

doi: 10.11823/j.issn.1674-5795.2022.01.02

最大熵原理及改进方法的研究现状

黄乾坤, 吴娅辉

(航空工业北京长城计量测试技术研究所, 北京 100095)

摘要: 熵作为信息论中的一个基本概念受到了广泛的关注与研究。以熵为基础的最大熵原理也在众多学者的研究与应用中不断发展, 逐渐形成了自有的理论体系并得到了广泛应用, 例如在使用贝叶斯方法进行测量不确定度评定时, 最大熵原理可以估计先验分布。本文以 Shannon 熵为主要对象, 对目前研究中涉及的不同约束下最大熵原理进行了归纳, 并整理为形式较为统一的模型。论述了利用转换函数法、密度核估计法等改善传统最大熵原理不足的方法, 具体介绍了它们的改进思路、理论模型及应用特点。最后结合实践, 从约束选择、评价指标和优化算法等方面对最大熵原理改进方法进行总结, 为最大熵原理的进一步研究及应用起到推动作用。

关键词: 最大熵原理; 优化求解; 矩约束; 秩约束; 转换函数; 密度核估计

中图分类号: TB94; TP391.41 **文献标识码:** A **文章编号:** 1674-5795(2022)01-0009-09

Research status of the maximum entropy principle and improved methods

HUANG Qiankun, WU Yahui

(Changcheng Institute of Metrology & Measurement, Beijing 10095, China)

Abstract: Entropy, as a basic concept in information theory, has been extensively studied. The principle of maximum entropy based on entropy has been constantly developing in the research and application of scholars, and has gradually formed its own theoretical system which has been widely applied. For example, when the Bayesian method is used to evaluate measurement uncertainty, the principle of maximum entropy can be used to estimate the prior distribution. In this paper, Shannon entropy is taken as the main research object, and the maximum entropy principle under different constraints involved in the current research is summarized and sorted out into a relatively unified model. This paper discusses the methods to improve the traditional maximum entropy principle using transformation function method and density kernel estimation method, and introduces in details the ideas for improvement, theoretical models and application characteristics. Finally, combined with practice, the improved methods of maximum entropy principle are summarized from the aspects of constraint selection, evaluation index and optimization algorithm, which will promote further research and application of maximum entropy principle.

Key words: maximum entropy principle; optimization solution; moment constraints; rank constraints; conversion function; density kernel estimation

0 引言

1865 年德国物理学家 Clausius 首先提出了“熵”的概念, 1948 年 Shannon 提出了“信息熵”的概念并将其作为热力学的度量工具^[1], 之后学

者们将“熵”的概念进一步推广到其他领域, 形成了所谓的“广义熵”。1957 年, 统计物理学家 Jaynes 利用非热力学场的熵增原理提出了最大熵的概念^[2-3], 并给出了采用最大熵原理定量解决问题的一般方法, 同时也利用最大熵原理证明了统计

收稿日期: 2021-10-25; 修回日期: 2021-11-30

基金项目: 国家自然科学基金(51775526)

引用格式: 黄乾坤, 吴娅辉. 最大熵原理及改进方法的研究现状[J]. 计测技术, 2022, 42(1): 9-17.

Citation: HUANG Q K, WU Y H. Research status of the maximum entropy principle and improved methods[J]. Metrology & measurement technology, 2022, 42(1): 9-17.



力学的一些分布函数。在1988年巴西物理学家C. Tsallis定义离散的Tsallis熵^[4]之前,熵的讨论主要集中在广延统计学中,最大熵原理也是以Shannon熵为基础的。而由于Tsallis熵的非广延性,使得熵进一步推广到了非广延统计学中,学者们也开始了非广延统计学中Tsallis熵的研究^[5-6]。这些研究使信息理论与统计物理建立起了相互联系,也使熵的概念从热力学中分离了出来。

应用最大熵原理的基本思想是:当只掌握有限信息时,应尽可能少地引入主观信息,也就是使熵最大化,并在满足已有信息作为熵约束的条件下,推导出随机变量集的概率分布。该方法被广泛应用于金融^[8]、图像处理^[9]、气象^[10]等领域中,并展现出了巨大的实用价值和研究意义。近年来,最大熵原理也被引入计量领域,许多学者利用最大熵原理建立测量数据的概率密度分布模型,或进一步与贝叶斯方法结合,证明了最大熵原理在测量不确定度评定上的可用性和有效性^[11-14]。

本文对存在和应用范围较广的Shannon最大熵原理进行阐述,介绍了最大熵原理改进方法的理论模型、计算公式、使用特点,并给出了最大熵原理约束选择、评价指标和优化算法等方面的总结与见解,为最大熵原理的进一步研究及应用提供指导与借鉴。

