基于分位数与耦合诊断平稳随机过程相依性

献给王梓坤教授 90 华诞

张世斌¹, 张新生^{2*}

1. 上海师范大学数学系, 上海 200234;

2. 复旦大学统计学系, 上海 200433

E-mail: zhangshibin1@gmail.com, xszhang@fudan.edu.cn

收稿日期: 2018-03-01; 接受日期: 2019-01-18; 网络出版日期: 2019-02-25; * 通信作者
国家自然科学基金 (批准号: 11571080 和 11671416) 资助项目

摘要 基于分位数和耦合的概念, 本文提出平稳随机过程分位数 - 耦合交叉协方差函数的概念. 该函数可以描述平稳过程的时间可逆性、状态相依性以及二阶矩不存在时的长程相依性等自协方差函数无法刻画的过程特征. 本文讨论了平稳随机过程分位数 - 耦合交叉协方差函数的一些基本性质; 基于离散抽样, 给出了分位数 - 耦合交叉协方差函数的估计量, 并证明了其相合性; 利用文中估计方法, 对具有相同一维平稳边际分布和自协方差函数的不同类型随机过程 (Gamma-OU (Ornstein-Uhlenbeck) 过程和 CIR (Cox-Ingersoll-Ross) 模型) 的分位数 - 耦合交叉协方差函数进行比较, 得到了其自相依性质的差别. 最后, 将估计方法应用于船体振动序列自相依性质分析, 为船体振动模型选择提供借鉴.

关键词 耦合 时间可逆性 长程相依性 状态相依性 平稳过程 OU 型过程 扩散

MSC (2010) 主题分类 60G10, 62M10

1 引言

设 $\{\Omega, \mathcal{F}, P\}$ 为概率空间, $\{\mathcal{F}_t, t \in \mathbb{R}\}$ 为定义在时间集 $\mathbb{R}$ 上的右连续 σ -代数流. 又设 $\{X_t, t \in \mathbb{R}\}$ 为关于 $\{\mathcal{F}_t, t \in \mathbb{R}\}$ 可测的严平稳随机过程. 长期以来, 基于二阶矩定义的随机过程自协方差函数 $C(u) = \text{Cov}(X_t, X_{t+u})$ 是描述过程自相依关系的一个重要工具. 例如, 二阶矩存在时过程长程相依性的定义^[1]、众多平稳过程中心极限定理相依性的条件^[2]等都通过协方差函数给出. 然而, 用自协方差函数 $C(u)$ 描述过程的相依性至少存在如下三个方面的问题.

第一, 自协方差函数不能刻画历史对现在的超“位置 - 尺度”影响. 平稳随机过程的自相依信息除了蕴含在其位置与尺度参数中之外, 还可能反映在其偏度和峰度等信息中, 而随机过程在这些方面的相依信息却无法用自协方差函数 $C(u)$ 加以刻画.

英文引用格式: Zhang S, Zhang X. Dependence diagnosis for stationary stochastic processes based on both quantiles and copulas (in Chinese). Sci Sin Math, 2019, 49: 591–606, doi: 10.1360/N012018-00043

第二, 自协方差函数关于随机过程不具有非线性变换稳健性. 自协方差函数对随机过程的非线性变换是不稳健的. 如果多个观测者观察同一过程的同一轨道, 但采用了不同的计量尺度, 例如, 如果他们分别考察 X_t 、 $\exp(X_t)$ 或 X_t^3 , 则他们将得到截然不同的描述同一观察过程相关性的自协方差函数.

第三, 自协方差函数的存在性依赖于二阶矩的存在性. 并不是所有随机过程的自协方差函数都存在, 对于由 α - 稳定 Lévy 过程 ($0 < \alpha < 2$) (参见文献 [3]) 驱动的扩散型过程 (即 X_t 为扩散型随机微分方程 $dX_t = \mu(X_{t-})dt + \sigma(X_{t-})dL_t$ 的解, 其中 L_t 为 α - 稳定 Lévy 过程) 而言, 其自协方差函数就不存在, 也就无法用自协方差函数刻画其自相依性质, 因此, 有很多学者借助于 Hurst 指数 (如文献 [1]) 刻画平稳随机过程自相依程度.

