y(\boldsymbol{x})-F_Y(\boldsymbol{x})|}_{(i)}+\underbrace{\sup_{\boldsymbol{x}\in\mathbb{R}^d}|\widehat{F}_{\widehat{H}_{\mathcal{S}}}(\boldsymbol{x})-F_{H_{\mathcal{S}}}(\boldsymbol{x})|}_{(ii)} \\ &\quad+\underbrace{\sup_{\boldsymbol{x}\in\mathbb{R}^d}\left|\left(\frac{1}{N_{\mathcal{S}}}\sum_{j\in\{1:N_{\mathcal{S}}\}}\mathbf{1}_{\{(X^\top_{\mathrm{ept},j,\mathcal{S}+}\widehat{\boldsymbol{B}}_{\mathcal{S}+})^\top\le\boldsymbol{x}\}}-F_{H_{\mathcal{S}}}(\boldsymbol{x})\right)\right|}_{(iii)}.\end{aligned}$$

注（i）根据多元 Glivenko-Cantelli 定理，当 $m\to\infty$ 时收敛于 0。（ii）根据引理 5.8 中的陈述 （iii），当 $m\to\infty$ 时收敛于 0。可使用引理 5.8 的陈述 （iii）的证明中的类似论证，证明（iii）当$N_{\mathcal{S}}\to\infty$时收敛于 0，此处不再重复。

A.4. 定理 5.7 的证明。为避免与 m在记号上的混淆，我们使用 t 表示探索样本的数量。探索率 m 相对于计数探索循环迭代次数的索引呈非线性增长，该循环位于 算法 5.2中。令 q 表示循环迭代索引，令 t_q 表示相应的探索率，即 $t_1=n+2$。令 $q(B)$ 表示 算法 5.2中的探索迭代步骤总数，该数量是随机的。根据定义，有 $t_{q(B)}=m(B)$，并且

$$n+1\le t_q\le t_{q+1}\le 2t_q \qquad 1\le q<q(B). \tag{A.5}$$

我们首先证明，$m(B)$ $B\to\infty$ 几乎处处发散。根据引理 5.8 中的陈述 (v)，

$\widehat{k}_1(\mathcal{S})\to k_1(\mathcal{S})$ $\widehat{k}_2(\mathcal{S})\to k_2(\mathcal{S})$ 对于 $\mathcal{S}\subseteq\{1:n\}$ 几乎处处成立。因此，对于几乎每个实现$\omega\in\Omega$（其中 Ω 表示探索样本的乘积空间），存在一个 $0<L(\omega)$ $L'(\omega)<\infty$，

$$\sup_{t>n+1}\max_{\mathcal{S}\subseteq\{1:n\}}\widehat{k}_1(\mathcal{S};\omega)<L(\omega)<\infty \qquad \inf_{t>n+1}\min_{\mathcal{S}\subseteq\{1:n\}}\widehat{k}_2(\mathcal{S};\omega)>L'(\omega)>0$$

算法算法 5.2的 cvMDL 探索停止准则要求

$$m(B;\omega)\ge\widehat{m}^*_{\mathcal{S}(B;\omega)}\ge\frac{B}{c_{\mathrm{epr}}+\sqrt{\frac{c_{\mathrm{epr}}L(\omega)}{L'(\omega)}}}\to\infty \qquad B\to\infty.$$

因此，

$$\lim_{B\to\infty}m(B;\omega)=\infty. \tag{A.6}$$

现在我们沿着一个固定实现 ω 进行工作，在该实现下，当 $B \to \infty$ 时， $m(B;\omega) \to \infty$ ，且 $k_1(\mathcal{S})$ $k_2(\mathcal{S})$ 收敛于真实参数。我们证明，对于这样的 ω，(5.14a) 和 (5.14b) 均成立。固定一个充分小的 $\delta < 1/2$ 。由于假设 $\mathcal{S}^*$ 是唯一的，连续性论证表明，存在充分大的 $T(\delta;\omega)$，使得对于所有 $t \geq T(\delta;\omega)$，

