Advances in Applied Mathematics 应用数学进展, 2025, 14(3), 192-202
Published Online March 2025 in Hans. https://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2025.143105

Hans汉斯

基于Fiducial预测密度的模型选择

范　晨

青岛大学数学与统计学院，山东　青岛

收稿日期：2025年2月11日；录用日期：2025年3月4日；发布日期：2025年3月14日

摘　要

模型选择作为统计分析的一个重要工具，它能选出候选模型集中拟合数据生成过程最好的那个模型。为了得到更好的预测标准，本文提出了一种基于Fiducial预测密度的模型选择方法，同时加入容许集来进一步压缩候选模型集，给出理论的同时也给出了MH算法去应用它。最后，对本文提出的模型选择进行了模拟研究与实例分析，结果均表明我们的方法优于其他方法。

关键词

模型选择，容许集，Fiducial预测密度

Model Selection Based on Fiducial Predictive Distribution

Chen Fan

School of Mathematics and Statistics, Qingdao University, Qingdao Shandong

Received: Feb. 11th, 2025; accepted: Mar. 4th, 2025; published: Mar. 14th, 2025

Abstract

Model selection is an essential tool in statistical analysis, enabling the identification of the best-fitting model for the data-generating process from a set of candidate models. To achieve better predictive performance, this paper proposes a model selection method based on Fiducial predictive density, incorporating an admissible set to further reduce the candidate model space. Theoretical foundations are provided, along with a Metropolis-Hastings (MH) algorithm for its implementation. Finally, simulation studies and empirical analyses are conducted to evaluate the proposed method. The results demonstrate that our method outperforms existing approaches in terms of both predictive accuracy and efficiency.

文章引用: 范晨. 基于 Fiducial 预测密度的模型选择[J]. 应用数学进展, 2025, 14(3): 192-202.
DOI: 10.12677/aam.2025.143105

Keywords

Model Selection, Admissible Set, Fiducial Predictive Distribution

1. 引言

随着大数据时代的到来，数据分析已经成为各行各业不可或缺的工作之一，如何有效地利用统计建模从数据中挖掘出有用的信息也越来越受到人们的关注。模型选择可以在建模过程中找到对响应变量最具有解释性的自变量集，提高模型的解释性和预测精度，因而也随之备受关注。模型选择在统计学与计量经济学中有着悠久的历史，研究者们提出了不同的模型选择方法与准则。目前可将其分为两大类：(1) 以优化理论为支撑的正则化模型选择；(2) 通过构建候选模型的概率描述来进行模型选择。

正则化变量选择因其计算速度快和对超高维度数据具有良好的适应性而得到迅速发展。常见的方法有 Lasso [1]、SCAD [2]、EN [3]、MCP [4]等方法，它们具有良好的稀疏性和收敛性。但以优化理论为支撑的正则化方法并不能给出所选模型为最优模型的概率保证，而且往往会导致过拟合现象。

量化的概率保证可以给从业者直观指导，这在实际应用中是十分重要的。其中一种概率保证是贝叶斯模型选择，有基于边际似然的最高模型后验准则 BIC [5]和基于偏差信息量的最高模型后验准则 DIC [6]，但前者计算复杂度往往很大且难以处理超高维问题，后者不能就近区分数据生成模型和过拟合模型[7]。另一种是频率学变量选择方法，常见的方法有 AIC 方法[8]，TIC 方法[9]、Mallows Cp 方法[10]、交叉验证[11]等，但它们都过于依赖参数个数。

近年来，基于 Fiducial 推断理论的模型选择开始得到关注。Fiducial 推断的思想是一种介于频率学派与贝叶斯学派的新思想，可以追溯到由 Fisher [12]提出了单参数的 Fiducial 分布的概念开始，并建议用 Fiducial 分布代替 Bayes 后验分布来获取参数的区间估计。对于单参数分布族，Fisher 的 Fiducial 置信区间与经典置信区间重合；对于多参数分布族，Fiducial 置信集的覆盖率接近名义水平，但在重复抽样频率意义上不精确。随后 Fiducial 推断得到进一步的发展，这其中包括 Dempster-Shafer 理论[13]，推断模型[14]，置信分布[15]等。这些现代 Fiducial 理论极大地推广了 Fisher 的 Fiducial 思想。Majumder 和 Hannig [16]建立了广义 Fiducial 推断的高阶渐近定义。Long 和 Xu [17]在研究模型分类问题中提出了 Fiducial 预测密度(FPD)的概念，并且证明了 FPD 拥有很好的依概率收敛的性质。

Fiducial 推断理论的快速发展，使得基于 Fiducial 推断构建模型选择准则成为可能。该准则不仅避免了贝叶斯模型选择方法依赖于先验信息的问题，还可以给出所选模型为最优模型的概率保证，并且为研究模型选择准则的大样本性质提供可能。Hannig 和 Lee [18]在 2009 年首次将 Fiducial 推断理论与模型选择相结合，导出了 Fiducial 模型选择范式，其主要思想是将模型 M 作为未知参数，引入结构方程中。但直接通过模型的 Fiducial 概率来选择最优模型往往会出现较为严重的过拟合现象，因此，Hannig 在此基础上施加最小描述长度惩罚(Minimum description length, MDL)使得结果有一定的改观。李涵等[19]提出基于 Fiducial 推断的 Kriging 模型变量选择方法，给出了如何从一般 Kriging 和正交 Kriging 模型选出最优模型。张淑芹等[20]给出了拟 Fiducial 推断的 Kriging 模型选择方法，发现相比于将 Lasso 和 EN 应用进 Kriging 模型进行模型选择，拟 Fiducial 推断的模型选择方法具有更高的拟合准确性和预测精度。