1 最大熵原理

在广延统计学和非广延统计学中,由于“熵”的定义不同,所推导得到的最大熵原理的内容也会不同;对于同一种“熵”,采用的约束类型不同,得到的最大熵原理的内容也会不同。本节首先介绍在两类统计学中各自主流“熵”下的最大熵模型,随后围绕Shannon熵介绍分数阶矩约束和秩约束下的最大熵原理。之后对求解最大熵优化问题的优化算法进行了阐述,经研究发现,优化算法的性能会对最大熵原理估计得到的结果产生影响,但由于其不是本文的论述重点,因此只进行必要的说明和整体的概述。

1.1 两种最大熵原理

1.1.1 Jaynes最大熵原理

Jaynes最大熵原理是建立在Shannon熵上的。设 X 为离散型随机变量,取值为 $\{x_1, x_2, \dots, x_n\}$,相应的概率为 $P(X = x_i) = p_i (i = 1, 2, \dots, n)$,则随机

变量 X 的Shannon熵^[6]为

$$H(x) = - \sum_{i=1}^n p_i \log p_i \quad (1)$$

且有

$$\begin{aligned} 0 \leq p_i \leq 1, i = 1, 2, \dots, n \\ \sum_{i=1}^n p_i = 1 \end{aligned} \quad (2)$$

当 X 为连续型随机变量时,设其概率密度函数为 $f(x)$,则对应的Shannon熵^[6]为

$$H(x) = - \int_{\Theta} f(x) \ln f(x) dx \quad (3)$$

式中: Θ 为积分空间,并有 $\int_{\Theta} f(x) dx = 1$ 。

由上述 $H(x)$ 的定义可知,熵的大小与概率的分布有关,在没有其他约束条件时,概率取均匀分布可以使 $H(x)$ 最大,但这显然不具备太大的实用意义。为获得更准确的概率分布,需要从所掌握的信息中提取出 p_i 或 $f(x)$ 要满足的其他条件。选择随机变量的整数阶样本原点矩作为约束条件,对于连续型随机变量,由此构造的优化问题^[15]为

$$\begin{aligned} \max \quad & H(x) = - \int_{\Theta} f(x) \ln f(x) dx \\ \text{s.t.} \quad & \int_{\Theta} f(x) dx = 1 \\ & \int_{\Theta} x^i f(x) dx = M_i, i = 1, 2, \dots, m \end{aligned} \quad (4)$$

式中: M_i 为第 i 阶样本原点矩; m 为采用的最高的样本原点矩的阶数; $s.t.$ 为使满足符号。对于式(4)的有约束优化问题,可采用拉格朗日乘子法将其转化为无约束优化问题,并可得到在熵值取得最大时, $f(x)$ 的解析式^[15]为

$$\hat{f}(x) = \exp(\lambda_0 + \sum_{i=1}^m \lambda_i x^i) \quad (5)$$

式中: λ_i 为采用拉格朗日乘子引入的待求解参数。将式(4)中的约束条件代入式(5)中,可进一步建立参数 $\{\lambda_1, \lambda_2, \dots, \lambda_m\}$ 解的方程组^[15],即

$$M_i = \frac{\int_{\Theta} x^i \exp(\sum_{i=1}^m \lambda_i x^i) dx}{\int_{\Theta} \exp(\sum_{i=1}^m \lambda_i x^i) dx} \quad (6)$$

记残差为 R_i ,可将上述方程组改写为非线性优化问题^[15],即

$$R_i = 1 - \frac{\int_{\Theta} x^i \exp(\sum_{i=1}^m \lambda_i x^i) dx}{M_i \int_{\Theta} \exp(\sum_{i=1}^m \lambda_i x^i) dx} \quad (7)$$

采用最大熵原理求解概率密度函数的问题转化为为了参数寻优问题,当残差平方和最小时,得到的 λ_i 为最优解,此时对应的熵也是在可求得解的空间中最大的。

这种建立在Shannon熵上,以整数阶样本原点矩为约束的Jaynes最大熵原理是最基本,且应用最为广泛的熵原理,一般称其为传统(经典)最大熵原理。

1.1.2 Tsallis最大熵原理

非广延统计学中,Tsallis熵的应用最为广泛,在非广延统计学中扮演着十分重要的角色。离散型的Tsallis熵^[3]为

$$S_q^T(X) = \frac{1}{1-q} \left(\sum_{i=1}^n P_i^q - 1 \right) \quad (8)$$

式中: q 为非广延参数,用于衡量系统非广延程度的程度。当 $q \rightarrow 1$ 时,Tsallis熵转化为Shannon熵^[16],即

$$\lim_{q \rightarrow 1} S_q^T(X) = - \sum_{i=1}^n p_i \ln p_i \quad (9)$$

Tsallis最大熵原理是在Jaynes最大熵原理的基础上发展而来的。为方便论述,采取如下记法^[16]

$$\ln_q(x) = \frac{x^{1-q} - 1}{1-q} \quad (10)$$

$$\exp_q(y) = [1 + (1-q)y]^{1/q}$$

式中: y 与 x 为变量; $\exp_q(\ln_q x) = \ln_q(\exp_q x) = x$ 。

于是离散型的Tsallis熵可以改写为如下形式^[16]

