
欧拉公式与二进制在数据科学中获新应用

作者：writer 来源：爱科学

本文原地址：<https://www.iikx.com/news/progress/13117.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

欧拉公式与二进制在数据科学中获新应用。欧拉公式是很多数学方向的根基，而二进制是计算机理解人类世界的语言。近日，美国北卡罗莱纳大学教堂山分校教授张凯、天普大学教授赵志根以及科罗拉多州立大学教授周文利用对连续随机变量的二进制逼近，在数据科学中对这两个经典理论做了结合和创新，特别是为均匀性检测以及随机向量的独立性检测提供了一个全新的非参数理论框架，并将现有理论方法统一在新的框架之下。相关成果预印本日前发表于arxiv.org。

欧拉公式 $e^{ix} = \cos x + i \sin x$

x 是数学发展史上最重要的里程碑之一。当取 x 为 π 时，该公式可以简化为 $e^{i\pi} + 1 = 0$ ，也就是著名的欧拉恒等式。这个恒等式将数学中最重要的5个常数0、1、圆周率 π 、自然对数的底 e 和虚数单位 i 用简单而深刻的方式联系起来。由于该公式是很多数学、物理和工程理论的基础，它在《费曼物理学讲义》一书中被评为最卓越的数学公式。

此次，研究人员在研究成果中首次从概率论角度发展出了基于两点分布的欧拉公式（BEE）。假设 A 是任意一个仅取值于-1和1的随机变量，则对任意的实数 x 有 $e^{iAx} = \cos x + iA \sin x$ 。该等式是经典的欧拉公式在概率意义下的直观的推广。

该论文进一步将上述公式应用于张凯于2019年发表的二进制展开检验（BET）的非参数统计框架，指出对于任意一个连续随机向量，首先将它的耦合分布的每一个维度进行二进制展开，再结合BEE可以得出初始分布的特征函数的二进制逼近（BEAUTY）。这个基于欧拉公式的二进制逼近为当前广泛应用于均匀性检测和独立性检测的众多非参数方法提供了统一的理论基础，并且为进一步改进和提高指明了方向。

在此基础上，该论文进一步构造了基于BEAUTY的BEAST检验方法。在统计方面，研究人员利用大量数据实证检验了该方法，证明BEAST与现有方法相比不仅大幅度地提高了检测的一致性和稳健性统计功效，而且提高了对各种复杂的关联结构的解释性。在计算方面，研究人员通过将观测值转化成二进制的途径，使得该算法可以直接对存储于机器中的数据进行位操作，从而大幅度提高了计算效率。在应用方面，由于该检验属于不受分布限制的非参数统计方法，所以可以被广泛地应用于生物、工程、医学、经济等自然和社会科学领域，为数据分析提供有力的新工具。

该文章在网上公布的同时，相关的统计应用软件R语言包BET已经发布在CRAN上可供免费使用。（来源：中国科学报唐凤）

相关论文信息：<https://arxiv.org/abs/2103.00674>

<https://doi.org/10.1080/01621459.2018.1537921>

版权声明：凡本网注明来源：中国科学报、科学网、科学新闻杂志的所有作品，网站转载，请在正文上方注明来源和作者，且不得对内容作实质性改动；微信公众号、头条号等新媒体平台，转载请联系授权。邮箱：shouquan@stimes.cn。
作者：张凯等 来源：JAM STAT ASSOC

更多科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](http://www.iikx.com)转发