

DOI: 10.7511/jslx20230525005

子集模拟的混合采样算法

廖子涵^{1,2}, 李宾宾^{*1,2}

(1. 浙江大学 伊利诺伊大学厄巴纳香槟校区联合学院, 海宁 314400; 2. 浙江大学 建筑工程学院, 杭州 310058)

摘要:子集模拟是当前可靠度计算领域常用的估计算法,相比于直接蒙特卡罗积分法,极大减少了函数调用的次数。子集模拟法一般使用单一采样器,然而不同采样器适用范围不同。如使用单一椭圆切片采样器,其遍历性较好但函数调用次数较多;而使用单一自适应条件采样时,其采样效率较高但样本容易陷入局部极值。单一采样器由于本身特性面对不同问题时失效概率的积分结果可能出现偏差,模拟效果不稳定。本文首次提出了一种混合采样子集模拟法,在子集模拟的前几层使用椭圆切片采样,此时失效区域收缩程度有限,函数调用次数在可接受的范围内,样本经过采样扩充后能充分探索参数空间,更有效地检测出所有失效区域。当失效区域收缩至一定限度后,使用自适应条件采样,此时种子样本继承了前几层样本较低的相关性,并在此基础上通过自适应条件采样更高效地增殖样本。本文通过四个模拟算例多种维度下的数值积分验证了该算法具有椭圆切片采样器较好的遍历性,同时采样效率位于椭圆切片采样与自适应条件采样之间,对于不同问题拥有良好的通用性。

关键词:结构可靠度;子集模拟;椭圆切片采样;混合算法

中图分类号: O213.2

文献标志码: A

文章编号: 1007-4708(2023)05-0693-08

1 引言

实际工程中,结构面临几何尺寸、材料属性、边界条件和荷载等不确定性因素。因此,可靠性分析对于结构设计优化和结构安全性评估等工作具有重要意义。其中结构失效概率的计算通常通过以下三种方式求解,(1)近似估计法,如基于功能函数泰勒展开的可靠度分析方法。该方法对于高度非线性问题与多失效点问题不适用。(2)蒙特卡洛积分法,包括直接蒙特卡洛模拟及马尔科夫链蒙特卡洛模拟 MCMC(Markov Chain Monte Carlo)法。此类方法具有良好的通用性与准确性,但计算代价较高。(3)代理模型法。常用的代理模型包括响应面法^[1]、高斯过程模型^[2]、Kriging 模型^[3]和支持向量机^[4,5]等。当参数维度较大时,该方法需要提取特征进行降维分析,这必然导致信息丢失且无法确保失效概率估计的无偏性。

为解决高维参数空间下失效区域概率估计问题,Au 等^[6]提出子集模拟算法,利用条件概率引入中间事件,自适应地将一个小的失效概率分解为一系列相对较大的条件概率的乘积。通过 MCMC 采样,将样本经由中间事件进行一系列筛选与生

成,滚动地不断接近参数空间中的失效区域。其作为一种较为高效的数值计算方法,广泛应用于各类结构可靠度的计算。此外,由于可靠度问题(小概率事件)、优化问题(极值事件)和贝叶斯更新问题(极限状态方程可以看作一类特殊的似然函数)之间存在联系,子集模拟近年也通过改造逐渐应用在多个领域。Suo 等^[7]将非支配排序引入样本排序的规则,实现多目标优化。广义子集模拟^[8]将子集模拟法的应用推广到多极限状态方程与时变可靠性分析领域^[9]。

子集模拟中,MCMC 算法的使用在增强采样效率的同时导致样本之间存在相关性。这种相关性不仅会增大条件概率的计算方差,还会使得算法的遍历性下降,样本无法对失效区域进行有效覆盖,进而影响单次模拟结果的可信度。目前降低样本相关性主要体现在参数优化和 MCMC 采样算法优化两个方面。其中相关参数的优化集中于条件概率的选取,条件概率越低,模拟收敛的速度越快,但马氏链之间相关性较高,遍历性下降。相反的,条件概率越高,算法遍历性越好,但层与层之间的相关性上升,模拟所需计算量增大,子集模拟算法的优势削弱。

采样器的更新优化是近年来子集模拟发展的主要方向。在最初的子集模拟中,Au 等^[6]采用的是改进 MH(Metropolis Hastings)算法。为了减少马尔科夫链内的相关性,Santoso 等^[10]提出重复

收稿日期:2023-05-25;修改稿收到日期:2023-07-20.

基金项目:浙江省自然科学基金(LY21E080025).

作者简介:李宾宾*(1986-),男,博士,研究员

(E-mail:binbinli@intl.zju.edu.cn).

采样版本的 MH 采样法,通过重复采样保证产生的待选样本与之前样本不同,然而该方法的人为选择性,导致 MH 算法中的可逆性条件不能满足,生成的样本虽然相关性降低,但分布与目标分布之间存在误差^[11]。此外,通过对种子样本进行拉伸采样^[13],也能从马尔科夫链的起始点降低链与链之间的相关性。更复杂地,Miao 等^[13]提出再生自适应子集模拟算法,通过随机再生建议分布以及延迟拒绝算法,克服 burn in 问题并且尽可能多地获得独立样本。然而,在子集模拟中马尔科夫链的种子分布已经自动满足目标分布,因此再生建议分布理论上对于解决 burn in 问题不起作用^[14]。为了方便在高维空间中对多种分布组成的概率密度函数进行蒙特卡罗采样,Papaioannou 等^[11]提出自适应条件采样 aCS(adaptive Conditional Sampling),将变量 θ 通过 Nataf 变换^[15]转化为独立和标准正态分布的变量 u ,并在标准正态空间 U 中进行采样与数值积分。该方法在简便性与准确性方面都较为突出,获得了广泛的应用。为了提升样本探索空间的效率,将 Hamiltonian 采样嵌入子集模拟法中^[16,17],该方法基于功能函数的梯度矩阵,提升样本向函数值较高的区域移动的速度,但同时会增加计算梯度矩阵的负担。连续空间转换^[18]通过控制变量方法先给出一个粗略的失效概率估计值,再通过修正项进行细化修正,通过扩大样本分布的方差以增加样本的遍历性。