$$S_q^T(x) = \sum_{i=1}^n p_i \ln_q \frac{1}{p_i} \quad (11)$$

式中: $\ln_q(p_i) = \frac{p_i^{1-q} - 1}{1-q}$ 。

对于连续的随机变量,Tsallis熵为^[17]

$$S^q(x) = \int_{\Theta} f(x) \ln_q \frac{1}{f(x)} dx \quad (12)$$

以连续的随机变量为例,在分布未知时,选择随机变量的样本原点矩作为约束条件,构造的Tsallis熵最大熵优化问题为^[17]

$$\begin{aligned} \max \quad & S^q(x) = \int_{\Theta} f(x) \ln_q \frac{1}{f(x)} dx \\ \text{s.t.} \quad & \int_{\Theta} f(x) dx = 1 \\ & \int_{\Theta} x^i f(x) dx = M_i, \quad i = 1, 2, \dots, m \end{aligned} \quad (13)$$

同样采用拉格朗日乘子法将其转化为无约束优化问题,可得到在熵值最大时,概率密度函数的解析式^[17]为

$$\hat{f}(x) = \frac{\exp_q(-1)}{\exp_q(-\lambda_0 - \sum_{i=1}^m \lambda_i x^i)} \quad (14)$$

吴贤东^[17]给出了当随机变量概率密度分布已知时,多约束条件下Tsallis最大熵的推导过程,并对常见的分布(正态分布、对数正态分等)进行了Tsallis最大熵分布的求解。

1.2 约束的选择

1.2.1 基于分数阶矩约束的最大熵方法

Tagliani等在研究Hausdorff矩问题时,采用少数分数阶矩作为最大熵的约束条件来求解有限的Hausdorff矩问题,结果显示利用较少分数阶矩就能保证结果的精度,还可以修正因为采用大量积分矩而产生的畸变^[18]。

以分数阶矩为约束条件,概率密度函数 $f(x)$ 具有如下形式^[18]

$$f(x) = \exp(-\lambda_0 - \sum_{i=1}^m \lambda_i x^{a_i}) \quad (15)$$

与经典最大熵原理类似,选择分数阶矩为约束条件的最大熵优化问题^[18]为

$$\begin{aligned} \max \quad & H(x) = - \int_{\Theta} f(x) \ln f(x) dx \\ \text{s.t.} \quad & \int_{\Theta} f(x) dx = 1 \\ & \int_{\Theta} x^{a_i} f(x) dx = E[x^{a_i}], \quad i = 1, 2, \dots, m \end{aligned} \quad (16)$$

式中: a_i 为统计矩阶数,是任意分数(或实数); $E[x^{a_i}]$ 为 x 的分数阶统计矩。对该优化问题进行求解便可得到式(15)的具体表达式。

Zhang X等^[19]将上述方法引入到结构可靠性分析中,结合降维处理实现了功能函数的分数阶矩估计,并进一步完善了分数阶矩约束下的最大熵原理求解概率密度函数的方法。

1.2.2 基于秩约束的最大熵方法

传统最大熵原理,如1.1.1节所述,通常采用随机变量的整数阶样本原点矩为约束条件来构造优化问题。虽然样本原点矩是总体矩的替代,但其仍包含了样本的统计信息。二阶矩反映了样本的均值和方差信息;三阶矩反映了样本的偏态信息;四阶矩反映了样本的峰态信息^[20]。通常在样本量较大时,选择二到五阶矩就能较好地估计样

本的概率密度分布。但在数据样本量很小时，样本矩所能提供的信息不够丰富和全面，所求得概率密度分布可能与实际存在较大偏差。吕文^[20]介绍了一种在小样本情况下采用秩来构造约束条件的最大熵方法，其中秩的分布^[20]为

$$f(r_i) = \frac{n!}{(i-1)!(n-i)!} r_i^{i-1} (1-r_i)^{n-i} \quad (17)$$

式中： r_i 为样本容量为 n 的顺序样本的 i 阶秩，其定义为

$$r_i = \int_{-\infty}^{o_i} f_x(x) dx \quad (18)$$

式中： o_i 为第 i 阶的顺序统计量； o_i 和 r_i 均为随机变量。

由于 $f(r_i)$ 只与 i 和 n 有关，且随着 n 的增加，其形状会变得瘦而高。为便于处理，利用 $f(r_i)$ 的这个特点对其两端进行等面积的截断处理，使其剩下的面积为 q ，即需对 $F(o_i)$ (F 为对 f 的积分)增加上下界^[20]

$$b_l \leq r_i \leq b_h \quad (19)$$

其中， b_l 和 b_h 由以下公式求出^[20]

$$\int_0^{b_l} f(r_i) dr_i = \frac{1-q}{2} \quad (20)$$

$$\int_{b_h}^1 f(r_i) dr_i = \frac{1-q}{2} \quad (21)$$

因此，当给定 q 时， b_l 和 b_h 也就相应确定了。根据以上边界条件，可得到基于秩约束的最大熵优化问题^[20]