$$\max_{(1-\delta)m^*_{\mathcal{S}^*} \leq m \leq (1+\delta)m^*_{\mathcal{S}^*}} \widehat{L}_{\mathcal{S}^*}(m;t) < \min_{\mathcal{S}\subseteq\{1:n\},\mathcal{S}\neq\mathcal{S}^*} \widehat{L}^*_{\mathcal{S}}(t). \tag{A.7}$$

$$1-\delta \leq \frac{\widehat{m}^*_{\mathcal{S}}(t;\omega)}{m^*_{\mathcal{S}}} \leq 1+\delta \qquad \forall \mathcal{S} \subseteq \{1:n\}, \tag{A.8}$$

其中 $\widehat{L}_{\mathcal{S}^*}(\cdot;t)$ 是使用 $\mathcal{S}^*$ t 探索样本得到的估计损失函数，而 $L^*_{\mathcal{S}}(t)$ 是使用 $L^*_{\mathcal{S}}$ （3.11）中的 t 探索样本得到的估计。

由于 $m^*_{\mathcal{S}}$ 在 B 中呈线性增长，而 $m(B;\omega)$ 随着 $B \to \infty$发散，因此存在足够大的 $B(\delta;\omega)$，使得对于 $B > B(\delta;\omega)$，

$$\min_{\mathcal{S}\subseteq\{1:n\}} m^*_{\mathcal{S}} > 4T(\delta;\omega) \tag{A.9}$$

$$t_{q(B)} = m(B;\omega) > 4T(\delta;\omega). \tag{A.10}$$

考虑 $q' < q(B)$，其满足 $t_{q'-1} < T(\delta;\omega) \leq t_{q'}$。这样的 q' 总是存在，这是由于（A.10），并且满足

$$t_{q'} \overset{(A.5)}{\leq} 2t_{q'-1} < 2T(\delta;\omega) \overset{(A.9)}{\leq} \frac{1}{2}\min_{\mathcal{S}\subseteq\{1:n\}} m^*_{\mathcal{S}} \overset{(A.8),\delta<1/2}{\leq} \widehat{m}^*_{\mathcal{S}}(t_{q'};\omega).$$

该不等式表明，在第 q' 次循环迭代中，对于所有 $\mathcal{S} \subseteq \{1:n\}$，相应的估计最优探索率大于当前探索率。在这种情况下，

$$\widehat{L}_{\mathcal{S}}(t_{q'} \vee \widehat{m}^*_{\mathcal{S}}(t_{q'};\omega);t_{q'}) = \widehat{L}_{\mathcal{S}}(\widehat{m}^*_{\mathcal{S}}(t_{q'};\omega);t_{q'}) = \widehat{L}^*_{\mathcal{S}}(t_{q'}) \qquad \forall \mathcal{S} \subseteq \{1:n\}.$$

这与（A.7）和（A.8）一起表明， $\mathcal{S}^*$ 是当前步骤中的估计最优模型，还需要进行更多探索。

为了确定 $t_{q'+1}$ 应为何值，我们考虑两种不同的情况。如果 $2t_{q'} \leq \widehat{m}^*_{\mathcal{S}^*}(t_{q'};\omega)$，那么

$$T(\delta;\omega) < t_{q'+1} = 2t_{q'} \leq \widehat{m}^*_{\mathcal{S}^*}(t_{q'};\omega) \leq (1+\delta)m^*_{\mathcal{S}^*},$$

这意味着

$$(1-\delta)m^*_{\mathcal{S}^*} \overset{(A.8)}{\leq} t_{q'+1} \vee \widehat{m}^*_{\mathcal{S}^*}(t_{q'+1};\omega) \leq (1+\delta)m^*_{\mathcal{S}^*}. \tag{A.11}$$

