

Monophyletic Representation of Random Vector

Shuo Zhuang, Jihui Ding, Baoxin Tang

School of Civil Science and Engineering, Yangzhou University, Yangzhou Jiangsu
Email: 337455377@qq.com

Received: Mar. 9th, 2017; accepted: Mar. 27th, 2017; published: Mar. 30th, 2017

Abstract

Based on orthogonal function theory and Karhunen-Loeve decomposition method, monophyletic random vector is constructed to solve the reduction of random variables in stochastic structural analysis. Based on the probability theory, the approximate independence of the monophyletic random vector and the construction method of the monophyletic random vector are studied. The results show that: 1) the monophyletic random vector is independent, the accuracy is sufficient to meet the engineering requirements; 2) the monophyletic random vector is simple and repeatable. Furthermore, it provides a new method for the stochastic analysis of structures.

Keywords

Structure Stochastic Analysis, Monophyletic, Random Vector Reduction, Approximate Independence

随机向量的单源表达方法

庄 硕, 丁继辉, 汤保新

扬州大学建筑科学与工程学院, 江苏 扬州
Email: 337455377@qq.com

收稿日期: 2017年3月9日; 录用日期: 2017年3月27日; 发布日期: 2017年3月30日

摘 要

基于正交函数理论和Karhunen-Loève分解法, 构造了一种单源随机向量, 从而实现了结构随机分析中随机变量的缩减。基于概率论方面的相关理论, 研究了单源随机向量的近似独立性和单源随机向量的构造方法。结果表明: 1) 单源随机向量近似独立, 精度满足工程要求; 2) 单源随机向量构造简单。这为

结构随机分析提供了一个新的手段。

关键词

结构随机分析, 单源, 随机向量缩减, 近似独立

Copyright © 2017 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

目前为减少实际工程基本随机变量个数, 一种可靠的方法是通过数学上的正交变换方法缩减随机变量个数。下面以随机过程的描述为例。Karhunen-Loève 分解法[1]将随机过程分解为各分量不相关的正交函数叠加, 大约需要 200~600 个基本随机变量。李杰等[2] [3]进一步研究基于标准正交基的随机过程展开法, 通过随机向量的相关分解对展开系数实施正交化, 基本随机变量一般为 10~20 个左右。汤保新[4] [5]对李杰等采用的不相关的基本随机变量再次采用单源随机向量(正交函数)表达, 随机过程的描述仅需 1 个基本随机变量。

本文给出了“单源随机向量”概念的更进一步解释, 并探讨了单源随机向量的近似独立性和构造方法。

2. 单源随机向量的概念

“单源”是遗传学上的概念, 它在模糊数学和力学领域已有借用的范例[6] [7]。

所谓单源随机向量[8] (monophyletic random vector), 是指以同一个随机变量为自变量的一组随机函数序列。如 $\xi = [\xi_1, \xi_2, \dots, \xi_n]^T$, 其中, $\xi_k = \xi_k(\theta), k = 1, 2, \dots, n$, θ 为唯一的随机源变量, 则称 ξ 为单源随机向量。可以构造一个分量服从反正弦分布的单源随机向量,

$$\xi_k = \sin(k\theta + \alpha_k), \quad k = 1, 2, \dots, n \quad (1)$$

式中, θ 为 $[-\pi, \pi]$ 上均匀分布的唯一随机变量, α_k 为任意确定的常量。容易推导, ξ_k 的概率密度函数形式同为,

$$f_{\xi_k}(x_k) = \frac{1}{\pi\sqrt{1-x_k^2}}, \quad -1 < x_k < 1 \quad (2)$$

概率分布函数为

$$F_{\xi_k}(x_k) = \frac{1}{2} + \frac{1}{\pi} \arcsin x_k = 1 - \frac{1}{\pi} \arccos x_k, \quad -1 \leq x_k \leq 1 \quad (3)$$

由此可见, ξ 的各分量服从同一分布, 均值 $E(\xi_k) = 0$, 均方差 $\sigma(\eta_k) = \frac{\sqrt{2}}{2}$ 。任意两个分量之间的相关系数 $\rho(\xi_p, \xi_q) = 0, p \neq q$, 即各分量之间不相关, 但并不相互独立。单源随机向量的性质将决定其应用, 下一节介绍其性质的研究。