$$\max H(x) = -\int_l^h f(x) \ln f(x) dx$$

$$s.t. \quad \int_l^h f(x) dx = 1 \quad (22)$$

$$b_l \leq \int_l^{d_i} f(x) dx \leq b_h, \quad i = 1, 2, \dots, n$$

式中： d_i 为顺序样本第 i 阶变量的 x 值； n 为样本容量。对于上述优化问题，无法通过数学推导得到 $f(x)$ 的理论分布，但可以采取合理近似法，将 $f(x)$ 拟合为与传统最大熵一样的指数函数的形式

$$f(x) = \exp(\lambda_0 + \sum_{i=1}^m \lambda_i x^i) \quad (23)$$

联合式(22)和式(23)，再经优化求解便可得到 $f(x)$ 。

1.3 优化求解方法

从1.1和1.2节中可以看出，无论基于何种熵、何种约束，优化问题的求解是最大熵原理应

用的关键。余秀美等^[21]采用基于非线性最小二乘法信赖域模型进行最大熵优化问题的求解；刘钰等^[22]在研究基于密度核估计的最大熵方法中，采用序列二次规划方法进行优化问题的求解；姜瑞等^[23]在研究贝叶斯方法在测量不确定度评定上的应用时，利用爬山算法实现了最大熵优化问题的求解；彭颖等^[24]通过仿真数据和高层建筑地震数据验证了遗传算法求解最大熵原理下概率密度函数的有效性。

最大熵优化问题是一个典型非线性方程组(Nonlinear Equation Systems, NESs)，因此针对NESs求解的算法基本都能用于对它的求解。由于NESs广泛存在于网络通信、机械制造、神经网络、图形处理等众多领域^[25-26]，其求解算法发展已较为成熟，并可大致分为两类：基于梯度信息进行迭代的传统优化方法和基于群体优化的智能优化算法^[27]。

传统优化算法包括牛顿法、共轭梯度法、最小二乘信赖域法、拟牛顿法、最速下降法等^[28-29]。传统优化算法有着完备的数学理论依据，并且由于起源较早，研究和应用都较为成熟。但它们的不足也很明显：比较依赖初始点的选取；因算法迭代需要梯度信息，只能求解可微函数问题；对于多极值问题，容易陷入局部最优解问题。这些不足限制了传统优化算法在复杂问题求解领域的应用。

智能优化算法包括遗传算法^[30-31]、模拟退火算法^[32]、粒子群算法^[33]等。智能优化算法通常从多点出发进行解的搜索，具有隐形的并行性。相比于传统优化算法，它们对初始点的选取往往不敏感；迭代过程也不使用梯度信息，对不可微NESs依旧适用；求解范围更广，不易陷入局部最优问题；并有着高效运算性和良好的鲁棒性。因此近年来智能优化算法受到了越来越多的关注，逐渐成为学者们的研究重点。然而这类算法也存在一些需要进一步解决的问题，包括：数学理论不够完善；算法收敛性的证明仍需继续研究；算法收敛速度还不够理想；NESs转化为约束问题的方法对求解的影响；多种算法混合应用以提高求解效率的研究等^[27]。罗宇婷等^[34]对智能优化算法进行了较为详尽的归纳和细分，并通过测试函数对几种应用广泛的主流算法进行了横向性能比较。

2 最大熵原理改进方法

2.1 基于转换函数法的最大熵原理

最大熵方法的计算过程涉及大量积分计算，而在使用计算机计算这些往往需要通过数值积分方法计算的积分时，积分区间的上下限无法像理论上那样取到无穷大。在实际计算时，如失效概率的计算，通常会选取一个以均值为中心，以十倍方差为半宽为的区间 $[\mu - 10\sigma, \mu + 10\sigma]$ 来进行近似计算，而这会直接引入截断误差。在概率密度函数形状集中于中心线附近时，截断处理所截掉的面积占比很小，即引入的误差很小，对结果的影响可以忽略不计。但当概率密度函数形状平而宽时，截断处理引起的误差往往是不能忽略的，甚至将导致计算结果不收敛。为了解决这个问题，李昊燃等^[35]提出了一种基于转换函数的最大熵方法。

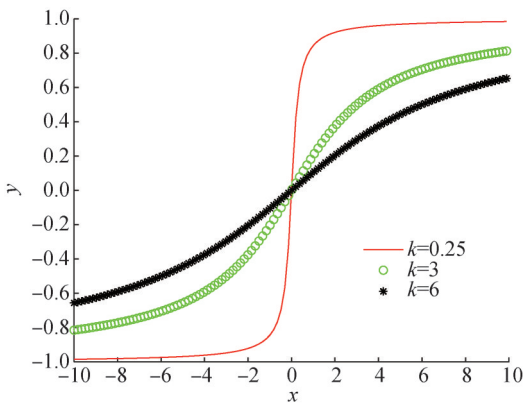


图1 不同 k 值下 x 与 y 的变换关系示意图^[35]

Fig.1 Schematic diagram of transformation relationship between x and y under different k values^[35]