为了减少候选模型的个数，从而在增强真实模型所反映的信号的同时，进一步减轻计算负担，Williams

等[21]引入了 ε-容许集的概念。该方法极大地压缩了候选模型集合的数量。这项工作带来了一种新的 Fiducial 模型选择方法，被称为 EAS (ε-admissible set，ε-容许集)。Williams 等[22]进一步将 EAS 方法推广到了向量自回归模型。赵勇超等[23]提出了一种基于 Fiducial 边际似然函数的模型不确定性度量新方法 FML，并将所提出的模型不确定性度量整合到 EAS 中，提出了一种 Fiducial 变量选择准则 FMC，该准则可以很好地处理高维甚至超高维的情形。

本文将基于 FPD 函数提出一个新的模型不确定性度量方法，该方法在进行模型选择时同时考虑了参数的 Fiducial 分布和未来观测值的 Fiducial 预测密度，同时考虑进 ε-容许集来加强对候选模型集的压缩，并记该模型选择方法为 FPMC，同时给出了 MH 算法去实现它。本文结构如下：第二节介绍了 Fiducial 预测模型选择准则 FPMC，第三节通过数值模拟及实例分析来说明本文方法的优越性。第四节是对本文的总结与展望。

2. Fiducial 预测密度模型选择

2.1. 高维线性回归模型

在本节中给出本文所考虑的模型及问题，并对部分符号作出说明。考虑 n 个独立的观测 (x_i, y_i)，$i=1,\cdots,n$ 其中 $x_i \in \mathbb{R}^p$，$y_i \in \mathbb{R}$。记 $y=(y_1,\cdots,y_n)'$。

令 $S=\{M \mid M \subset \{1,2,\cdots,p\}\}$ 是所有候选模型所构成的集合，则集合 S 中总共包括了 2^p 个候选模型。$|M|$ 是模型 M 的模型长度。对于候选模型 M，本文考虑回归系数为 β，误差项为 $\varepsilon_M \sim N_n(0,\sigma_M^2 I_n)$ 的多元正态线性回归模型，即

$$y = X_M \beta_M + \varepsilon_M \tag{1}$$

其中设计阵 X_M 定义为仅由索引集 M 对应的 X 的那些列组成的矩阵，β_M 是由 M 索引的 β 对应的 $|M|$ 维向量。令 $\theta_M=(\beta_M,\sigma_M^2)\subset\Theta_M$，其中 Θ_M 是第 M 个模型的参数空间。

在本文的后续研究中，允许协变量个数大于观测值个数，即 $p>n$，但此时回归模型的回归系数并不存在唯一的最小二乘估计。事实上，现有文献的各种高维变量选择问题的研究主要都是施加一定的稀疏性假设，使得高维问题在数学上可处理。尽管稀疏性假设有时并不一定正确，但就目前而言这是必要的。为了方便在数学框架下讨论高维变量选择问题，本文研究也是在一定的稀疏性假设下进行的。

令 M_o 为真实模型，假设 $M_o \in S$，即候选模型集合中包含真实模型，则模型选择的目标分为两种，第一种是识别，即在 2^p 个候选模型中识别出真实模型 M_o。第二种是预测，即并不需要识别出真实模型，只需要找到可以达到最优预测效果的模型即可，该模型可能只是真实模型的一个有效替代。相比较后者，前者的要求更高，因为完全识别出真实模型是很难的，事实上在实际案例中无法明确哪一个模型是真实模型，即 $M_o \notin S$。但后者所选的替代模型可能是个非常复杂的模型，即存在较多的冗余信息。为此，本文综合的来考虑两个目的，即在识别效果可比的前提下，是否可以达到较优的预测效果。

2.2. 参数的 Fiducial 密度

Fiducial 推断的基本思想源于对 Fisher 提出的 Fiducial 思想的理解。对于模型 M，首先将数据 y 和参数 θ_M 的关系表示为

$$y = G(U,\theta_M)$$

其中 G 是一个确定的函数，称为数据生成方程，U 是分布已知的随机部分，并且与参数 θ_M 独立。给定 y 的情况下，如果对于任意的 θ_M，G^{-1} 均存在，则

$$\theta_M = G^{-1}(y,u)$$

为了简单起见，本文只考虑简单情况，即数据生成函数 G 可逆，对于不可逆的情况，Hannig [24]也给出了相应的解决方法，在这里不多赘述。在已知 U 的分布下，多次重复抽样 U，通过结构方程的逆可以得到一组关于 θ_M 的随机样本，我们称之为 θ_M 的 Fiducial 样本，与之对应的密度称为 θ_M 的 Fiducial 密度，记为 $r(\theta_M|y)$。

对于线性回归模型 M，参数 $\theta_M = (\beta_M, \sigma_M^2)$ 的充分统计量为

$$(\hat{\beta}_M, \hat{\sigma}_M^2) = \left((X_M' X_M)^{-1} X_M' Y, \frac{RSS_M}{n-|M|} \right)$$