3. 单源随机向量的近似独立性

因为随机源相同, 单源随机向量的各分量之间通过唯一随机源相联系, 所以各分量之间是不独立的,

存在一定的相依关系。这种关系, 经典概率论一般采用独立性和线性相关系数描述, 但这种描述存在一定局限性[9]。

这里仅研究三角函数形式的单源随机向量 $\xi_k = \sin k\theta$ 或 $\xi_k = \cos k\theta$ 的近似独立性, 这里 $k \in N$, θ 服从 $[-\pi, \pi]$ 上的均匀分布(以下记为 $\theta \sim U[-\pi, \pi]$)。

1) 先考虑用联合概率分布是否等于各边缘分布的乘积来判断。

公式推导发现联合概率分布不等于各边缘分布的乘积。在严格的概率意义上, $[\xi_p, \xi_q]$ 是不相互独立的。但从数值计算结果看, $F_{\xi_p \xi_q}(x, y)$ 与 $F_{\xi_p}(x)F_{\xi_q}(y)$ 很接近, 见图 1。误差大小跟 p, q 的取值有关, 当 $p=4, q=7$ 时, 最大误差 0.0063; 当 $p=17, q=23$ 时, 最大误差 0.0067。由于 $F_{\xi_p \xi_q}(x, y) \approx F_{\xi_p}(x)F_{\xi_q}(y)$, 所以, 在不太严格的概率意义上说, $[\xi_p, \xi_q]$ 是接近相互独立的, 不妨称之为近似独立。

2) 其次考虑按统计学方法通过样本检验来判断。

这里采用二维列联表的独立性 χ^2 检验[10]。将 $[\xi_p, \xi_q]$ 的值域离散为 $r \times c = n$ 个区格, 抽样统计每个区格的频数 O_{ij} , 与每个区格的理论频数 E_{ij} 进行比较, 计算统计量

$$\chi_n^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (4)$$

当该统计量服从自由度为 $d = (r-1) \times (c-1)$ 的 χ^2 分布时, 说明抽样数据相互独立。取显著水平为 α , 检验结果见表 1。由表可见, $\chi_n^2 < \chi_n^2(\alpha, d)$, 所以无法拒绝抽样数据的独立性。

3) 最后按高阶相关系数度量单源随机向量的近似独立性。

基于多边矩阵理论[11], 罗纯、王晓迪、张应山[12]给出了两组随机向量相互独立和相互观测独立的定义, 推出了框架的正交性与随机变量的独立性等价定理。崔瑞峰、牛新军[13]以高阶相关系数描述随机变量之间的非线性关系, 定义了两个随机变量 X, Y 的 n 阶协方差矩阵,

$Cov_n(X, Y) = [Cov(X^s, Y^t)]_{s, t=1, 2, \dots, n}$, 当 $Cov_n(X, Y) = O_{n \times n}$ 时, 称 X, Y 为 n 阶不相关。研究证明, 对两个连续型随机变量, 任意高阶不相关与相互独立等价。

由于协方差的计算会受到量纲的影响, 这里定义两个随机变量 X, Y 的 n 阶相关系数矩阵(无量纲), $\rho_n(X, Y) = [\rho(X^s, Y^t)]_{s, t=1, 2, \dots, n}$, 那么, 可将高阶相关系数矩阵接近不相关矩阵 $O_{n \times n}$ 的程度作为两个随机变量逼近相互独立的判据。本文研究了两个随机变量的情形。由定义

$$\rho(s, t) = \rho(X^s, Y^t) = \frac{E(X^s Y^t) - E(X^s)E(Y^t)}{\sigma(X^s)\sigma(Y^t)} \quad (5)$$

不难推导各阶相关系数。

这里以四阶相关系数($s+t=4$)为例。此时有 $s=1, t=3$, 或 $s=2, t=2$, 或 $s=3, t=1$ 三种情形。利用对称性, 仅需讨论前 2 种情形, 结果如下:

$$\rho(1, 3) = \begin{cases} 0 & p \neq 3q \\ -1/\sqrt{10} & p = 3q \end{cases}, \quad \rho(2, 2) = 0, \quad \rho(3, 1) = \begin{cases} 0 & q \neq 3p \\ -1/\sqrt{10} & q = 3p \end{cases} \quad (6)$$