2 子集模拟法

2.1 子集模拟法

定义 n 维结构参数 $\theta \in \mathbb{R}^n$ 及其概率密度函数 $f_n(\theta)$,对于代表结构失效模式的极限状态方程 $g(\theta)$,结构失效概率可表达为

$$\Pr(F) = \int f_n(\theta) I_F(\theta) d\theta \quad (1)$$

式中 $I_F(\theta)$ 为失效事件 $F = \{\theta \in \mathbb{R}^n : g(\theta) \leq 0\}$ 的指示函数,即

$$I_F(\theta) = \begin{cases} 1 & (g(\theta) \leq 0) \\ 0 & (g(\theta) > 0) \end{cases} \quad (2)$$

式(1)可通过直接蒙特卡罗采样积分方式近似求解,然而结构失效一般为小概率事件,失效域仅占参数空间的极小部分,大量样本点落在安全域,造成计算资源的严重浪费,特别是在参数 θ 的维度 n 很大的情况下。

子集模拟算法将结构失效事件 F 表示为一系列逐级包含的失效事件的交集

$$F = \bigcap_{j=1}^M F_j \quad (3)$$

$$F_1 \supset F_2 \supset \dots \supset F_M = F \quad (4)$$

式中 $F_j = \{\theta \in \mathbb{R}^n : g(\theta) \leq b_j\}$, $b_1 > b_2 > \dots > b_M = 0$ 。利用条件概率的性质,原结构失效概率转化为一列条件失效概率的乘积

$$P_f = \Pr(F) = \Pr(\bigcap_{j=1}^M F_j) = \prod_{j=1}^M \Pr(F_j | F_{j-1}) \quad (5)$$

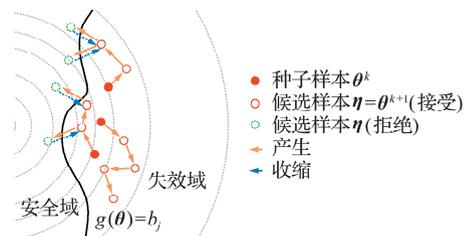
式中 $F_0 = \{\theta \in \mathbb{R}^n : g(\theta) \leq \infty\}$ 为确定事件,即 $\Pr(F_0) = 1, \Pr(F_1 | F_0) = \Pr(F_1)$ 。每一层的条件失效概率计算可参照式(1)进行,

$$\Pr(F_j | F_{j-1}) = \int f_n(\theta | F_{j-1}) I_{F_j}(\theta) d\theta \quad (6)$$

对应的条件概率密度函数为

$$f_n(\theta | F_{j-1}) = \frac{f_n(\theta)}{\Pr(F_{j-1})} I_{F_{j-1}}(\theta) \quad (7)$$

由于条件概率密度函数 $f_n(\theta | F_{j-1})$ 不是标准的概率分布,如何实现其快速有效采样成为实现失效概率估计的关键。基于这一观察,子集模拟可看作是一种自适应重要性采样算法,从初始概率密度函数 $f_n(\theta)$ 出发,通过给定条件失效概率 $p_0 = \Pr(F_1 | F_0)$ (一般取为 0.1 或 0.2) 确定分界点 b_1 和对应失效域 F_1 ,然后以失效域 F_1 内采样点作为种子,基于 MCMC 原理实现 $f_n(\theta | F_1)$ 的采样并进而确定新的分界点 b_2 和失效域 F_2 ,以此类推,直至达到目标失效域 $F_M = F$ 。从上述过程可以看出,子集模拟选用 MCMC 采样,每一层样本基于上一层传递下来的种子样本,从而实现条件概率密度函数 $f_n(\theta | F_{j-1})$ 的自适应采样,且种子本身满足分布 $f_n(\theta | F_{j-1})$,大大提升了候选样本的接受率。



(注:图中圆弧形虚线为样本概率密度函数的等密度线)

图1 样本运动

Fig. 1 Sample movement

值得注意的是,子集模拟法中的 MCMC 采样通常对 $f_n(\theta) I_{F_{j-1}}(\theta)$ 的采样分为两步,即产生步与收缩步。如图 1 所示,从种子样本 θ^k 出发进行移动,产生候选样本 η ,以 $f_n(\theta)$ 为目标函数计算接受率 $\hat{\alpha}_1(\eta, \theta^k)$

$$\hat{\alpha}_1(\eta, \theta^k) = \min \left\{ 1, \frac{T_n(\theta^k | \xi) f_n(\eta)}{T_n(\eta | \theta^k) f_n(\theta^k)} \right\} \quad (8)$$