首先为了将概率密度函数的定义域由 $(-\infty, +\infty)$ 转化为有限区间，引入具有单调性的三角函数来对定义域进行非线性变换。引入的三角变换如下^[35]

$$y = \frac{2}{\pi} \operatorname{acr} \tan(x/k) \quad (24)$$

式中： $x \in (-\infty, +\infty)$ ， $y \in (-1, 1)$ 。其逆变换为

$$x = g^{-1}(y) = k \tan\left(\frac{\pi}{2} y\right) \quad (25)$$

式中： k 为变换系数。调整 k 的值可以改变 x 和 y 的变换关系，如图1所示。可以看出，随着 k 值的增大，变换曲线逐渐变缓， y 的取值也相应变小。故

可根据所求问题来调节 k 值。

在此变换下，概率密度函数变换的关系为^[35]

$$f_y(y) = \frac{\partial g^{-1}(y)}{\partial y} f_x(x) \quad (26)$$

将上述三角变换代入概率密度函数中，完成变量的替换，此时 $\Theta \in (-a, a)$ ， a 为正实数，经函数变换后，最大熵原理优化问题表述如下

$$\begin{aligned} \max H(y) &= -\int_{-a}^a f(y) \ln f(y) dy \\ \text{s.t.} \quad &\int_{-a}^a f(y) dy = 1 \\ &\int_{-a}^a y^i f(y) dy = M_i, \quad i = 1, 2, \dots, m \end{aligned} \quad (27)$$

这种改进方法既避免了截断误差的引入，又能在一定程度上提高计算精度和计算结果的稳定性，通常不引入过高阶矩信息即可得到较精确的结果。

2.2 基于密度核估计的最大熵原理

传统最大熵原理在 m 取值较大时，除了求解困难会增加外，所估计得到 $f(x)$ 也会更不易满足概率密度函数的基本条件

$$\begin{aligned} f(x) &> 0 \\ \int_{-\infty}^{+\infty} f(x) dx &= 1 \end{aligned} \quad (28)$$

如图2所示，在3阶矩约束下得到概率密度函数曲线，无论数据的偏度如何，都会出现某一侧曲线陡增的情况，而在这些地方 $f(x)$ 显然是不可积的，但截断处理又会额外增加截断误差。

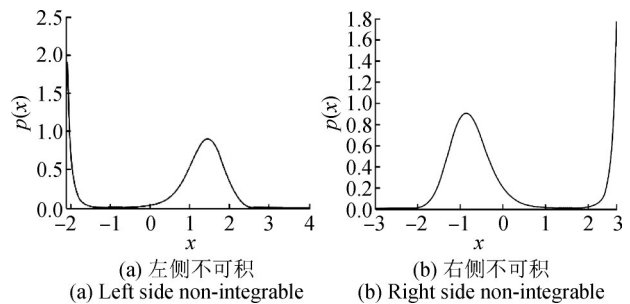


图2 3阶矩最大熵密度函数不可积示意图^[22]

Fig.2 Schematic diagram of non-integrability of maximum entropy density function of third-order moment^[22]

如果能使求 $f(x)$ 满足式(28)的条件，可以在很大程度上减少截断误差。为此可以引入密度函数核估计^[22]。与转换函数法变换积分区间不同，密度核函数的引入可以在原理上保证所求的概率

密度函数满足其基本性质。该方法对规则或不规则、单峰或多峰分布都能够有较好的估计，且在样本量足够丰富时，能以任意精度收敛到任何未知复杂分布^[36]。具体原理如下：

设 $K(u)$ 为 R 上的给定概率密度函数，其满足概率密度函数的基本条件^[37]

$$p_n(x) = \sum_{i=1}^n \alpha_i \left[\frac{1}{h_i} K\left(\frac{x-x_i}{h_i}\right) \right] \quad (29)$$

式中： $p_n(x)$ 为总体密度函数 $p(x)$ 的一个核估计； n 为核密度的数量； $h_i > 0$ 为窗宽； α_i 为各个核函数的权重^[38-39]，并有 $0 < \alpha_i < 1$ ， $\sum_{i=1}^n \alpha_i = 1$ ， $i=1, 2, \dots, n$ 。Parzen E^[36] 证明了总体密度函数 $p(x)$ 是一个满足式 (28) 条件的合理函数。

将核估计函数 $p_n(x)$ 代入到经典最大熵原理信息熵的定义中，则有^[22]

$$H(p_n) = - \int_{\Theta} \sum_{i=1}^n \alpha_i \left[\frac{1}{h_i} K\left(\frac{x-x_i}{h_i}\right) \right] \ln \left(\int_{\Theta} \sum_{i=1}^n \alpha_i \left[\frac{1}{h_i} K\left(\frac{x-x_i}{h_i}\right) \right] dx \right) dx \quad (30)$$

选择整数阶样本原点矩作为约束，则最大熵优化问题为

$$\begin{aligned} & \max H(p_n) \\ & s.t. \int_{\Theta} x^i p_n(x) dx = M_i, \quad i = 1, 2, \dots, m \end{aligned} \quad (31)$$