由抽样分布定理和基本的数理运算可得

$$\hat{\beta}_M = \beta_M + \sigma_M (X_M' X_M)^{-\frac{1}{2}} Z$$

$$(n-|M|)\hat{\sigma}_M^2 = \sigma_M^2 U$$

其中 $Z \sim N_{|M|}(0, I_{|M|})$，$U \sim \chi^2_{n-|M|}$，并且两者相互独立。当给定观测值 y 和 (z,u) 时，该结构方程的逆存在唯一解：

$$\beta_M = \hat{\beta}_M + \sigma_M (X_M' X_M)^{-\frac{1}{2}} z, \quad \sigma_M^2 = \frac{(n-|M|)\hat{\sigma}_M^2}{u}$$

因此 $\theta_M = (\beta_M, \sigma_M^2)$ 的联合 Fiducial 密度为

$$r(\theta_M|y) = \frac{\pi^{-\frac{|M|}{2}}}{2^{\frac{n}{2}}\Gamma\left(\frac{n-|M|}{2}\right)} RSS_M |X_M' X_M|^{\frac{1}{2}} (\sigma_M^2)^{-\frac{n}{2}-1} \exp\left(-\frac{\|y - X_M \beta_M\|}{2\sigma_M^2} \right)$$

而且容易求得 σ_M^2 的边际密度函数是 $IGa\left(\frac{n-|M|}{2}, \frac{RSS_M}{2}\right)$ 的概率密度函数，为

$$r(\sigma_M^2|y) = \frac{\left(\frac{RSS_M}{2}\right)^{\frac{n-|M|}{2}}}{\Gamma\left(\frac{n-|M|}{2}\right)} (\sigma_M^2)^{-\frac{n-|M|}{2}-1} \exp\left(\frac{RSS_M}{2\sigma_M^2} \right)$$

β_M 的边际密度函数为

$$\begin{aligned} r(\beta_M|y) &= \int_0^{+\infty} r(\theta_M|y) \mathrm{d}\sigma_M^2 \\ &= \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-|M|}{2}\right)} \frac{RSS_M^{\frac{n-|M|}{2}}}{\pi^{\frac{|M|}{2}}} |X_M' X_M|^{\frac{1}{2}} \left(\|y - X_M \beta_M\|_2^2 \right)^{-\frac{n}{2}} \\ &= \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-|M|}{2}\right)} \frac{RSS_M^{\frac{-|M|}{2}}}{\pi^{\frac{|M|}{2}}} |X_M' X_M|^{\frac{1}{2}} \left[1 + \frac{RSS_M \left\| X_M (\beta_M - \hat{\beta}_M) \right\|_2^2}{(n-|M|)^2} \right]^{-\frac{n}{2}} \end{aligned} \tag{2}$$

即 β_M 服从多元 t 分布 $t_{n-|M|}\left(\hat{\beta}_M, \frac{RSS_M}{n-|M|}\left(X_M' X_M\right)^{-1}\right)$。

2.3. 候选模型集压缩方法

在高维模型下经常会出现 $p \gg n$ 的情况，包括 Fiducial 模型选择在内的概率描述类模型择方法无法实现。其原因一方面是遍历整个候选模型集合来分别计算各自的模型概率度量具有极高的计算复杂度；另一方面，模型概率度量在整个候选模型集合上取值为 1，当候选模型集中包含了较多模型时，包括真实模型在内的每个模型所分摊的概率将会很小，这导致真实模型所反映出的信号变的很弱，从而无法达到模型选择的目的。为此合理且有效的压缩候选模型集是必要的。

Williams 和 Hannig [21]提出如下的 ε-容许集的概念。

给定 $\varepsilon > 0$，一个带有指标集 $M \subset \{1,2,\cdots,p\}$ 的系数向量 β_M 被称为是 ε-容许的，当且仅当 $h(\beta_M)=1$，其中

$$h(\beta_M) := I\left(\frac{1}{2}\left\|X'\left(X_M\beta_M - Xb_{\min}\right)\right\|_2^2 \geq \varepsilon\right)$$

其中 $b_{\min}$ 是在 $\|b\|_0 \leq |M|-1$ 的条件下，下述优化问题的解

$$\min_{b\in\mathbb{R}^p} \frac{1}{2}\left\|X'\left(X_M\beta_M - Xb\right)\right\|_2^2$$

候选模型集的压缩效果依赖于 ε，即一个较小的 ε 会产生一个相对稀疏的容许集合，但过大的 ε 可能会导致真实模型都变为不被容许。ε 偏大或偏小都会降低模型选择的精确度。在本文中，建议使用 Williams 和 Hannig [21]相同的默认规则来确定 ε：

$$\varepsilon_M = \Lambda_M \hat{\sigma}_M^2 \left(\frac{n^{0.51}}{9} + |M|\frac{\log(p\pi)^{1.1}}{9} + 1 - p_o\right)_+$$