式中 $T_n(\cdot|\cdot)$ 为转移概率密度。通过随机数与接受率比较后筛选得到候选样本 ξ 。此时,候选样本的总体分布介于 $f_n(\theta|F_{j-1})$ 与 $f_n(\theta)$ 之间,近似于从 $f_n(\theta|F_{j-1})$ 扩散到 $f_n(\theta)$ 的过渡阶段。第二步,基于指示函数 $I_{F_{j-1}}(\theta)$ 对候选样本 ξ 进行二次筛选,将超出边界的样本点收缩回初始点,得到更新后的样本 θ^{k+1} ,即

$$\theta^{k+1} = \begin{cases} \xi & (g(\xi) \leq b_{j-1}) \\ \theta^k & (g(\xi) > b_{j-1}) \end{cases} \quad (9)$$

对应接受率 $\hat{\alpha}_2(\xi) = I_{F_{j-1}}(\xi)$ 。以上两步采样本质上可理解为标准 MCMC 采样的拆解形式,因为其最终接受率满足

$$\hat{\alpha}(\xi, \theta^k) = \min \left\{ 1, \frac{T_n(\theta^k|\xi)f_n(\xi|F_{j-1})}{T_n(\xi|\theta^k)f_n(\theta^k|F_{j-1})} \right\} = \min \left\{ 1, \frac{T_n(\theta^k|\xi)f_n(\xi)}{T_n(\xi|\theta^k)f_n(\theta^k)} \frac{I_{F_{j-1}}(\xi)}{I_{F_{j-1}}(\theta^k)} \right\} = \hat{\alpha}_1(\eta, \theta^k) \hat{\alpha}_2(\xi) \quad (10)$$

即满足细致平衡条件。式(10)利用 $I_{F_{j-1}}(\theta^k) = 1$ 和函数 $\min\{\cdot\}$ 的性质。需要指出的是,直接应用 MCMC 采样并不适用于高维随机采样,也就是存在高维灾难问题,本文将简要介绍两种常用的高维采样算法,从而使子集模拟能够解决高维可靠度估计问题。

2.2 高维随机采样

2.2.1 改进 MH 采样

改进 MH 采样是 Au 和 Beck 在子集模拟^[6]提出的算法。该算法首先通过 Nataf 变换将随机变量 θ 转换到标准正态空间 U ,使得每个维度相互独立且都满足标准正态分布,即 $u_i \sim \phi(u_i) = N(0, 1)$,相应的每步采样目标由 $f_n(\theta|F_{j-1})$ 变换为 n 维截断正态分布 $\phi_n(u|F_{j-1})$ 。然后,依次从 $1 \sim n$ 逐维进行如下操作,基于第 i 维样本 u_i^k ,通过建议分布 $q_i(\cdot|u_i^k)$ 产生候选样本 η_i ,依据接受概率

$$\hat{\alpha}_1(\eta_i, \theta_i^k) = \min \left\{ 1, \frac{q_i(u_i^k|\eta_i)\phi(\eta_i)}{q_i(\eta_i|u_i^k)\phi(u_i^k)} \right\} \quad (11)$$

确定是否接受 η_i ,并将所有接受样本重新映射回原参数空间以确定其是否位于失效域,进行相应收缩操作。由于该算法将参数空间转换为独立标准正态空间,可证明其满足马尔科夫链采样的细致平衡条件^[1]。不同于传统 MH 采样,该算法采用单维多次方式,从而提升了候选样本的接受率,适用于高维采样问题。

2.2.2 自适应条件采样

自适应条件采样由 Papaioannou 等^[11]提出,同样在标准正态空间进行。为实现 $\phi_n(u|F_{j-1})$ 的

采样,给定种子样本 u^k ,该算法采用

$$\eta = \rho u^k + \lambda \sigma du \quad (12)$$

产生候选样本,其中 $\sigma = \text{diag}(\sigma_1, \dots, \sigma_n)$,一般取为种子样本的标准差; $\rho_i = \sqrt{1 - (\lambda\sigma_i)^2}$, $\rho = \text{diag}(\rho_1, \dots, \rho_n)$; $du \sim N_n(0, 1)$,即 n 维独立标准正态分布; λ 为缩放系数。式(12)计算出的样本自动满足标准正态分布,因此接受率 $\hat{\alpha}_1(\eta, \theta^k) = 1, \xi = \eta$ 且 $\xi \neq u^k$,即每步都有新样本产生。当然,候选样本 ξ 还需收缩运算,因此无法保证 $u^{k+1} \neq u^k$ 。利用多维正态分布的性质,条件采样式(12)将 n 维采样问题成功转化为一维采样问题,整个采样过程由一个标量缩放系数 λ 控制,从而克服了马尔科夫链采样的高维灾难。为达到最优采样,即满足最终接受率 $\hat{\alpha}(\xi, \theta^k) = 0.44$, Papaioannou 等^[11]提出自适应策略,在采样过程中自动选取缩放系数 λ ,具体步骤参见文献[11]。需要指出的是,本文经验表明,即使使用自适应策略,对大部分问题而言最终接受率仍很难达到 0.44,因为随着结构失效概率的降低,采到新的样本变得愈发困难。总体而言,该方法在遍历性与高效性之间取得了较好的平衡。