如果 $t_{q'} \leq \widehat{m}^*_{\mathcal{S}^*}(t_{q'};\omega) < 2t_{q'}$，那么

$$t_{q'+1} = \left\lceil \frac{t_{q'} + \widehat{m}^*_{\mathcal{S}^*}(t_{q'};\omega)}{2} \right\rceil \leq \widehat{m}^*_{\mathcal{S}^*}(t_{q'};\omega) \overset{(A.8)}{\leq} (1+\delta)m^*_{\mathcal{S}^*},$$

这也意味着（A.11）。但是，A.11）与（A.7）和（A.8）结合可推出， $\mathcal{S}^*$ 再次成为（$q'+1$）次循环迭代中的估计最优模型。应用上述论证

通过归纳可证明 $\mathcal{S}(B) = \mathcal{S}^*$，即（5.14b）。注意，（A.11）在算法终止之前始终成立；结合终止条件 $t_{q(B)} \geq \widehat{m}^*_{\mathcal{S}^*}(t_{q(B)}; \omega) \geq (1-\delta) m^*_{\mathcal{S}^*}$可推出

$$1 - \delta \leq \frac{m(B;\omega)}{m^*_{\mathcal{S}^*}} = \frac{t_{q(B)}}{m^*_{\mathcal{S}^*}} \leq 1 + \delta.$$

（5.14a）现由注意到 δ 可以任意小而得出。

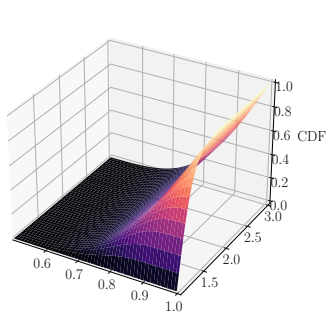
最后，给定 B，按照（5.10）选取，其中 $\mathcal{S} = \mathcal{S}^*$，$m = m^*_{\mathcal{S}^*}$且 $N_{\mathcal{S}} = (B - c_{\text{epr}} m^*_{\mathcal{S}^*}) / c_{\mathcal{S}^*}$。注意，$m$和 $N_{\mathcal{S}}$ 都是确定的，并且当 $B \to \infty$时发散。由三角不等式，

$$\sup_{\boldsymbol{x} \in \mathbb{R}^d} |\widetilde{F}(\boldsymbol{x}; B) - F_Y(\boldsymbol{x})| \leq \sup_{\boldsymbol{x} \in \mathbb{R}^d} |\widetilde{F}(\boldsymbol{x}; B) - \widetilde{F}'(\boldsymbol{x}; B)| + \sup_{\boldsymbol{x} \in \mathbb{R}^d} |\widetilde{F}'(\boldsymbol{x}; B) - F_Y(\boldsymbol{x})|.$$

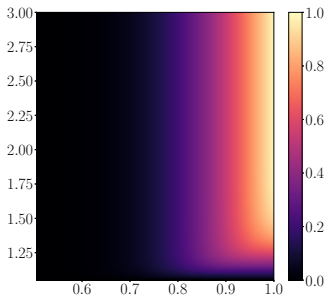
由于 $B \to \infty$，右侧第一项由定理 5.7中的（5.14a）和（5.14b）收敛于 0，而右侧第二项由定理 5.10收敛于 0。这证明了（5.14c）。

附录 B。其他数值结果。

B.1. 第 6.2 小节中几何布朗运动的补充结果第 6.2 小节。我们展示两幅图，以补充第 6.2 小节中的实验结果。图中展示了 Oracle 累积分布函数，见图11（左、中）。预言模型损失和探索样本数见图 11（右）。图12展示了一个热力图实例，其中呈现了 ECDF、cv MDL 和 cvMDL-sorted 的绝对估计误差，当 $B = 10^6$时，为 cvMDL 比 ECDF 更准确提供了佐证，适用于 $\mathcal{T}$。



(a)