其中 $\Lambda_M := \|H_M X\|_{Frobenius}^2$，$H_M := X_M\left(X_M' X_M\right)^{-1} X_M'$ 以及 $\hat{\sigma}_M^2 := \mathrm{RSS}_M / \left(n-|M|\right)$。$p_o$ 反映的是真实模型 M_o 的稀疏程度，我们用 Zhu 等人[25]提出的自适应最佳子集选择(Adaptive Best Subset Selection, ABESS)算法来获得真实模型 M_o 的稀疏度估计。不同数据集下的线性回归都可以适用该 ε 规则。

2.4. Fiducial 预测密度模型选择准则

类似于 Bayes 预测密度的定义，Long 和 Xu [17]将参数的 Fiducial 密度替代 Bayes 后验密度，给出了 Fiducial 预测密度定义。

对于线性回归模型(1)，如果我们已知 $|M|$ 维向量 x 的数值，且 x 独立于设计矩阵 X_M，则预测变量 z 与 y 都相互独立，并且 z 的分布为 $N\left(x'\beta_M, \sigma_M^2\right)$，即 z 的概率密度函数为

$$f\left(z|\theta_M, M\right) = \frac{1}{\sqrt{2\pi}\sigma_M}\exp\left(-\frac{\left(z-x'\beta_M\right)^2}{2\sigma_M^2}\right)$$

将 θ_M 的 Fiducial 密度 $r\left(\theta_M|y\right)$ 限制到那些被容许的集合上，得

$$r_\varepsilon\left(\theta_M|y\right) = r\left(\theta_M|y\right)\cdot h\left(\beta_M\right)$$

则 z 的 Fiducial 预测密度(简记为 FPD)为

$$\hat{f}\left(z|y,M\right)=\int f\left(z|\theta_M,M\right)r_{\varepsilon}\left(\theta_M|y\right)\mathrm{d}\theta_M$$

$$=\int_{\mathbb{R}^{|M|}}\int_0^{+\infty}h\left(\beta_M\right)\frac{\pi^{-\frac{|M|+1}{2}}RSS_M^{\frac{n-|M|}{2}}\left|X_M'X_M\right|^{\frac{1}{2}}}{2^{\frac{n+1}{2}}\Gamma\left(\frac{n-|M|}{2}\right)}\left(\sigma_M^2\right)^{-\frac{n+3}{2}}\exp\left\{-\frac{\left(z-x'\beta_M\right)^2+\left\|y-X_M\beta_M\right\|_2^2}{2\sigma_M^2}\right\}\mathrm{d}\sigma_M^2\,\mathrm{d}\beta_M$$

为了简化起见，先计算下面的积分

$$\int_0^{+\infty}\left(\sigma_M^2\right)^{-\frac{n+3}{2}}\exp\left\{-\frac{\left(z-x'\beta_M\right)^2+\left\|y-X_M\beta_M\right\|_2^2}{2\sigma_M^2}\right\}\mathrm{d}\sigma_M^2 \tag{3}$$

令 $t=\dfrac{\left(z-x'\beta_M\right)^2+\left\|y-X_M\beta_M\right\|_2^2}{2\sigma_M^2}$，则

$$(3)\text{式}=2^{\frac{n+1}{2}}\Gamma\left(\frac{n+1}{2}\right)\left[\left(z-x'\beta_M\right)^2+\left\|y-X_M\beta_M\right\|_2^2\right]^{-\frac{n+1}{2}}$$

因此

$$\hat{f}\left(z|y,M\right)=\int_{\mathbb{R}^{|M|}}h\left(\beta_M\right)\frac{RSS_M^{\frac{n-|M|}{2}}\left|X_M'X_M\right|^{\frac{1}{2}}\Gamma\left(\frac{n+1}{2}\right)}{\pi^{\frac{|M|+1}{2}}\Gamma\left(\frac{n-|M|}{2}\right)}\left[\left(z-x'\beta_M\right)^2+\left\|y-X_M\beta_M\right\|_2^2\right]^{-\frac{n+1}{2}}\mathrm{d}\beta_M$$

$$=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{n}{2}\right)}\int_{\mathbb{R}^{|M|}}h\left(\beta_M\right)\frac{1}{\sqrt{\left(z-x'\beta_M\right)^2+\left\|y-X_M\beta_M\right\|_2^2}}\left[\frac{\left(z-x'\beta_M\right)^2}{\left\|y-X_M\beta_M\right\|_2^2}+1\right]^{-\frac{n}{2}}r\left(\beta_M|y\right)\mathrm{d}\beta_M$$

$$\propto E\left\{h\left(\beta_M\right)\left(\left\|y-X_M\beta_M\right\|_2^2\right)^{-\frac{1}{2}}\left[\frac{\left\|y-X_M\beta_M\right\|_2^2}{\left(z-x'\beta_M\right)^2+\left\|y-X_M\beta_M\right\|_2^2}\right]^{\frac{n+1}{2}}\right\}$$

其中最后一行是关于 β_M 的期望。

Vehtari 等人[26]给出了一种对数逐点预测密度期望的贝叶斯留一估计，我们基于此提出对数逐点 Fiducial 预测密度估计，即对于模型 M 和数据集 y，

$$elfpd_{loo}=\sum_{i=1}^{n}\log\hat{f}\left(y_i|y_{-i},M\right)\triangleq h\left(y,M\right)$$