3 混合采样算法

3.1 椭圆切片采样

椭圆切片采样 ES(Elliptical slice sampling)^[19]为切片采样在正态空间的一种特殊形式。与条件采样和哈密顿采样类似,首先对现有样本 u 进行扰动

$$v = u \cos\beta + du \sin\beta \quad (13)$$

式中 $\beta \sim U[0, 2\pi], du \sim N_n(0, 1)$ 。设定下一步角度选取的上下界 $[\beta_{\min}, \beta_{\max}] = [\beta_1 - 2\pi, \beta_1]$ 。使用公式对候选样本进行筛选,若拒绝,则将 β_1 设定为分布的上界或下界

$$\begin{aligned} \beta_{\min} &= \beta & (\beta < 0) \\ \beta_{\max} &= \beta & (\beta \geq 0) \end{aligned} \quad (14)$$

基于更新后的上下界均匀随机取值 β 并重复以上步骤,直到样本接受。该方法本质上为一种收缩条件采样(或正态分布下的轨迹收缩哈密顿采样)。将边界条件的筛选步骤与新样本的生成结合在一起,保证新产生的样本在满足边界条件的前提下互相不重复,增加了等效独立样本的数量,并且每次对角度的搜索从全分布出发,增强了样本的遍历性。相对的,椭圆切片采样收缩过程中拒绝步骤的嵌入也会带来较大的计算量。如图 2 所示,椭圆切片采样在单次产生步骤的样本范围大于自适应条件采样,带来较好遍历性的同时降低收缩步的接受概率。

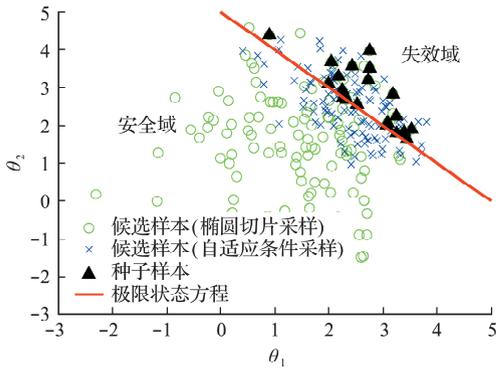
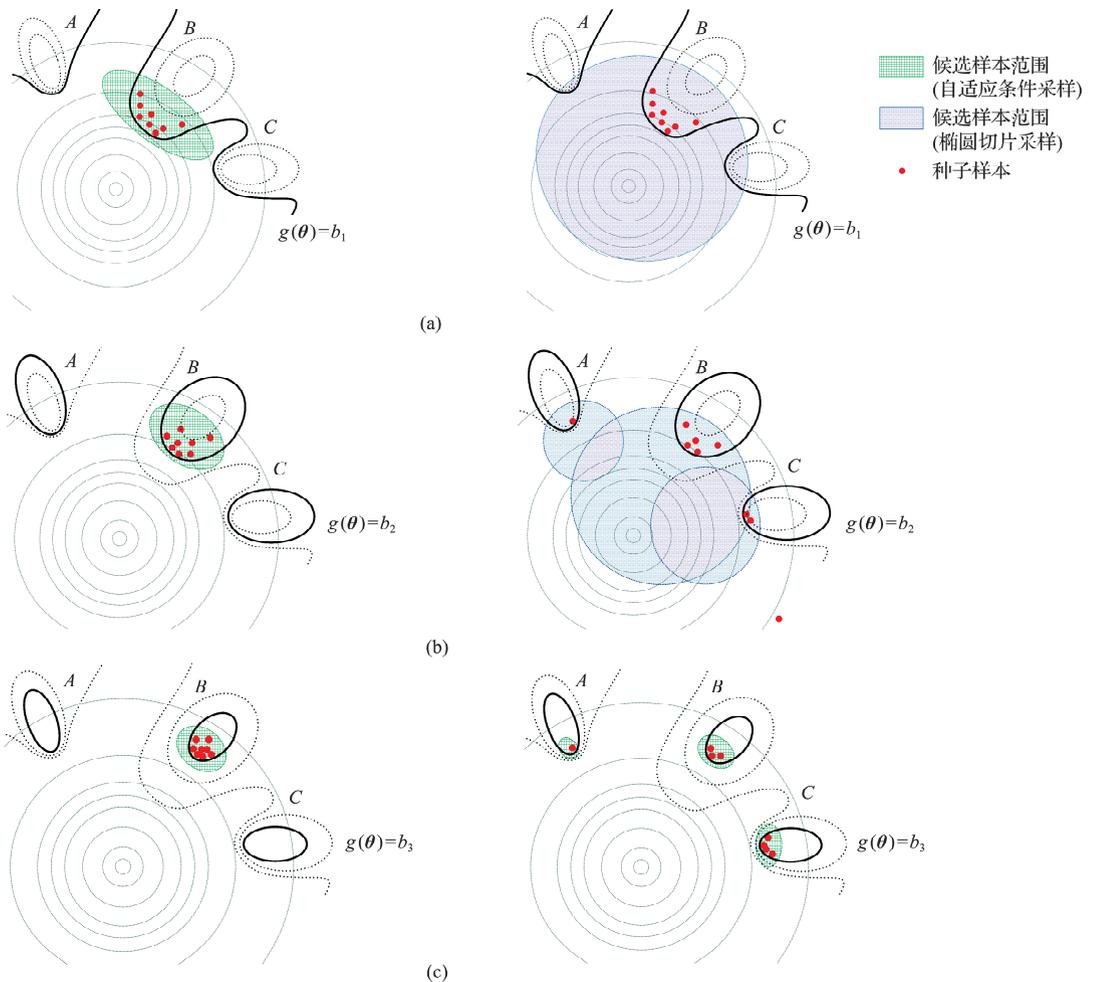