以服从对数正态分布的概率密度函数为例说明实施该原理的一般过程^[37]，即

$$f_{LN}(x) = \frac{1}{\sqrt{2\pi} \sigma x} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \quad (32)$$

式中： $x > 0$ ； $-\infty < \mu < \infty$ ； $\sigma > 0$ 。选用正态窗函数^[22]

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} \quad (33)$$

则总体 $p_n(x)$ 的核估计为^[37]

$$p_n(x) = \sum_{i=1}^n \alpha_i \left[\frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} \right] \quad (34)$$

为保证 α_i 满足其取值要求，同时使待优化参数的取值空间为实数空间，进行之下变换^[37]

$$\alpha_i = \frac{\exp(\beta_i)}{\sum_{i=1}^n \exp(\beta_i)} \quad (35)$$

式中： $\beta_i \in R$ 。最终总体密度函数 $f_{ln}(x)$ 的核估计变换为^[37]

$$p_n(x) = \sum_{i=1}^n \frac{\exp(\beta_i)}{\sum_{i=1}^n \exp(\beta_i)} \left[\frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} \right] \quad (36)$$

选取适当的 n 后，可确定未知参数的个数，再通过优化算法求解即可确定 $p_n(x)$ 的具体表达式。理论上 n 值越大，结果估计的越好，但同时也会增加算法的复杂度。

通过引入密度核的方法，来解决传统最大熵方法求得的概率密度函数不严格可积的问题，在避免截断误差的同时，也能在一定程度上提高估计结果的精度。

除以上两种方法，还有一类面对具体分布的改进方法：利用共轭分布法与最大熵原理结合的共轭最大熵。Savchuk V. P 和 Martz H. F^[40] 就二项分布情形，讨论了共轭分布与最大熵原理相结合确定先验分布的方法；詹昊可等^[41] 在 Savchuk V. P 的方法不适合指数分布的情况下，给出了一种新的解决思路，在多元先验信息下，讨论了共轭最大熵先验下的贝叶斯估计。

3 结论

本文根据目前已有的研究情况，对以 Shannon 熵为基础的最大熵原理进行了介绍，阐述了经典最大熵原理的基本内容，讨论不同约束条件下最大熵优化问题的一般形式，分析了针对经典最大熵不足的改进方法，对最大熵原理的进一步研究应用起到了促进作用。最后结合实践对最大熵原理的研究与应用现状提出以下见解：

1) 在约束的选择上，目前的应用依旧以整数阶样本原点矩为主。分数阶矩约束虽已被证明有效，但应用场景比较单一，需其在更多样问题上的应用研究，以进一步检验或改善其性能。秩约束虽然在理论上能解决样本矩在样本量过小时无法提供足够样本信息的问题，但提出该方法的学者也认为秩约束尚缺乏实际验证，因此应多开展相应的实用性研究。此外，在样本量较少时，利用统计学方法对样本数据进行扩充后再选取矩约束是否可取得理想的结果也缺乏相关研究。

2) 在估计结果 $f(x)$ 的评价上，目前主要有两种评价方法：①考察估计的 $f(x)$ 与理论曲线接近程度；②直接考察 $f(x)$ 统计量（如均值、方差）与理论值的接近程度。方法①适用于样本量较大

且理论曲线平而宽的对象,并能够得到均值、方差等信息。对于方法②,由于理论值往往是未知的,因此还需要利用残差平方和从侧面考察 $f(x)$ 。但对于不同的问题,残差平方和的大小是不同的,因此需要进行具体研究,以指导使用者通过残差平方和的值来判断求得的 $f(x)$ 是否已达到最优。

3) 最大熵原理得到的结果与优化问题的求解关系甚大。目前传统优化算法虽然能够求解最大熵优化问题,但存在初始值选择较繁琐,结果不稳定的问题,并且容易出现局部极小值问题。因此需要开展智能算法求解最大熵优化问题的研究。智能算法在很大程度上突破了传统优化算法的局限性,但在求解密度核最大熵优化问题时,由于其参数较多,且不宜转化为无约束优化问题,智能算法尚未得到实际应用,因此未来需要开展相关研究,以便更有效地求解密度核最大熵优化问题。

4) 在计量领域,学者们已经展开了最大熵原理在测量不确定度评定中的应用研究。但随着研究的深入和评定对象复杂度的增加,经典最大熵原理的局限性会对评定结果产生影响,因此需要展开最大熵原理改进方法在测量不确定度评定领域的应用研究,以扩大适用范围、提升评定结果的可靠性。