进而有

$$h\left(y,M\right)\propto\sum_{i=1}^{n}E\left\{h\left(\beta_M\right)\left(\left\|y_{-i}-X_{M,-i}\beta_M\right\|_2^2\right)^{-\frac{1}{2}}\left[\frac{\left\|y_{-i}-X_{M,-i}\beta_M\right\|_2^2}{\left\|y-X_M\beta_M\right\|_2^2}\right]^{\frac{n+1}{2}}\right\}$$

假设所有模型的先验分布 $\pi\left(M\right)$ 为均匀分布，由此我们定义一种新的模型不确定性度量

$$P_r\left(M|y\right)=\frac{h\left(y,M\right)\pi\left(M\right)}{\sum_{M'\in S}h\left(y,M'\right)\pi\left(M'\right)}\propto h\left(y,M\right)$$

于是基于 Fiducial 预测密度的选择准则给出，即

$$\hat{M} = \arg\max_{M \in S} P_r\left(M \middle| y\right) = \arg\max_{M \in S} h\left(y, M\right)$$

将该模型选择方法成为 Fiducial 预测密度模型选择方法，简记为 FPMC。

该方法也具有相合性，即在部分正则化条件下，真实模型 M_o 满足

$$P_r\left(M \middle| y\right) = \frac{h\left(y, M\right)\pi\left(M\right)}{\sum_{M' \in S} h\left(y, M'\right)\pi\left(M'\right)} \propto h\left(y, M\right)$$

其中 P_y 是指与观测值 Y 的抽样分布有关的概率度量。该性质的证明与 William 和 Hannig [21]文中证明 GFI 概率下的相合性的思路近乎相同，所以证明过程此处不展开讨论。

3. FPMC 的实现

仿照赵勇超[23]文中采用分组独立的 Metropolis-Hastings 算法实现 FMC 的思路，我们也采取该算法并将其应用进 FPMC。基本思路是首先通过(2)式引入潜变量从而得到一个解析表达式。然后抽取一个模型 M，并从 M 对应的多元 t 分布中抽取 N_{MC} 样本来计算期望的蒙特卡罗估计，记为 B。经过上述过程可以得到一个联合的马尔可夫链 $\left(M, B\right)$，通过仅关注 M 部分，便可从 M 的伪边际分布中获取样本。通过计算模型在稳定后的马尔可夫链中的后验概率，选择后验概率最大的模型 M 作为最优模型。算法的具体实现步骤见表 1：

Table 1. Distribution-independent MH algorithm
表 1. 分组独立的 MH 算法

输入： y， X， N_{step}， N_{burnin}， N_{MC}，一个初始的模型 M^0，一个提议概率函数 $q\left(\tilde{M} \middle| M^0\right)$，一个马尔科夫链 $S_{\text{MCMC}} = \phi$ 以及默认的 ε。

步骤一： 执行如下循环，对于 $k = 1, \cdots, N_{\text{step}}$，

(1) 自 M^0 开始通过随机游走的方式搜索下一个模型 $M \in S$，记为 $\tilde{M}$。

$$\tilde{M} = \begin{cases} M^0 \cup \{\text{一个新的协变量}\} & \omega = 1/3 \\ M^0 / \{\text{一个已存在的协变量}\} & \omega = 1/3 \\ M^0 / \{\text{一个已存在的协变量}\} \cup \{\text{一个新的协变量}\} & \omega = 1/3 \end{cases}$$

(2) 通过(4) 对 $\left(\beta_{M^0}, \sigma^2_{M^0}\right)$ 和 $\left(\beta_{\tilde{M}}, \sigma^2_{\tilde{M}}\right)$ 进行估计，分别记为 $\left(\hat{\beta}_{M^0,-i}, \hat{\sigma}^2_{M^0,-i}\right)$ 和 $\left(\hat{\beta}_{\tilde{M},-i}, \hat{\sigma}^2_{\tilde{M},-i}\right)$。

(3) 分别从分布

$$t_{n-|M|}\left(\hat{\beta}_M, \hat{\sigma}^2_M \left(X'_M X_M\right)^{-1} X'_M Y\right), \quad M \in \left\{M^0, \tilde{M}\right\}$$

抽取样本 $\beta^j_{M^0}$ 和 $\beta^j_{\tilde{M}}$， $j = 1, 2, \cdots, N_{MC}$。

(4) 对于所有 $j = 1, 2, \cdots, N_{MC}$，计算

$$\mathrm{H}^j_M \triangleq \sum_{i=1}^{n} h\left(\beta^j_M\right)\left(\left\|y_{-i} - X_{M,-i}\beta^j_M\right\|_2^2\right)^{-\frac{1}{2}} \left[\frac{\left\|y_{-i} - X_{M,-i}\beta^j_M\right\|_2^2}{\left\|y - X_M\beta^j_M\right\|_2^2}\right]^{\frac{n+1}{2}}, \quad M \in \left\{M^0, \tilde{M}\right\}$$

其中 β^j_M 可由显式 L_0 最小化算法计算。

(5) 计算(17)式的蒙特卡洛估计

$$\hat{P}_r\left(M \middle| y\right) \propto \frac{1}{N_{MC}} \sum_{j=1}^{N_{MC}} \mathrm{H}^j_M, \quad M \in \left\{M^0, \tilde{M}\right\}$$