平稳随机过程在两时间点 t 和 $t-u$ 处的时间相依特征完全包含在 (X_t, X_{t-u}) 的联合分布中. 事实上, (X_t, X_{t-u}) 的联合分布共包含两个方面的信息: 一是一维边际分布 $F(x)$ 的信息, 二是 $(F(X_t), F(X_{t-u}))$ 的联合分布信息, 其中 $F(\cdot)$ 为随机过程的一维边际分布函数. 可以看出, 平稳随机过程在两个时间点的时间相依特征完全包含在 $F(X_t)$ 和 $F(X_{t-u})$ 的相依信息中. 同时, 为了能够刻画历史对现在的超“位置 - 尺度”影响, 本文借用分位数分析的思想, 对任意的 $(\tau_1, \tau_2) \in (0, 1)^2$, 定义刻画随机过程自相依特征的量

$$\gamma^{(\tau_1, \tau_2)}(u) := \text{Cov}(\mathbb{I}_{(0, \tau_1]}(U_t), \mathbb{I}_{(0, \tau_2]}(U_{t-u})) = \text{Cov}(\mathbb{I}_{(-\infty, q_{\tau_1}]}(X_t), \mathbb{I}_{(-\infty, q_{\tau_2}]}(X_{t-u})), \quad (1.1)$$

其中 $U_t := F(X_t)$ 为由平稳分布 F 诱导的随机过程, q_τ 为一维边际分布 F 的 τ - 分位点. (1.1) 实际上是两个取值为 0-1 的随机过程 $\mathbb{I}_{(0, \tau_1]}(U_t)$ 与 $\mathbb{I}_{(0, \tau_2]}(U_{t-u})$ (或 $\mathbb{I}_{(-\infty, q_{\tau_1}]}(X_t)$ 与 $\mathbb{I}_{(-\infty, q_{\tau_2}]}(X_{t-u})$) 的交叉协方差, 其定义一方面用到了耦合 (copula)^[4] 的思想; 另一方面用到了平稳过程一维边际分布的分位数, 因此, 称 (1.1) 为随机过程 $\{X_t\}$ 的 (τ_1, τ_2) - 分位数 - 耦合交叉协方差函数. 显然, (1.1) 定义的量不仅可以克服上面提及的自协方差函数在描述过程相依性方面存在的三个弊端, 而且通过变换 τ_1 与 τ_2 的值能够全面描述随机过程 $\{X_t\}$ 在任意两时刻处的自相依性质.

在时间序列的相依性刻画方面, 借助分位数的分析方法在近年来尤其得到关注. 例如, 利用基于分位数的谱, 文献 [5] 给出了两类检验序列相依性的方法; 为了处理重尾 (heavy-tailed) 噪声及数据获得和传输中的非线性扭曲问题, 文献 [6] 提出了分位数周期图 (quantile periodogram) 的概念. 最近, 文献 [7] 将分位数和耦合的思想融合, 提出了一个既拥有分位数分析的优点又能描述时间序列中传统谱分析方法难以刻画的相依性质的工具—耦合谱密度核 (copula spectral density kernel). 在时间序列谱分析领域, 该量非常具有吸引力. 相继, 文献 [8] 对其一类非参数估计量进行了渐近分析; 文献 [9] 基于 Whittle 型似然给出了其 Bayes 统计推断方法; 文献 [10] 将该概念推广到了局部平稳时间序列的情形. 上述研究都是将时间序列转化到频域中进行分析. 对时间序列而言, 在频域中研究序列的自相依性质是方便的, 相应谱密度 (如自协方差函数的 Fourier 变换) 的定义域为有界区间 $(-\pi, \pi]$, 实现了数据的压缩. 但是, 对于时间连续的随机过程而言, 频域分析的定义域仍为无界区域, 即 $\mathbb{R}$, 相对于在时域内对过程部分自相依性质的研究, 其优势并不十分明显.

基于分位数 - 耦合交叉协方差函数, 本文从时域角度探讨其在平稳随机过程时间可逆性、状态相依性和时间相依性强弱比较等过程相依特征描述与研究方面的应用. 基于离散抽样, 分位数 - 耦合交叉协方差函数的经验版本 (估计量) 十分易于获得, 并且关于真值具有相合性. 进一步, 通过模拟实验, 验证了所提出的估计量的小样本可靠性. 借助于分位数 - 耦合交叉协方差函数的估计量, 本文不仅对一维平稳边际分布和自协方差函数完全相同的两类随机过程 (Gamma-OU 过程与 CIR 模型) 进行研究并发现了其自相依性质的若干差别, 而且考察了船体振动过程的自相依性质并提供了对其建模的建议.