其它各阶相关系数最大值见表 2。

研究发现, ρ 均与 p, q 的取值有关。当 p, q 经过约简后, 如果 $p^2 q^2$ 值较大, 则各阶相关系数较小。可以通过适当的 p, q 选值, 达到一定的接近独立的程度。

4. 任意分布单源随机向量的构造

由均匀分布单源随机向量

$$u_k = \frac{1}{2} + \frac{1}{\pi} \arcsin \sin(k\theta + \alpha_k) \tag{7}$$

可构造任意类型的概率分布 F ,

$$v_k = F^{-1} \left(\frac{1}{2} + \frac{1}{\pi} \arcsin \sin(k\theta + \alpha_k) \right) \tag{8}$$

由式(7)所构造的均匀分布函数近似独立, 如图 2 所示。

用正态分布作为媒介, 通过变换, 进行单源构造后, 可将实际随机变量的数量减少到 1 个。

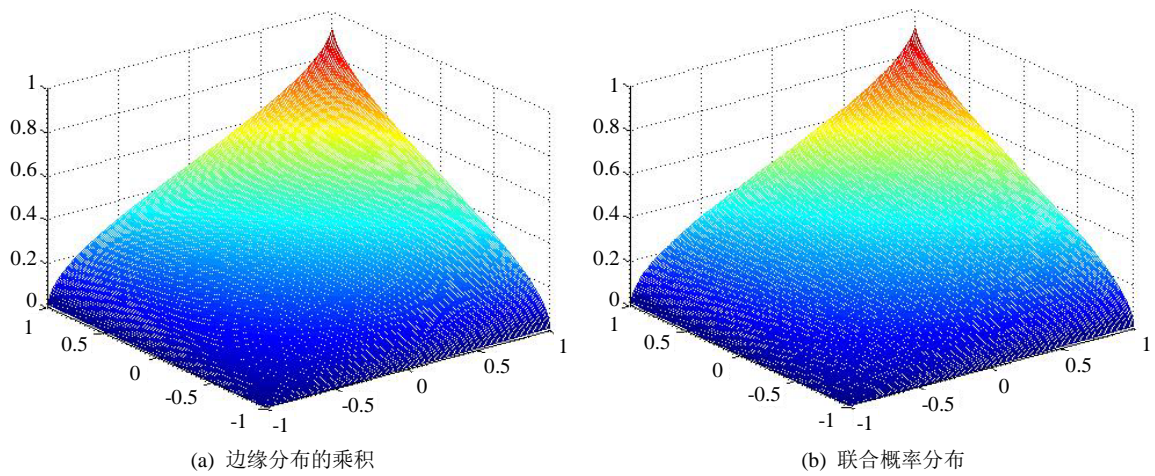


Figure 1. Two dimensional joint probability distribution function
图 1. 二维联合概率分布函数

Table 1. The independence test of two dimensional contingency table
表 1. 二维列联表的独立性检验

$\alpha = 0.05$ ($\alpha = 0.001$)	$\xi_p = \sin p\theta, \xi_q = \sin q\theta \ (p, q \in N, p \neq q)$			
	$p = 4, q = 7$		$p = 17, q = 23$	
区格数 $r \times c = n$	$r \times c = 5 \times 5$	$r \times c = 10 \times 10$	$r \times c = 5 \times 5$	$r \times c = 10 \times 10$
统计量 χ_n^2	0.0190	2.1012	0.0013	0.6438
临界值 $\chi_n^2(\alpha, d)$	7.9616 (3.9416)	61.2615 (47.2770)	7.9616 (3.9416)	61.2615 (47.2770)
独立性检验	独立	独立	独立	独立

Table 2. Higher order correlation coefficients of two random variables
表 2. 两个随机变量的高阶相关系数

阶数	2	3	4	5	6	7	8
ρ_{\max}	0	0	$\frac{1}{\sqrt{10}}$	0	$\frac{5}{\sqrt{126}}$	0	$\frac{5}{\sqrt{70}}$