图2 产生步样本分布

Fig.2 Sample distribution in Generation step VS aCS sampler

通过分析可知,在可靠度分析领域,椭圆切片采样范围缩减次数的上限与问题维度和失效概率相关。每一次范围的缩减,由于均匀取值的特性,每个维度缩减比例的期望为 0.5。若该问题失效区域在每个维度上占比相等,子集模拟最后一层内(所有种子样本皆位于失效区域,所需范围缩减次数最大)采样范围缩减次数的期望可以表示为

$$\log_{0.5} P_f = \log_{0.5} P_f / n \quad (15)$$



(注:图中左下部分圆形虚线为样本概率密度函数的等密度线,右上部分实(虚)曲线则为极限状态方程的等值线)

图3 自适应条件采样器和混合采样器
Fig.3 aCS sampler VS Mixed sampler

式中 n 为问题的维度。可以看出,椭圆切片采样的效率随着问题的维度上升而升高(满足上述假定的前提下),但若存在极端情况,即无论问题维度,失效区域的占比始终集中于一维,其余维度采样范围不受目标函数影响,则椭圆切片采样的效率将不随维度变化。

为保证算法效率,在椭圆切片采样中设置范围缩减次数的上限 N_{max} ,以避免目标函数的过多调用。考虑到上述极端情况的存在与结构失效概率的量级,本文设置 $N_{max} = 20$,其在极端情况下仍能采样到失效区域($P_f \geq 0.5^{20} \approx 9.54 \times 10^{-7}$)内的样本。

3.2 混合采样子集模拟

本文基于上述采样方法提出了一种搭配混合采样器的子集模拟方法。与传统子集模拟始终调用一种采样器不同,该方法在一次模拟内先后调用椭圆切片采样与自适应条件采样两种采样器。如图3所示的示意性算例中,条件失效区域分别标记为A,B和C,随着模拟的三个区域不断缩小

$(g(\theta) < b_1 \rightarrow g(\theta) < b_3 = 0)$ 。假设传统子集模拟法(自适应条件采样器)与本文方法(混合采样器)第一层的种子样本相同,都位于区域 B 。本文提出的新方法先使用椭圆切片采样器,提升样本的遍历性,使得候选样本范围能够到达 A 和 C 区域(图 3(a)),避免陷入极限状态方程的局部极值。第二层中(图 3(b)),此时失效区域已经缩减到较小的范围,椭圆切片采样的效率降低,因此新方法在第三层启用效率更高的自适应条件采样(图 3(c)),在较少的函数调用次数下,更高效地探索后续层的参数空间。可以看出,新算法通过对不同采样器的先后使用,较好地利用了不同采样器的优势,同时一定程度上弥补了各自的局限性。

为避免使用椭圆切片采样器的层数过多导致计算资源浪费,需要设置自适应的采样器切换边界以提升新算法的通用性。本文中当椭圆切片采样器函数调用次数大于样本数的 4 倍时,判断椭圆切片采样器效率过低,切换为自适应条件采样器。

4 算例分析

援用 4 个算例(3 个可自行控制维度的高维算例以及一个高度非线性算例),从不同角度检验新算法的特性并对比与其他算法的优缺点。

4.1 n 维线性超平面^[20]

极限状态方程为

$$g_1(x_i) = \beta\sqrt{n} - \sum_{i=1}^n x_i \quad (16)$$

$$X_{1 \sim n} \sim N(0, 1)$$

式中 $X_i (i = 1:n)$ 为独立的标准正态分布变量。该算例中 β 值控制失效概率的大小,且失效概率不随维度变化,又因为该算例为线性方程,使用 FORM 可以找到失效概率的解析值,给算法的测试提供了较为便利的环境,可较方便地用来研究算法在不同概率水平和不同维度下的性能。本文取 $\beta = 6$, 对应的失效概率 $P_f = 9.87 \times 10^{-10}$ ^[20]。

4.2 n 维非线性超曲面 1^[21]

极限状态方程的非线性集中于前两维

$$g_2(x_i) = \beta_0 - \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i - \frac{k}{4} (x_1 - x_2)^2 \quad (17)$$

$$X_{1 \sim n} \sim N(0, 1)$$

式中 $X_i (i = 1:n)$ 为独立的标准正态分布变量。 k 为控制函数非线性的曲率参数, k 越大,失效概率越小,对算法要求越大。本文取 $k = -10$, $\beta_0 = 4$, 对应的失效概率 $P_f = 4.29 \times 10^{-6}$ ^[21]。

4.3 n 维非线性超曲面 2^[22]

第二个非线性算例的极限状态方程为以下形式,除第 n 维外,其他维度皆存在非线性。

$$g_3(x_i) = \frac{2}{n-1} \sum_{i=1}^{n-1} x_i^2 - x_n \quad (18)$$

$$X_{1 \sim n-1} \sim N(0, 1)$$

$$X_n \sim N(1, 0.2)$$

式中 $X_i (i = 1:n-1)$ 为独立的标准正态分布变量, X_n 满足均值为 1, 标准偏差为 0.2 的正态分布。与 4.1 节线性超平面算例不同的是,该算例的失效概率与维度相关且失效概率的值无法通过计算直接得到。因此本文将多次蒙特卡罗积分计算结果的平均值作为失效概率的准确值^[22]。

4.4 高度非线性问题^[23]