续表

(6) 如果 $\min\left\{\dfrac{\hat{P}_r\left(\tilde{M}\mid y\right)q\left(M^0\mid\tilde{M}\right)}{\hat{P}_r\left(M^0\mid y\right)q\left(\tilde{M}\mid M^0\right)},1\right\}>u^*, u^*\sim U(0,1)$，

那么 $M^0=\tilde{M}$ 并且 $S_{\text{MCMC}}=\left\{S_{\text{MCMC}},\tilde{M}\right\}$；

否则 $M^0=M^0$ 并且 $S_{\text{MCMC}}=\left\{S_{\text{MCMC}},\phi\right\}$。

终止循环。

步骤二： 舍弃马尔可夫链 S_{MCMC} 上前 N_{burnin} 个样本。计算所有模型在剩余 $N_{\text{step}}-N_{\text{burnin}}$ 个样本上的后验概率。选择具有最高后验概率的模型作为最优模型 M_{best}。

步骤三： 输出最优模型 M_{best}。

4. 数值模拟

本节对提出的 FPMC 进行数值模拟，并将其模拟结果与 Lasso 和 EN 进行比较。比较的指标主要包括两方面：一方面是变量识别的准确性，主要有积极变量识别率的平均(AEIR)；消极变量识别率的平均(IEIR)；稀疏度估计误差(SLE)。另一方面是预测精度，用到的是均方根误差(RMSE)。设 M_o 和 M_o^c 为真实模型 M_o 中包含的积极效应和消极效应所构成的集合，M_i 为第 i 次试验中选择的最优模型，N 为重复实验次数。则各个指标的定义如下

$$\text{SLE}=\frac{1}{N}\sum_{i=1}^{N}\left|\left|M_i\right|-\left|M_o\right|\right|,\quad \text{RMSE}=\sqrt{\frac{1}{n_{test}}\sum_{i=1}^{n_{test}}\left(y_i-\hat{y}_i\right)^2}$$

$$\text{AEIR}=\frac{1}{N}\sum_{i=1}^{N}\frac{\left|M_i\cap M_o\right|}{\left|M_o\right|}\times 100\%,\quad \text{IEIR}=\frac{1}{N}\sum_{i=1}^{N}\frac{\left|M_i\cap M_o^c\right|}{\left|M_o^c\right|}\times 100\%$$

AEIR 越大，说明所选模型中包含的积极变量越多，故 AEIR 越大越好(越接近 1 越好)；IEIR，RMSE 越小越好，SLE 越接近 0 越好。

4.1. 模拟 1

本例与 Williams 和 Hannig [21]的研究中的模拟设定 1 保持一致。从(1)式中生成数据，其中设计阵 $X_{n\times p}$ 服从多元正态分布 $N_p\left(\mathbf{0}_p,\Sigma\right)$。真实模型的参数向量为

$$\beta_{M_o}=\left\{-1.5,-1,-0.8,-0.6,0.6,0.8,1,1.5,0_{p-8}\right\}$$

其中 $p\in\left\{40,80,120\right\}$。$\Sigma\left(=\left(\sigma_{ij}\right)_{p\times p}\right)$ 被设定为单位矩阵，训练集大小为 $n=100$，测试集大小为 $n_{\text{test}}=30$，误差项的方差 σ^2 固定为 1。在赵勇超[23]的模拟 1 中已通过 ABESS 算法得出了此时 p_o 的估计值取 $p_o=8$ 最合适。下面的表 2、表 3 和表 4 给出了 FPMC、LASSO 和 EN 的表现。

Table 2. The performances of three methods when p = 40
表 2. p = 40 时三种方法的表现

	SLE	RMSE	AEIR	IEIR
FPMC	1.19	3.6589	0.936	0.053
Lasso	0.53	3.6743	0.828	0.0597
EN	1.23	3.6740	0.875	0.0698

Table 3. The performances of three methods when $p = 80$
表 3. $p = 80$ 时三种方法的表现

	SLE	RMSE	AEIR	IEIR
FPMC	1.10	3.7422	0.930	0.023
Lasso	0.62	3.7501	0.821	0.028
EN	1.52	3.7592	0.884	0.034

Table 4. The performances of three methods when $p = 120$
表 4. $p = 120$ 时三种方法的表现

	SLE	RMSE	AEIR	IEIR
FPMC	0.95	3.6385	0.927	0.014
Lasso	1.00	3.6736	0.845	0.020
EN	1.87	3.6772	0.896	0.024

模拟 1 结果表明，无论变量个数 p 取 40，80 还是 120，FPMC 的 RMSE 总是最小的，即无论 p 小于 n 还是大于 n，我们方法的预测准确度都是最高的。随着 p 值的增加新方法的 SLE 数值越来越接近 0，并且当 p 大于 n 时 FPMC 是三种方法 SLE 值最小的。从被选模型的识别效果和可解释性方面来看，FPMC 的 AEIR 值在三种情形下都是最大值、IEIR 值都是最小值，即 FPMC 方法能更精确地识别积极效应、更精确地压缩消极效应。综合来说，本文提出的 FPMC 在预测和识别方面都更胜一筹。

4.2. 模拟 2

采用与 Williams 和 Hannig [21]的研究中模拟设定 2 相同的数据生成方式，其中 $n = 30$。数据生成过程为

$$Y \sim N_n\left(1\cdot x^{(1)} + 1\cdot x^{(2)} + \cdots + 1\cdot x^{(9)}, I_n\right)$$