注: ρ_{\max} 为相关系数绝对值的最大值。

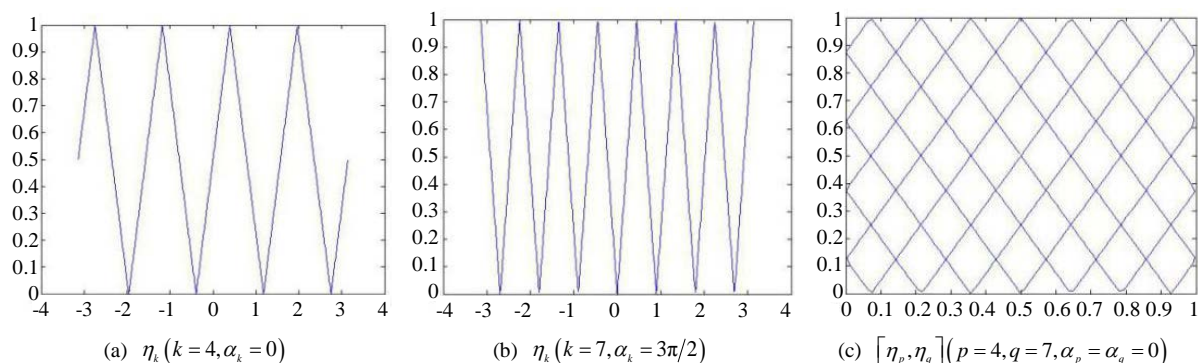


Figure 2. The two-dimensional uniform rectilinear grid

图 2. 二维均匀分布直线网格

5. 结论

综上所述, 单源随机向量 $[\xi_p, \xi_q]$ 在概率论上本质是不独立的, 但相当接近于独立。高阶相关系数足够小且统计检验又无法拒绝抽样数据的独立性, 可以说, 单源随机向量所表现的统计意义上的观测独立性达到了“以假乱真”的效果, 甚或掩盖了概率意义上的不相互独立的本质。对数值模拟来说, 如果单源随机向量近似独立, 当抽样次数不多时亦可满足样本点的独立性要求, 这为有效减少抽样次数提供了方法论上的依据, 也为结构随机分析提供了新的手段。

基金项目

国家自然科学基金资助项目(51541805)。

参考文献 (References)

- [1] Loève, M. (1977) Probability Theory. Springer-Verlag, Berlin.
- [2] 李杰, 刘章军. 基于标准正交基的随机过程展开法[J]. 同济大学学报: 自然科学版, 2006, 34(10): 1279-1283.
- [3] 刘章军, 李杰. 地震动随机过程的正交展开[J]. 同济大学学报: 自然科学版, 2008, 36(9): 1153-1159.
- [4] Tang, B.X., Cheng, K.H. and Li, Q. (2011) Simulating Stochastic Process with a Monophyletic Random Vector. *Sustainable Development of Urban Environment and Building Material*, Tran Tech Publications, Zurich-Durnten, 1698-1703.
- [5] 汤保新. 随机结构的单源随机向量表达[J]. 扬州大学学报, 2011, 14(4): 64-68.
- [6] 王柏生. 单源模糊数及运算[J]. 模糊系统与数学, 1998, 12(2): 49-53.
- [7] 刘长虹, 陈虬. 单源模糊数的模糊随机有限元方程的解法[J]. 应用数学与力学, 2000, 21(11): 1147-1150.
- [8] 汤保新. 单源随机向量及其在随机分析中的应用[J]. 工程力学, 2012, 29(12): 51-55.
- [9] 严忠权. 随机变量相依关系的度量[J]. 黔南民族师范学院学报, 2008(6): 34-38.
- [10] 盛骤, 谢式千, 潘承毅. 概率论与数理统计[M]. 第3版. 北京: 高等教育出版社, 2001.
- [11] 张应山. 多边矩阵理论[M]. 北京: 中国统计出版社, 1993.
- [12] 罗纯, 王晓迪, 张应山. 正交性与独立性——处理复杂系统的新思维系列之三[J]. 上海应用技术学院学报: 自然科学版, 2010, 10(4): 271-277.
- [13] 崔瑞峰, 牛新军. 随机变量相互独立与不相关之间关系的推广[J]. 西安电子科技大学学报, 1999, 26(2): 262-264.

期刊投稿者将享受如下服务：

1. 投稿前咨询服务 (QQ、微信、邮箱皆可)
2. 为您匹配最合适的期刊
3. 24 小时以内解答您的所有疑问
4. 友好的在线投稿界面
5. 专业的同行评审
6. 知网检索
7. 全网络覆盖式推广您的研究

投稿请点击：<http://www.hanspub.org/Submission.aspx>

期刊邮箱：aam@hanspub.org