2 分位数 - 耦合交叉协方差函数

2.1 分位数 - 耦合交叉协方差函数的优点

显然, 平稳随机过程的自协方差函数仅仅描述其在任意两时刻处的线性自相依性. 为全面描述平稳随机过程在任意两时刻处的自相依性, 同时融合分位数和耦合两个分析角度, 本文提出由 (1.1) 定义的分位数 - 耦合交叉协方差函数 $\gamma^{(\tau_1, \tau_2)}(u)$. 不同概率对下分位数 - 耦合交叉协方差函数的集合

$$\{\gamma^{(\tau_1, \tau_2)}(u) \mid (\tau_1, \tau_2) \in (0, 1)^2, u \in \mathbb{R}\} \quad (2.1)$$

可以全面描述一平稳随机过程在任意两时刻处的自相依性质.

对任意的 $u \in \mathbb{R}$ 和 $(\tau_1, \tau_2) \in (0, 1)^2$, 分位数 - 耦合交叉协方差函数 $\gamma^{(\tau_1, \tau_2)}(u)$ 的值仅由 $\mathbb{I}_{(0, \tau_1]}(U_t) = \mathbb{I}_{(-\infty, q_{\tau_1}]}(X_t)$ 和 $\mathbb{I}_{(0, \tau_2]}(U_{t-u}) = \mathbb{I}_{(-\infty, q_{\tau_2}]}(X_{t-u})$ 定义. 与经典的自协方差函数不同, 对所有的平稳随机过程而言, $\gamma^{(\tau_1, \tau_2)}(u)$ 始终都是存在的. 并且, 对于固定的 $u \in \mathbb{R}$, $\{\gamma^{(\tau_1, \tau_2)}(u) \mid (\tau_1, \tau_2) \in (0, 1)^2\}$ 完全刻画了 (U_t, U_{t-u}) 的联合分布信息; 反之, (U_t, U_{t-u}) 的联合分布信息完全能够描述 $\{\gamma^{(\tau_1, \tau_2)}(u) \mid (\tau_1, \tau_2) \in (0, 1)^2\}$. 平稳随机过程任意两时刻处的自相依信息完全可以由 (U_t, U_{t-u}) 的联合分布所提供的信息来刻画. 现在, 分位数 - 耦合交叉自协方差函数将平稳随机过程任意两时刻处的自相依信息从二维平稳边缘分布中剥离出来, 研究随机过程任意两时刻处的自相依信息只需研究集合 (2.1) 即可.

2.2 平稳随机过程的自相依特征

因为 $E[\mathbb{I}_{(0, q_{\tau_1}]}(X_t)\mathbb{I}_{(0, q_{\tau_2}]}(X_{t-u})] = P(X_t \leq q_{\tau_1}, X_{t-u} \leq q_{\tau_2})$ 并且 $0 \leq P(X_t \leq q_{\tau_1}, X_{t-u} \leq q_{\tau_2}) \leq \min\{P(X_t \leq q_{\tau_1}), P(X_{t-u} \leq q_{\tau_2})\} = \min(\tau_1, \tau_2)$, 故

$$-\tau_1\tau_2 \leq \gamma^{(\tau_1, \tau_2)}(u) \leq \min(\tau_1, \tau_2) - \tau_1\tau_2$$

对所有的 $u \in \mathbb{R}$ 和 $(\tau_1, \tau_2) \in (0, 1)^2$ 都成立. 并且, 平稳随机过程的自相依性与分位数 - 耦合交叉协方差函数存在如下关系.

命题 2.1 (自相依性与分位数 - 耦合交叉协方差函数) 假设平稳随机过程 $\{X_t, t \in \mathbb{R}\}$ 的一维边缘分布在状态空间上具有严格单调递增的分布函数, 则

(1) 平稳随机过程 $\{X_t, t \in \mathbb{R}\}$ 是一关于时间两两独立的随机过程 (即对于任意的 $t_1, t_2 \in \mathbb{R}$, X_{t_1} 与 X_{t_2} 相互独立) 的充分必要条件是 $\gamma^{(\tau_1, \tau_2)}(u) = (\min(\tau_1, \tau_2) - \tau_1\tau_2)\mathbb{I}_{\{0\}}(u)$ 对所有的 $(\tau_1, \tau_2) \in (0, 1)^2$ 都成立;

(2) 对所有的 $u \in \mathbb{R}$, $(X_t, X_{t+u}) \stackrel{d}{=} (X_t, X_{t-u})$ 成立的充分必要条件是对任意的 $(\tau_1, \tau_2) \in (0, 1)^2$, $\gamma^{(\tau_1, \tau_2)}(u)$ 关于 u 为偶函数, 其中 “ $\stackrel{d}{=}$ ” 表示依概率分布相等, 当 $\tau_1 = \tau_2 = \tau \in (0, 1)$ 时, $\gamma^{(\tau, \tau)}(u)$ 关于 u 一定是偶函数.