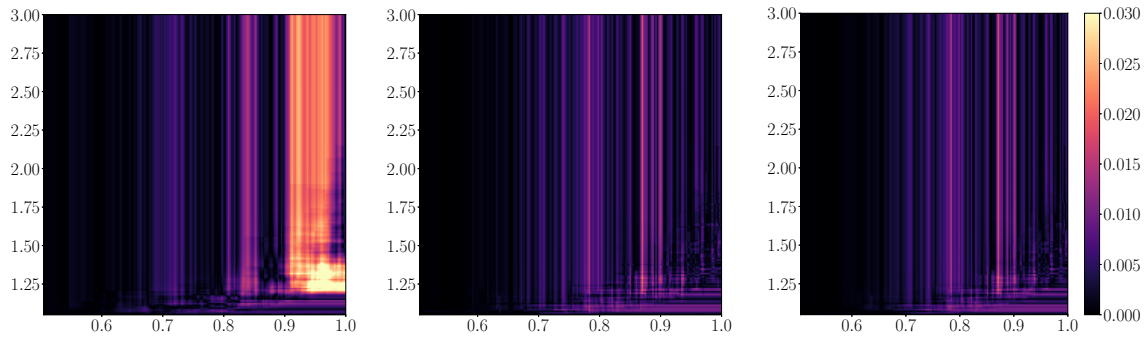


(b)

模型 $\mathcal{S}$	$\gamma_{\mathcal{S}}$	$m^*_{\mathcal{S}}$
{1}	**11.3**	**613**
{2}	13.7	23.2
{3}	23.2	902
{1, 2}	12.2	596
{1, 3}	11.6	606
{2, 3}	14.2	790
{1, 2, 3}	12.4	581

(c)

图 11：几何布朗运动。（a）-（b）：使用 10^5 个蒙特卡洛样本计算得到的高保真模型中（$S_{\min}$，$S_{\max}$）的基准累积分布函数。（c）：预算为 $B = 10^6$时，使用 50,000 个样本计算得到的预言机缩放损失 $\gamma_{\mathcal{S}}$（3.11）和最优探索样本数 $m^*_{\mathcal{S}}$（3.11）。



(a) (b) (c)

图12：几何布朗运动。预算为 $B = 10^6$ 时，(a) 经验累积分布函数、(b) cvMDL 和 (c) cvMDL-sorted 的绝对逐点估计误差示例。

B.2. 第 6.3 小节中脆性断裂的补充结果第 6.3 小节。我们提供了更多实验细节，以补充第 6.3 小节中的内容。

回顾式（6.1）和（6.2）中的边值问题。作用于 Γ_N和 Γ_D 上的边界条件为

$$\begin{aligned}[(1-\phi_d(\mathbf{x}))^2+q]\nabla\cdot\boldsymbol{\sigma}(\mathbf{x}) &= \mathbf{v}_n, \quad \mathbf{x} \text{ on } \Gamma_N,\\ \mathbf{u}(\mathbf{x}) &= \mathbf{0}, \quad \mathbf{x} \text{ on } \Gamma_D,\\ \nabla\phi_d(\mathbf{x})\cdot\mathbf{n} &= 0, \quad \mathbf{x} \text{ on } \Gamma_N,\end{aligned}$$

其中 $q \ll 1$, $\boldsymbol{\sigma}(\mathbf{x}) = \dfrac{\partial\Psi}{\partial\boldsymbol{\epsilon}(\mathbf{x})}(\boldsymbol{\epsilon}(\mathbf{x}))$ 是柯西应力张量，而 $\Psi(\boldsymbol{\epsilon}(\mathbf{x})) = \dfrac{1}{2}\lambda(\mathrm{tr}(\boldsymbol{\epsilon}(\mathbf{x})))^2 + \mu\mathrm{tr}(\boldsymbol{\epsilon}(\mathbf{x})^2)$是弹性能量密度，其中 μ 和 λ 是 Lam´e 常数，即，

$$\lambda = \frac{\nu\kappa}{(1+\nu)(1-2\nu)}, \qquad \mu = \frac{\kappa}{2(1+\nu)}$$