其中 $x^{(1)}, x^{(2)}, x^{(3)} \overset{i.i.d}{\sim} N_n(0, I_n)$，$x^{(4)} \sim N_n\left(0.25\cdot x^{(1)}, 0.1^2 I_n\right)$，$x^{(5)} \sim N_n\left(0.50\cdot x^{(2)}, 0.1^2 I_n\right)$，$x^{(6)} \sim N_n\left(-0.75\cdot x^{(3)}, 0.1^2 I_n\right)$，$x^{(7)} \sim N_n\left(1\cdot x^{(1)} + 1\cdot x^{(3)}, 0.1^2 I_n\right)$，$x^{(8)} \sim N_n\left(1\cdot x^{(2)} - 1\cdot x^{(3)}, 0.1^2 I_n\right)$，$x^{(9)} \sim N_n\left(1\cdot x^{(1)} + 1\cdot x^{(2)} + 1\cdot x^{(3)}, 0.1^2 I_n\right)$。

由此我们建立线性模型，经过 10 折交叉验证我们取 FPMC 方法中的 p_o 估计值为 6。进行 1000 次重复模拟实验，三种方法 RMSE 和模型长度的数值如下：

Table 5. The performances of three methods in Simulation 2
表 5. 模拟 2 中三种方法的表现

	FPMC	Lasso	EN
RMSE	4.6854	5.3463	5.6291
Length of model	6.1	7.0	6.9

由表 5 可知 FPMC 方法的 RMSE 数值最小，即它的预测准确度最高。而且它在保证准确度的情况下没有过多的选择变量，控制了所选模型的复杂程度。

5. 实例分析

5.1. 实例一

活塞拍击的声音是一种由活塞二次运动引起的发动机噪音。为了降低活塞拍击的噪音，选取了 6 种

对噪音影响较大的因素进行分析，希望通过改变这 6 种因素达到减少噪音的目的。6 种因素分别为活塞和气缸套之间的设定间隙 x_1，峰值压力位置 x_2，裙部长度 x_3，裙部轮廓形状 x_4，裙部椭圆度 x_5，活塞销偏置 x_6。数据集来源于 Huang 等[27]并进行了归一化处理，共包含 100 个观测样本，每个样本有 6 个输入变量，模型中可能包含所有的线性主效应、二次主效应以及正交多项式编码下的所有双因素相互作用，因此共有 72 个基变量。我们从 100 个样本中取 80 个作为训练集，剩下的 20 个做测试集，模拟结果在表 6 中给出。

Table 6. Data simulation results of a piston slap noise example
表 6. 活塞拍击噪声实例的数据模拟结果

	FPMC	Lasso	EN
RMSE	0.5088	0.6462	0.6368
Length of model	16	39	32

由表 6 可以得出，FPMC 所选的模型长度不仅比 Lasso 和 EN 的要小，RMSE 数值也是三种方法中最小的。总体来看，FPMC 方法能够在有效简化模型同时具有较好的预测效果。

5.2. 实例 2

关于影响维生素 B 生产速率的基因位点识别，是一个典型的基于超高维线性回归模型的模型选择问题。该数据集首次公开于 Buhlmann 等人的研究，其中包含 71 个观测值和 4088 个基因表达，其中响应变量为枯草芽孢杆菌的维生素 B 生产速率的对数。我们采用线性回归模型来探索数据集中响应变量与协变量之间的关系，因此该数据集符合模型 2-(1)的设定。首先用 ABESS 算法得到该真实数据集的 p_o 的估计值为 $\hat{p}_o = 6$。然后用 BESS (Best Subset Selection)算法进行初筛并保留 $p = 71$ 个基因表达进行分析。需要注意的是，BESS 和 ABESS 算法都源自 Zhu 等人[25]的研究。区别在于前者需要指定稀疏度(在这里稀疏度指定为 $p = 71$)，而后者是依靠数据驱动的自适应版本。对于最终保留的 71 个基因表达，分别采用 FPMC、Lasso、EN 来进行变量选择，基本的设置与第四节保持一致。

Table 7. Data simulation results of a gene recognition of vitamin B example
表 7. 维生素 B 基因识别实例的数据模拟结果

	FPMC	Lasso	EN
RMSE	0.292	0.331	0.324
Length of model	8	10	9

由表 7 可以得出，FPMC 所选的模型长度不仅比 Lasso 和 EN 的要小，RMSE 数值也是三种方法中最小的。总体来看，FPMC 方法表现更好。

6. 结论

本文提出了高斯回归模型的 Fiducial 预测模型选择方法 FPMC，并同时给出它的理论步骤与实践算法。之后其与 Lasso 和 EN 相比较，通过两个模拟和两个实例我们得到三种方法中最好的是 FPMC，不仅预测误差总是最小的，而且模型拟合精度和复杂度都表现良好。

参考文献

[1] Tibshirani, R. (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society Series*

B: *Statistical Methodology*, **58**, 267-288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

[2] Fan, J. and Li, R. (2001) Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association*, **96**, 1348-1360. https://doi.org/10.1198/016214501753382273