最后一个算例为白噪声激励下具有不确定阻尼振子的双自由度系统,如图 4 所示,极限状态方程为

$$g_3(x_i) = F_s - 3k_s \sqrt{\frac{\pi S_0}{4\zeta_s \omega_s^3}} \left[\frac{\zeta_a \zeta_s}{\zeta_p \zeta_s (4\zeta_a^2 + \theta^2) + \gamma \zeta_a^2} \frac{(\zeta_p \omega_p^3 + \zeta_s \omega_s^3) \omega_p}{4\zeta_a \omega_a^4} \right] \quad (19)$$

式中 $\omega_p = \sqrt{k_p/m_p}$, $\omega_s = \sqrt{k_s/m_s}$, $\omega_a = (\omega_p + \omega_s)/2$, $\zeta_a = (\zeta_p + \zeta_s)/2$, $\gamma = m_s/m_p$, $\theta = (\omega_p - \omega_s)/\omega_a$, 涉及参数分布列入表 1。该算例的特点是高度非线性以及较低的失效概率,失效概率的准确值参考文献^[23]。

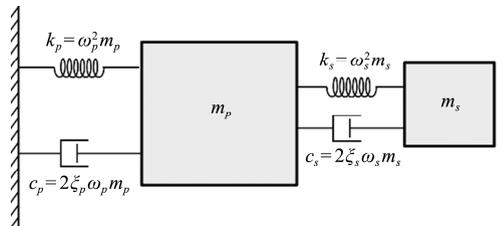


图 4 白噪声激励下的双自由度系统
Fig. 4 Two degrees-of-freedom system under white noise excitation

表 1 算例 4 涉及参数分布

Tab. 1 Distribution of parameters in example 4

参数	m_p	m_s	k_p	k_s	ζ_p	ζ_s	F_s	S_0
分布	对数正态分布							
均值	1.5	0.01	1	0.01	0.05	0.02	27.5	100
变异系数	0.1	0.1	0.2	0.2	0.4	0.5	0.1	0.1

由表 2 可知,所有算例在各类采样器下运行 5000 次,保证失效概率浮动基本不影响偏差判断,

结果列入表 2。对于算例 1,除了改进 MH 采样外,其余算法基本不存在偏差,椭圆切片器由于其采样效率较低,结果的变异系数较大。混合采样器的变异系数位于自适应条件采样与椭圆切片采样之间。在该算例中,所有维度的采样范围都随着模拟层数的增加不断缩减,因此即使该算例对应的失效概率较低,椭圆切片采样仍能取得较好的效果。对于自适应条件采样,该算例的极限状态方程为线性,对于遍历性要求较低,因此也能取得较好的效果。

算例 2 中,自适应条件采样的偏差与变异系数都明显好于椭圆切片采样,混合采样偏差与自适应条件采样接近,变异系数位于两者之间,明显优于传统的改进 MH 采样。在该算例中,极限状态方程的非线性与失效区域的收缩集中在前两个维度,对于自适应条件采样较为有利,而对于椭圆切片采

表 2 各算法数值积分结果

Tab. 1 Results of different MCMC algorithms

维度	改进 MH 采样	自适应条件 采样	椭圆切片 采样	混合 采样
算例 1 线性超平面				
50	+16.33	+4.84	+6.25	-0.92
	±100.27 (8113)	±70.88 (8530)	±103.74 (8381)	±87.17 (8371)
100	+18.64	+1.90	+7.17	-0.41
	±104.52 (8079)	±68.25 (8530)	±106.61 (8449)	±93.78 (8359)
200	+19.12	+2.10	+0.86	-2.00
	±106.93 (8071)	±67.33 (8527)	±90.33 (8493)	±82.69 (8358)
算例 2 非线性超曲面 1				
50	+15.60	+12.42	+20.28	+9.47
	±75.52 (5287)	±65.71 (5394)	±86.72 (5168)	±72.34 (5212)
100	+15.82	+9.36	+14.32	+7.10
	±73.86 (5289)	±59.52 (5397)	±82.08 (5217)	±66.19 (5321)
200	+17.31	+5.68	+20.78	+7.02
	±69.26 (5280)	±54.48 (5411)	±87.24 (5193)	±69.81 (5218)
算例 3 非线性超曲面 2				
101	+9.14	+14.84	-0.27	+0.70
	±28.02 (3171)	±27.05 (3075)	±34.18 (3102)	±33.57 (3069)
150	+12.77	+25.77	+2.33	+0.19
	±36.86 (3691)	±34.85 (3691)	±43.27 (3607)	±42.09 (3696)
200	+16.37	+36.23	+4.73	+2.21
	±44.86 (3788)	±41.42 (3742)	±51.71 (3675)	±49.35 (3762)
算例 4 高度非线性问题				
8	+3.72	-2.04	+13.49	+1.72
	±121.59 (6258)	±79.74 (6336)	±142.75 (6350)	±104.14 (6230)

注:表中数字表示:偏差/%±变异系数/(函数调用次数)。

样的效率不利。在模拟的最后几层中,椭圆切片采样在前两维的采样范围缩减严重,造成样本的大量重复,也导致其模拟结果较差。

对于第三个算例,椭圆切片采样的偏差相较于自适应条件采样更优,变异系数则较大。混合采样偏差与椭圆切片采样接近,变异系数位于两者之间。在该算例中,极限状态方程的非线性与失效区域的收分散于所有维度,对于采样算法的遍历性要求较高,混合采样在保证遍历性的前提下较椭圆切片采样效率更高,因此数值积分结果的变异系数更低。