其中杨氏模量为 κ，泊松比为 ν，且 $\boldsymbol{\epsilon}(\mathbf{x}) = \dfrac{1}{2}\left[\nabla\mathbf{u}(\mathbf{x}) + \nabla\mathbf{u}(\mathbf{x})^\top\right]$ 为小应变张量。在（6.2）中，历史变量 $H(\mathbf{x})$ 定义如下：

$$H(\mathbf{x}) = \begin{cases}\Psi(\boldsymbol{\epsilon}(\mathbf{x})), & \Psi(\boldsymbol{\epsilon}(\mathbf{x})) < H_i(\mathbf{x})\\ H_i(\mathbf{x}), & \text{otherwise}\end{cases}, \quad i = 1, 2, \ldots, n,$$

其中，$H_i(\mathbf{x})$是在离散化载荷的第 i步计算的应变能，该步对应于迭代求解器阶段 $\bar{v}_i \cdot e_2$，其中 $\bar{v}_i \in [0, \bar{v}]$。

参考文献

[1] H. Amor, J.-J. Marigo, and C. Maurini, *Regularized formulation of the variational brittle fracture with unilateral contact: Numerical experiments*, 《力学与固体物理学杂志》, 57（2009）, 第1209–1229页。
[2] E. Andreassen, A. Clausen, M. Schevenels, B. S. Lazarov, and O. Sigmund, *Efficient topology optimization in matlab using 88 lines of code*, 《结构与多学科优化》, 43（2011）, 第1–16页。
[3] Q. Ayoul-Guilmard, S. Ganesh, S. Krumscheid, and F. Nobile, *Quantifying uncertain system outputs via the multi-level Monte Carlo method–distribution and robustness measures* , 《国际不确定性量化杂志》, 13（2023）。[4] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth, *Learnability and the vapnik-chervonenkis dimension* 36 1989 929 965 , 《ACM学报》（JACM）, （）, 第–页。[5] S. Bobkov and M. Ledoux, *One-dimensional empirical measures order statistics and kantorovich transport distances* 261 2019 , , 《美国数学学会回忆录》, （）。[6] B. Bourdin, G. A. Francfort, and J.-J. Marigo, *Numerical experiments in revisited brittle fracture*, 《力学与固体物理学杂志》, 48（2000）, 第797–826页。
[7] P. Bühlmann and S. Van De Geer, *Statistics for high-dimensional data: methods theory and applications* 2011 , , 施普林格科学与商业媒体, 。[8] M. Croci, K. Willcox, and S. Wright, *Multi-output multilevel best linear unbiased estimators via semidefinite programming*, 《计算机在应用力学与工程中的方法》, 413（2023）, 第116130页。[9] I.-G. Farcas, *Context-aware model hierarchies for higher-dimensional uncertainty quantification*, 博士论文, Technische Universit¨at M¨unchen, 2020年。[10] I.-G. Farcas, B. Peherstorfer, T. Neckel, F. Jenko, and H.-J. Bungartz, *Context-aware learningof hierarchies of low-fidelity models for multi-fidelity uncertainty quantification*, 《计算机在应用力学与工程中的方法》, 406（2023）, 第115908页。[11] M. B. Giles, *Multilevel Monte Carlo path simulation*, 《运筹学》, 56（2008）, 第607–617页。[12] M. B. Giles, T. Nagapetyan, and K. Ritter, *Multilevel Monte Carlo approximation of distribution functions and densities* 3 2015 267 295 , SIAM/ASA不确定性量化期刊, （）, 第–页。[13] M. B. Giles, T. Nagapetyan, and K. Ritter, *Adaptive multilevel Monte Carlo approximation of distribution functions* 1706 06869, 2017 , arXiv预印本arXiv:。（）。[14] P. Glasserman, *Monte Carlo methods in financial engineering*, 第53卷, 施普林格, 2004年。
[15] P. Glasserman, P. Heidelberger, and P. Shahabuddin, *Efficient Monte Carlo methods for value-at-risk* 2 2000 5 18 , 《风险精通》, （）, 第–页。[16] A. A. Gorodetsky, G. Geraci, M. S. Eldred, and J. D. Jakeman, *A generalized approximate controlvariate framework for multifidelity uncertainty quantification*, 《计算物理学杂志》, 408（2020）, 第109257页。[17] A. Gruber, M. Gunzburger, L. Ju, and Z. Wang, *A multifidelity monte-carlo method for realistic computational budgets* 94 2023 2 , 《科学计算杂志》, （）, 第 页。[18] M. Heinkenschloss, B. Kramer, and T. Takhtaganov, *Adaptive reduced-order model constructionfor conditional value-at-risk estimation*, SIAM/ASA不确定性量化期刊, 8（2020）, 第668–692页。[19] M. Heinkenschloss, B. Kramer, T. Takhtaganov, and K. Willcox, *Conditional-value-at-riskestimation via reduced-order models*, SIAM/ASA不确定性量化期刊, 6（2018）, 第1395–1423页。[20] T. C. Hesterberg and B. L. Nelson, *Control variates for probability and quantile estimation*, 《管理科学》, 44（1998）, 第1295–1312页。[21] S. Krumscheid and F. Nobile, *Multilevel Monte Carlo approximation of functions*, SIAM/ASA不确定性量化期刊, 6（2018）, 第1256–1293页。
[22] T. Lattimore and C. Szepesvári, *Bandit algorithms*, 剑桥大学出版社, 2020年。[23] D. Lu, G. Zhang, C. Webster, and C. Barbier, *An improved multilevel monte carlo method for estimating probability distribution functions in stochastic oil reservoir simulations*, 《水资源研究》, 52（2016）, 第9642–9660页。