证明 因平稳随机过程 $\{X_t, t \in \mathbb{R}\}$ 的一维边缘分布在状态空间上具有严格单调递增的分布函数, 故对于任意的 $(\tau_1, \tau_2) \in (0, 1)^2$, 存在状态空间内唯一的两点 x_1 和 x_2 满足 $q_{\tau_i} = F^{-1}(\tau_i) = x_i$ ($i = 1, 2$), 使得

$$P(X_t \leq q_{\tau_1}, X_{t-u} \leq q_{\tau_2}) = P(X_t \leq x_1, X_{t-u} \leq x_2), \quad u \in \mathbb{R}. \quad (2.2)$$

(1) 平稳随机过程 $\{X_t, t \in \mathbb{R}\}$ 是关于时间两两独立的随机过程的充分必要条件是, 对任意状态空间中的两点 x_1 和 x_2 , 满足

$$P(X_t \leq x_1, X_{t-u} \leq x_2) = \begin{cases} P(X_t \leq x_1)P(X_{t-u} \leq x_2), & \text{若 } u \neq 0, \\ P(X_t \leq \min(x_1, x_2)), & \text{若 } u = 0. \end{cases} \quad (2.3)$$

由 (2.2), (2.3) 等价于对任意的 $(\tau_1, \tau_2) \in (0, 1)^2$, 有

$$P(X_t \leq q_{\tau_1}, X_{t-u} \leq q_{\tau_2}) = \begin{cases} P(X_t \leq q_{\tau_1})P(X_{t-u} \leq q_{\tau_2}) = \tau_1 \tau_2, & \text{若 } u \neq 0, \\ P(X_t \leq \min(q_{\tau_1}, q_{\tau_2})) = \min(\tau_1, \tau_2), & \text{若 } u = 0. \end{cases} \quad (2.4)$$

再根据分位数 - 耦合交叉协方差函数的定义 (1.1) 可知, (2.4) 等价于对任意的 $(\tau_1, \tau_2) \in (0, 1)^2$, 有

$$\begin{aligned} \gamma^{(\tau_1, \tau_2)}(u) &= P(X_t \leq q_{\tau_1}, X_{t-u} \leq q_{\tau_2}) - P(X_t \leq q_{\tau_1})P(X_{t-u} \leq q_{\tau_2}) \\ &= \begin{cases} 0, & \text{若 } u \neq 0, \\ \min(\tau_1, \tau_2) - \tau_1 \tau_2, & \text{若 } u = 0. \end{cases} \end{aligned} \quad (2.5)$$

(2) 对所有的 $u \in \mathbb{R}$, $(X_t, X_{t+u}) \stackrel{d}{=} (X_t, X_{t-u})$ 成立的充分必要条件是, 对任意的状态空间中的两点 x_1 和 x_2 , 满足

$$P(X_t \leq x_1, X_{t-u} \leq x_2) = P(X_t \leq x_1, X_{t+u} \leq x_2), \quad u \in \mathbb{R}. \quad (2.6)$$

由 (2.2), (2.6) 等价于对于任意的 $(\tau_1, \tau_2) \in (0, 1)^2$, 有

$$P(X_t \leq q_{\tau_1}, X_{t-u} \leq q_{\tau_2}) = P(X_t \leq q_{\tau_1}, X_{t+u} \leq q_{\tau_2}). \quad (2.7)$$

由 (2.5) 即知, (2.7) 定价于 $\gamma^{(\tau_1, \tau_2)}(u) = \gamma^{(\tau_1, \tau_2)}(-u)$ 对所有的 $u \in \mathbb{R}$ 成立, 即 $\gamma^{(\tau_1, \tau_2)}(u)$ 关于 u 为偶函数.