第四个算例为低维高非线性低失效概率算例,所有采样器除椭圆切片采样外偏差接近,其中自适应条件采样的变异系数最小。该算例说明自适应采样对于低维非线性问题依旧具有良好的遍历性。与算例 2 类似,椭圆切片采样对此类问题存在样本点难以移动的问题,导致最终模拟结果不理想。混合采样的效果优于椭圆切片采样与传统的改进 MH 采样。

5 结 论

为解决单一采样器带来的不同问题下通用性较差的缺点,本文提出了一种混合采样的子集模拟法。经各类算例验证,混合采样器相比于其他单一类型的采样器具有以下优势。

(1) 同时具有椭圆切片采样器遍历性较好与自适应条件采样器采样效率高的优点。

(2) 通过调用次序不同,可有效避开两种采样器的短板,提升算法的鲁棒性。

(3) 通过自适应的切换机制,自动选择各阶段适合的采样器,对于不同问题具有良好的通用性。

(4) 继承了两种采样器对于高维问题的适用性,对于算例的各类高维问题都有较优秀的表现。

然而该算法依旧存在以下有待发展的方面。

(1) 椭圆切片采样部分依旧不够高效,导致同等函数调用次数下,积分结果的方差依旧大于单一自适应条件采样,未来可考虑用替代模型减少函数调用次数,增强算法效率。

(2) 自适应的采样器切换边界只基于函数调用次数与样本数的比值,较粗糙。此外,层内采样过程也可考虑调用不同采样器以达到更佳的效果。

目前算法只包含椭圆切片采样与自适应条件采样两种采样器,其他方法如哈密顿采样和拉伸采样等具有各自不同特点的采样器也可以加入采样器候选库中,使得在子集模拟过程中,可针对模型各阶段的特性选择最适合的采样方法。

参考文献(References):

- [1] 武清玺,卓家寿. 二序列响应面法分析重力坝的动力可靠度[J]. 振动工程学报, 2001, **14**(2): 224-227. (WU Qing-xi, ZHUO Jia-shou. Dynamic reliability analysis of gravity dams using a second-order sequence response surface method[J]. *Journal of Vibration Engineering*, 2001, **14**(2): 224-227. (in Chinese))
- [2] Zhou T, Peng Y B. Kernel principal component analysis-based Gaussian process regression modelling for high-dimensional reliability analysis[J]. *Computers & Structures*, 2020, **241**: 106358.
- [3] Huang X X, Chen J Q, Zhu H P. Assessing small failure probabilities by AK-SS: An active learning method combining Kriging and Subset Simulation[J]. *Structural Safety*, 2016, **59**: 86-95.
- [4] 赵安龙,李洪双. 基于概率支持向量机的可靠性分析与设计方法[J]. 应用力学学报, 2017, **34**(1): 50-56, 195. (ZHAO An-long, LI Hong-shuang. Reliability analysis and optimization design based on probabilistic support vector machine [J]. *Chinese Journal of Applied Mechanics*, 2017, **34**(1): 50-56, 195. (in Chinese))
- [5] Bourinet J M, Deheeger F, Lemaire M. Assessing small failure probabilities by combined subset simulation and support vector machines [J]. *Structural Safety*, 2011, **33**(6): 343-353.
- [6] Au S K, Beck J L. Estimation of small failure probabilities in high dimensions by subset simulation[J]. *Probabilistic Engineering Mechanics*, 2001, **16**(4): 263-277.
- [7] Suo X S, Yu X Q, Li H S. Subset simulation for multi-objective optimization[J]. *Applied Mathematical Modelling*, 2017, **44**: 425-445.
- [8] Li H S, Ma Y Z, Cao Z J. A generalized subset simulation approach for estimating small failure probabilities of multiple stochastic responses[J]. *Computers & Structures*, 2015, **153**: 239-251.
- [9] Li H S, Wang T, Yuan J Y, et al. A sampling-based method for high-dimensional time-variant reliability analysis[J]. *Mechanical Systems and Signal Processing*, 2019, **126**: 505-520.
- [10] Santoso A M, Phoon K K, Quek S T. Modified Metropolis-Hastings algorithm with reduced chain correlation for efficient subset simulation[J]. *Probabilistic Engineering Mechanics*, 2011, **26**(2): 331-341.
- [11] Papaioannou I, Betz W, Zwirgmaier K, et al. MCMC algorithms for subset simulation [J]. *Probabilistic Engineering Mechanics*, 2015, **41**: 89-103.
- [12] Shields M D, Giovanis D G, Sundar V S. Subset simulation for problems with strongly non-Gaussian, highly anisotropic, and degenerate distributions [J]. *Computers & Structures*, 2021, **245**: 106431.
- [13] Miao F, Ghosn M. Modified subset simulation method for reliability analysis of structural systems [J]. *Structural Safety*, 2011, **33**(4-5): 251-260.
- [14] Au S K, Beck J L, Zuev K M, et al. Discussion of paper by F. Miao and M. Ghosn "Modified subset simulation method for reliability analysis of structural systems", *Structural Safety*, 33: 251-260, 2011 [J]. *Structural Safety*, 2012, **34**(1): 379-380.
- [15] Hohenbichler M, Rackwitz R. Non-normal dependent vectors in structural safety [J]. *Journal of the Engineering Mechanics Division*, 1981, **107**(6): 1227-1238.
- [16] Wang Z Q, Broccardo M, Song J. Hamiltonian Monte Carlo methods for subset simulation in reliability analysis [J]. *Structural Safety*, 2019, **76**: 51-67.
- [17] 贾少敏,王子琦,陈华霆,等. 多点激励下减震桥梁结构抗震可靠度分析的哈密顿蒙特卡洛子集模拟法 [J]. 振动工程学报, 2021, **34**(2): 357-363. (JIA Shao-min, WANG Zi-qi, CHEN Hua-ting, et al. Hamiltonian Monte Carlo based subset simulation for reliability analysis of energy-dissipation bridge structures with viscous dampers under multi-support seismic excitations [J]. *Journal of Vibration Engineering*, 2021, **34**(2): 357-363. (in Chinese))
- [18] Rashki M. SESC: A new subset simulation method for rare-events estimation [J]. *Mechanical Systems and Signal Processing*, 2021, **150**: 107139.
- [19] Murray I, Adams R, MacKay D. Elliptical slice sampling [A]. Proceedings of the thirteenth international conference on artificial intelligence and statistics [C]. 2010.
- [20] Englund S, Rackwitz R. A benchmark study on importance sampling techniques in structural reliability [J]. *Structural Safety*, 1993, **12**(4): 255-276.
- [21] Ghalehnovi M, Rashki M, Ameryan A. First order control variates algorithm for reliability analysis of engineering structures [J]. *Applied Mathematical Modelling*, 2020, **77**: 829-847.
- [22] Dang C, Xu J. Unified reliability assessment for problems with low-to high-dimensional random inputs using the Laplace transform and a mixture distribution [J]. *Reliability Engineering & System Safety*, 2020, **204**: 107124.
- [23] Bourinet J M. Rare-event probability estimation with adaptive support vector regression surrogates [J]. *Reliability Engineering & System Safety*, 2016, **150**: 210-221.