[24] A. J. McNeil, R. Frey, and P. Embrechts, *Quantitative risk management: concepts techniques and tools-revised edition*, 普林斯顿大学出版社, 2015。[25] S. Natarajan and R. K. Annabattula, *A FEniCS implementation of the phase field method for quasi-static brittle fracture*, 《结构与土木工程前沿》, 13 (2019), 380–396。[26] A. B. Owen, *Monte Carlo theory, methods and examples*, 斯坦福, 2013年。[27] B. Peherstorfer, *Multifidelity monte carlo estimation with adaptive low-fidelity models*, SIAM/ASA不确定性量化期刊, 7 (2019), 579–603。[28] B. Peherstorfer, B. Kramer, and K. Willcox, *Multifidelity preconditioning of the cross-entropymethod for rare event simulation and failure probability estimation*, SIAM/ASA不确定性量化期刊, 6 (2018), 737–761。[29] B. Peherstorfer, K. Willcox, and M. Gunzburger, *Optimal model management for multifidelity Monte Carlo estimation*, SIAM科学计算期刊, 38 (2016), 3163–3194。[30] E. Qian, B. Peherstorfer, D. O'Malley, V. V. Vesselinov, and K. Willcox, *Multifidelity MonteCarlo estimation of variance and sensitivity indices*, SIAM/ASA不确定性量化期刊, 6 (2018), 683–706。[31] C. R. Rao, C. R. Rao, M. Statistiker, C. R. Rao, and C. R. Rao, *Linear statistical inference and its applications*, 2, 1973, 第卷, Wiley, 纽约, 。[32] D. Schaden and E. Ullmann, *On multilevel best linear unbiased estimators*, SIAM/ASA不确定性量化期刊, 8 (2020), 601–635。[33] D. Schaden and E. Ullmann, *Asymptotic analysis of multilevel best linear unbiased estimators*, SIAM/ASA不确定性量化期刊, 9 (2021), 953–978。[34] R. Vershynin, *High-dimensional probability: An introduction with applications in data science*, 第47卷, 剑桥大学出版社, 2018年。[35] Y. Xu, V. Keshavarzzadeh, R. M. Kirby, and A. Narayan, *A bandit-learning approach to multifidelity approximation*, SIAM科学计算期刊, 44 (2022), 150–175。[36] Y. Xu and A. Narayan, *Budget-limited distribution learning in multifidelity problems*, 《数值数学》, 153 (2023), 171–212。