[3] Zou, H. and Hastie, T. (2005) Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society Series B*: *Statistical Methodology*, **67**, 301-320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

[4] Zhang, C. (2010) Nearly Unbiased Variable Selection under Minimax Concave Penalty. *The Annals of Statistics*, **38**, 894-942. https://doi.org/10.1214/09-aos729

[5] Schwarz, G. (1978) Estimating the Dimension of a Model. *The Annals of Statistics*, **6**, 461-464. https://doi.org/10.1214/aos/1176344136

[6] Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and Van Der Linde, A. (2002) Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society Series B*: *Statistical Methodology*, **64**, 583-639. https://doi.org/10.1111/1467-9868.00353

[7] Maity, A.K., Basu, S. and Ghosh, S. (2021) Bayesian Criterion-Based Variable Selection. *Journal of the Royal Statistical Society Series C*: *Applied Statistics*, **70**, 835-857. https://doi.org/10.1111/rssc.12488

[8] Akaike, H. (1973) Maximum Likelihood Identification of Gaussian Autoregressive Moving Average Models. *Biometrika*, **60**, 255-265. https://doi.org/10.1093/biomet/60.2.255

[9] Takeuchi, K. (1976) Distribution of Informational Statistics and a Criterion of Model Fitting. *Suri-Kagaku*, **153**, 12-18.

[10] Mallows, C.L. (1973) Some Comments on CP. *Technometrics*, **15**, 661-675. https://doi.org/10.1080/00401706.1973.10489103

[11] Allen, D.M. (1974) The Relationship between Variable Selection and Data Agumentation and a Method for Prediction. *Technometrics*, **16**, 125-127. https://doi.org/10.1080/00401706.1974.10489157

[12] Fisher, R.A. (1922) On the Mathematical Foundations of Theoretical Statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, **222**, 309-368.

[13] Dempster, A.P. (2008) The Dempster-Shafer Calculus for Statisticians. *International Journal of Approximate Reasoning*, **48**, 365-377. https://doi.org/10.1016/j.ijar.2007.03.004

[14] Martin, R., Zhang, J. and Liu, C. (2010) Dempster-Shafer Theory and Statistical Inference with Weak Beliefs. *Statistical Science*, **25**, 72-87. https://doi.org/10.1214/10-sts322

[15] Xie, M. and Singh, K. (2013) Confidence Distribution, the Frequentist Distribution Estimator of a Parameter: A Review. *International Statistical Review*, **81**, 3-39. https://doi.org/10.1111/insr.12000

[16] Majumder, A.P. and Hannig, J. (2016) Higher Order Asymptotics of Generalized Fiducial Distribution.

[17] Long, Y. and Xu, X. (2020) Classification by Fiducial Predictive Density Functions. *Communications in Statistics—Theory and Methods*, **51**, 5187-5203. https://doi.org/10.1080/03610926.2020.1836218

[18] Hannig, J. and Lee, T.C.M. (2009) Generalized Fiducial Inference for Wavelet Regression. *Biometrika*, **96**, 847-860. https://doi.org/10.1093/biomet/asp050

[19] 李涵, 赵建昕, 王晓, 李新民. 计算机试验下 Kriging 模型选择的比较[J]. 应用数学进展, 2021, 10(3): 694-700.

[20] 张淑芹, 李涵, 李新民. 基于 Fiducial 推断的 Kriging 模型选择[J]. 应用数学进展, 2024, 13(2): 684-691.

[21] Williams, J.P. and Hannig, J. (2019) Nonpenalized Variable Selection in High-Dimensional Linear Model Settings via Generalized Fiducial Inference. *The Annals of Statistics*, **47**, 1723-1753. https://doi.org/10.1214/18-aos1733

[22] Williams, J.P., Xie, Y. and Hannig, J. (2022) The EAS Approach for Graphical Selection Consistency in Vector Autoregression Models. *Canadian Journal of Statistics*, **51**, 674-703. https://doi.org/10.1002/cjs.11726

[23] Zhao, Y., Li, X. and Liang, H. (2022) Fiducial Marginal Likelihood-Based Variable Selection for High-Dimensional Regression (in Chinese). *Science China Mathematics*, **52**, 1-21.

[24] Hannig, J., Iyer, H., Lai, R.C.S. and Lee, T.C.M. (2016) Generalized Fiducial Inference: A Review and New Results. *Journal of the American Statistical Association*, **111**, 1346-1361. https://doi.org/10.1080/01621459.2016.1165102

[25] Zhu, J., Wen, C., Zhu, J., Zhang, H. and Wang, X. (2020) A Polynomial Algorithm for Best-Subset Selection Problem. *Proceedings of the National Academy of Sciences*, **117**, 33117-33123. https://doi.org/10.1073/pnas.2014241117

[26] Vehtari, A., Gelman, A. and Gabry, J. (2016) Practical Bayesian Model Evaluation Using Leave-One-Out Cross-Validation and Waic. *Statistics and Computing*, **27**, 1413-1432. https://doi.org/10.1007/s11222-016-9696-4

[27] Huang, H., Lin, D.K.J., Liu, M. and Zhang, Q. (2019) Variable Selection for Kriging in Computer Experiments. *Journal of Quality Technology*, **52**, 40-53. https://doi.org/10.1080/00224065.2019.1569959.