参考文献

- [1] SHANNON C E. A mathematical theory of communication [J]. John wiley and sons, Ltd, 1948, 27 (4) : 623-656.
- [2] JAYNES E T. Information theory and statistical mechanics [J]. Physical review, 1957, 106(4) : 620-640.
- [3] JAYNES E T. Information theory and statistical mechanics II [J]. Physical review, 1957, 108(2) : 171.
- [4] TSALLIS C. Possible generalization of Boltzmann-Gibbs statistics [J]. Journal of statistical physics, 1988, 52 (1) : 479-487.
- [5] PLASTINO A R, PLASTINO A. Non-extensive statistical mechanics and generalized Fokker-Planck equation [J]. Physica A : statistical mechanics and its applications, 1995, 222(1-4) : 347-354.
- [6] PLASTINO F, TSALLIS C. Nonadditive entropy reconciles the area law in quantum systems with classical thermodynamics [J]. Physical review E (statistical, nonlinear, and soft matter physics), 2008, 78(2) : 21102.
- [7] SHANNON C E. A mathematical theory of communication [J]. John wiley & sons, Ltd, 1948, 27 (4) : 379-423.
- [8] 曾晓华,袁持平. 基于最大熵原理的金融系统风险测度及实证研究 [J]. 东莞理工学院学报, 2019, 26(3) : 94-98, 124.
- ZENG X H, YUAN C P. Measurement and empirical study of financial system risk based on maximum entropy principle [J]. Journal of Dongguan institute of technology, 2019, 26(3) : 94-98, 124. (in Chinese)
- [9] GRANZIOL D, RU R, ZOHREN S, et al. A maximum entropy approach to massive graph spectra [J]. 2019.
- [10] NEUPER M, EHRET U. Quantitative precipitation estimation with weather radar using a data-and information-based approach [J]. Hydrology and earth system sciences discussions, 2019: 1-36.
- [11] 刘恩斌,周国模,葛宏立. 基于最大熵原理的浙江毛竹胸径分布及测量不确定度评定 [J]. 生态学报, 2009, 29(1) : 86-91.
- LIU E B, ZHOU G M, GE H L. Examination of moso bamboo's diameter probability distribution and evaluation of measurement uncertainty with the maximum entropy theory [J]. Acta ecologica sinica, 2009, 29(1) : 86-91. (in Chinese)
- [12] CHEN B Y, KOU Y, ZHAO D, et al. Maximum entropy distribution function and uncertainty evaluation criteria [J]. China ocean engineering, 2021, 35 (2) : 238-249.
- [13] 孙永厚,周洪彪,黄美发. 基于最大熵的测量不确定度的贝叶斯评估方法 [J]. 统计与决策, 2008 (12) : 143-145.
- SUN Y H, ZHOU H B, HUANG M F. Bayesian evaluation method of measurement uncertainty based on maximum entropy [J]. Statistics and decision making, 2008 (12) : 143-145. (in Chinese)
- [14] 姜瑞,陈晓怀,王汉斌,等. 基于贝叶斯信息融合的测量不确定度评定与实时更新 [J]. 计量学报, 2017, 38 (1) : 123-126.
- JIANG R, CHEN X H, WANG H B, et al. Evaluation and real time updating of measurement uncertainty based on bayesian information fusion [J]. Journal of measurement, 2017, 38(1) : 123-126. (in Chinese)
- [15] JAYNES E T. Papers on probability, statistics and statistical physics [M]. Dordrecht: Springer netherlands, 1989.
- [16] 吴贤东,汪加梅,李婉晴. Tsallis 最大熵原理及其逆问题 [J]. 重庆工商大学学报(自然科学版), 2019, 36 (2) : 31-36.
- WU X D, WANG J M, LI W Q. Tsallis maximum en-