Subset simulation method with mixed sampler

LIAO Zi-han^{1,2}, LI Bin-bin^{*1,2}

(1. ZJU-UIUC Institute, Zhejiang University, Haining 314400, China;

2. College of Civil Engineering and Architecture, Zhejiang University, Hangzhou 310058, China)

Abstract: Subset simulation method is a widely used Monte Carlo integration method in failure probability estimation. Compared with direct Monte Carlo integration method, it greatly reduces the number of function calls. Subset simulation method generally uses a single sampler, but different samplers have different application ranges. If a single elliptical slice sampler is used, its ergodicity is better, but the number of function calls is higher; When using single adaptive conditional sampling, its sampling efficiency is high, but it is easy for the samples to fall into a local extremum. When a single sampler faces different problems due to its own characteristics, the integration results of failure probability may deviate, and the simulation effect is unstable. In this paper, a mixed sampling set simulation method is proposed for the first time. Elliptic slice sampling is used in the first few layers of subset simulation. At this time, the shrinkage of failure area is limited, and the number of function calls is within an acceptable range. After sampling expansion, the sample can fully explore the parameter space and detect all failure areas more effectively. When the failure area shrinks to a certain limit, adaptive conditional sampling is used. At this time, the seed samples inherit the lower correlation of the previous layers of samples, and on this basis, the samples are multiplied more efficiently through adaptive conditional sampling. In this paper, the numerical integration of four simulation examples in multiple dimensions verifies that the algorithm has good ergodicity of elliptic slice sampler, and the sampling efficiency is between elliptic slice sampling and adaptive conditional sampling. It has good universality for different problems.

Key words: subset simulation; Monte Carlo simulation; slice sampling; conditional sampling

会议简讯

《计算力学学报》第八届编委会召开第二次工作会议

2023年8月21日,《计算力学学报》第八届编委会在大连召开第2次工作会议。会议由主编大连理工大学亢战教授主持。副主编西安交通大学王铁军教授,副主编大连理工大学李刚教授,副主编大连理工大学郭旭教授,编委河北工业大学韩旭教授,西北工业大学邓子辰教授,河海大学章青教授,清华大学张雄教授,厦门大学王东东教授,中科院大学王智慧教授,兰州大学刘小靖教授,湖南大学侯淑娟教授,北京理工大学廉艳平教授,大连理工大学陈颺松教授,专职副主编冯颖编审全程参加了会议,学报编辑部编辑刘燕老师列席会议。

亢战主编首先感谢各位编委的出席,感谢编委为学报发展所做的工作。然后按照设定的会议日程逐项进行。

专职副主编、编辑部主任冯颖编审汇报了学报近三年在全体编委的大力支持下开展的工作、工作成效以及学报在发展过程中遇到的瓶颈问题、面临的机遇和挑战,以详实的数据和图表介绍和分析了期刊的现状。感谢编委对提升学报影响力做出的贡献,未来将继续秉持内容为王,质量优先原则,做好期刊定位和发展规划,纽约优质稿源,坚持学术标准,加强数字化建设,进一步提升期刊质量和社会影响力。

讨论环节,各位参会编委讨论热烈,分别就提高稿源质量、优化审稿流程、缩短发表周期、提升期刊影响力等方面提出了具有建设性的意见。会议历时一个半小时,在亢战主编总结讲话后结束。

(本刊编辑部)

引用本文/Cite this paper:

廖子涵,李宾宾.子集模拟的混合采样算法[J].计算力学学报,2023,40(5):693-700.

LIAO Zi-han,LI Bin-bin. Subset simulation method with mixed sampler[J]. *Chinese Journal of Computational Mechanics*, 2023, 40(5): 693-700.