- ropy principle and its inverse problem [J]. Journal of Chongqing technology and business university (natural science edition), 2019, 36(2): 31-36. (in Chinese)
- [17] 吴贤东. Tsallis最大熵原理的若干问题[D].安徽:安徽工业大学,2019.
WU X D. Some problems of Tsallis maximum entropy principle[D]. Anhui: Anhui university of technology, 2019.(in Chinese)
- [18] GZYL H, TAGLIANI A. Hausdorff moment problem and fractional moments[J]. Applied mathematics and computation, 2010, 216(11):3319-3328.
- [19] ZHANG X, PANDEY M D. Structural reliability analysis based on the concepts of entropy, fractional moment and dimensional reduction method [J]. Structural safety, 2013, 43:28-40.
- [20] 吕文. 最大熵原理与贝叶斯方法在测量数据处理中的应用[D].成都:电子科技大学,2006.
LV W. Application of maximum entropy principle and bayesian method in measurement data processing[D]. Chengdu: University of electronic science and technology of china, 2006.(in Chinese)
- [21] 余秀美,童玲,胡学海. 基于MATLAB最大熵分布的最优[C]//中国仪器仪表学会.2004全国测控、计量与仪器仪表学术年会论文集(上册).中国仪器仪表学会, 2004:6.
YU X M, TONG L, HU X H. Evaluation of uncertainty in measurement based on the maximum entropy method[C]// Chinese academy of instrumentation. 2004 national measurement and control, metrology and instrumentation symposium proceedings (Volume 1.Chinese academy of instrumentation, 2004:6. (in Chinese)
- [22] 刘钰,韩峰,王玉恒,等. 一种基于密度核估计的最大熵方法[J]. 工程数学学报, 2011, 28(3): 285-292.
LIU Y, HAN F, WANG Y H, et. al. A new maximum entropy method based on kernel density estimate [J]. Journal of engineering mathematics, 2011, 28(3): 285-292. (in Chinese)
- [23] 姜瑞. 现代不确定度评定方法及应用[D].安徽:合肥工业大学, 2017.
JIANG R. Modern uncertainty evaluation method and its application[D]. Anhui: Hefei university of technology, 2017.(in Chinese)
- [24] 彭颖,朱南海. 基于遗传算法的数据最大熵概率分布计算[J]. 南昌大学学报(工科版), 2020, 42(1): 40-45.
PENG Y, ZHU N H. Calculation of maximum entropy probability distribution of data based on genetic algorithm [J]. Journal of nanchang university (engineering edition), 2020, 42(1): 40-45. (in Chinese)
- [25] SMALE S. Mathematical problems for the next century[J]. The mathematical intelligencer, 1998, 20(2): 7-15.
- [26] GUO D, NIE Z, YAN L. The application of noise-tolerant zd design formula to robots' kinematic control via time-varying nonlinear equations solving [J]. IEEE transactions on systems, man, and cybernetics: systems, 2017: 2188-2197.
- [27] 高卫峰,罗宇婷,原杨飞. 求解非线性方程组的智能优化算法综述[J]. 控制与决策,2021,36(4):769-778.
GAO W F, LUO Y T, YUAN Y F. Overview of intelligent optimization algorithms for solving nonlinear equation systems[J]. Control and decision making, 2021, 36(4): 769-778. (in Chinese)
- [28] BROYDEN C G. A class of methods for solving nonlinear simultaneous equations [J]. Mathematics of computation, 1965, 19(92): 577-593.
- [29] RAMOS H, MONTEIRO M T T. A new approach based on the Newton's method to solve systems of nonlinear equations[J]. Journal of computational and applied mathematics, 2017, 318: 3-13.
- [30] GAAFAR L K, MASOUD S A. Genetic algorithms and simulated annealing for scheduling in agile manufacturing [J]. International journal of production research, 2005, 43(14): 3069-3085.
- [31] CORANA A, MARCHESI M, MARTINI C, et al. Minimizing multimodal functions of continuous variables with the "simulated annealing" algorithm [J]. ACM transactions on mathematical software, 1987, 13(3): 262-280.
- [32] MHETRE P S. Genetic algorithm for linear and nonlinear equation[J]. International journal of advanced engineering technology, 2012(2): 114-118.
- [33] KENNEDY J, EBERHART R C. Particle swarm optimization [C]//Proceedings of the 1995 IEEE international conference on neural networks. Perth: IEEE service center, 1995: 1942-1948.
- [34] 罗宇婷. 求解非线性方程组的智能优化算法研究[D].陕西:西安电子科技大学,2020.
LUO Y T. Research on intelligent optimization algorithms to solve nonlinear equation systems[D]. Shaanxi: Xidian university, 2020.(in Chinese)
- [35] 李昊燃. 结构可靠性分析的改进最大熵方法[D].辽宁:大连理工大学, 2015.
LI H R. The improved maximum entropy method of structure reliability analysis[D]. Liaoning:Dalian University of Technology, 2015.(in Chinese)

- [36] PARZEN E. On estimation of a probability density function and model [J]. *Annals of mathematical statistics*, 1962, 33(7): 316-327.
- [37] 卢钰坤. 密度核估计最大熵方法在控制图设计中的应用[D]. 河北: 燕山大学, 2019.
- LU Y K. Maximum entropy method based on density kernel estimation in control chart design[D]. Hebei: Yanshan university, 2019. (in Chinese)
- [38] 边肇祺, 张学工. 模式识别[M]. 北京: 清华大学出版社, 2000.
- BIAN Z Q, ZHANG X G. *Pattern recognition*[M]. Beijing: Tsinghua university press, 2000. (in Chinese)
- [39] 郭照庄, 宋向东, 张良勇, 等. 变窗宽密度核估计的构造及均方相合性[J]. *佳木斯大学学报(自然科学版)*, 2006, 24(3): 423-425.
- GUO Z Z, SONG X D, ZHANG L Y, et al. Structure and consistent of the variable bandwidth for kernel density estimation [J]. *Journal of Jiamusi university (natural science edition)*, 2006, 24(3): 423-425. (in Chinese)
- [40] SAVCHUK V P, MARTZ H F. Bayes reliability estimation using multiple sources of prior information: binomial sampling [J]. *IEEE transactions on reliability*, 2002, 43(1): 138-144.
- [41] 詹昊可, 姜礼平. 共轭最大熵先验下的贝叶斯估计 [J]. *运筹与管理*, 2002(6): 27-31.
- ZHAN H K, JIANG L P. Bayesian estimation based on maximum-entropy conjugate priors [J]. *Operations research and management*, 2002(6): 27-31. (in Chinese)

(本文编辑: 刘圣晨)