而当 $\tau_1 = \tau_2 = \tau \in (0, 1)$ 时, 由随机过程的严平稳性, (2.7) 是一定成立的, 所以, $\gamma^{(\tau, \tau)}(u)$ 关于 u 必定是偶函数. 证毕. $\square$

对大多数随机过程而言, 在不同概率对 $(\tau_1, \tau_2) \in (0, 1)^2$ 处, 分位数 - 耦合交叉协方差函数 $\gamma^{(\tau_1, \tau_2)}(u)$ 往往呈现出不同的特征, 即概率 $P(X_t \leq q_{\tau_1}, X_{t-u} \leq q_{\tau_2})$ 呈现不同的特征, 也就是随机过程在不同平稳分布分位数对 (q_{τ_1}, q_{τ_2}) 处的自相依性质不同. 下面将讨论这些由平稳分布分位数对界定的自相依性质.

2.2.1 时间可逆性

定义 2.1 设 $\{X_t, t \in \mathbb{R}\}$ 为平稳随机过程. 若对于固定的 $(\tau_1, \tau_2) \in (0, 1)^2$, $\gamma^{(\tau_1, \tau_2)}(u)$ 关于 u 为偶函数, 则称该随机过程在平稳分布分位数对 (q_{τ_1}, q_{τ_2}) 处具有时间可逆性.

注 2.1 对任意的 $\tau \in (0, 1)$, 平稳随机过程在平稳分布分位数对 (q_τ, q_τ) 处一定具有时间可逆性, 即平稳随机过程关于其平稳分布的任意分位数自身都是时间可逆的.

现考虑两类 OU 过程模型, 它们都是 OU 型随机微分方程

$$dX_t = -\lambda X_t dt + dL_t \quad (2.8)$$

的解, 其中 $\lambda > 0$ 为均值反演参数. 在模型 A 中, $\{L_t, t \in \mathbb{R}\}$ 为标准 Brown 运动; 而在模型 B 中, $\{L_t, t \in \mathbb{R}\}$ 为标准的 Cauchy-Lévy 过程, 即 L_t 具有特征函数 $E[e^{izL_t}] = \exp\{-t|z|\}$ (参见文献 [3]). 实际上, 模型 B 为文献 [11] 中所研究的由对称 α - 稳定运动驱动的 OU 型过程当 $\alpha = 1$ 时的特例. 模型 A 具有几乎处处连续的轨道, 而模型 B 是一纯跳过程. 模型 A 和 B 的一维平稳边际分布分别为 $N(0, (2\lambda)^{-1})$ 和 $S_1(\lambda^{-1}, 0)$, 其中 $N(\mu, \sigma^2)$ 代表均值为 μ 方差为 σ^2 的正态分布, $S_1(\sigma, \mu)$ 代表特征函数为 $\exp\{\sigma|z| + i\mu z\}$ 的概率分布. 并且, 在给定 $X_t = x$ 的条件下, 模型 A 和 B 转移到 X_{t+u} 的分布分别为

$$N\left(e^{-\lambda u}x, \frac{1 - e^{-2\lambda u}}{2\lambda}\right) \quad \text{和} \quad S_1\left(\frac{1 - e^{-\lambda u}}{\lambda}, e^{-\lambda u}x\right).$$

模型 A 的自协方差函数存在但模型 B 的并不存在. 对任意的概率对 $(\tau_1, \tau_2) \in (0, 1)^2$, 两模型的分位数 - 耦合交叉协方差函数都是存在的, 因此, 可以用其描述两模型的自相依性并对它们的自相依性进行比较. 取 $\lambda = 0.5$, 图 1 给出了在几个典型概率对下模型 A 与 B 的分位数 - 耦合交叉协方差函数

$$\gamma^{(\tau_1, \tau_2)}(u), \quad \tau_1, \tau_2 = 0.05, 0.50, 0.95 \quad \text{且} \quad \tau_2 \geq \tau_1 \quad (2.9)$$

的图像, 其中 “—” 为模型 A 的分位数 - 耦合交叉协方差函数, “...” 为模型 B 的分位数 - 耦合交叉协方差函数. 在图 1 及以下各图中, γ_{ij} 代表 $\gamma^{(\tau_{1i}, \tau_{2j})}(u)$ 的取值, 其中当 $i = 1, 2, 3$ 时, τ_{1i} 分别为 0.05、0.50 和 0.95; 同样, 当 $j = 1, 2, 3$ 时, τ_{2j} 也分别为 0.05、0.50 和 0.95.

图 1 模型 A 与 B 在不同概率对下分位数 - 耦合交叉协方差函数比较

由图 1 可见, 模型 A 在其平稳分布任意分位数对 (q_{τ_1}, q_{τ_2}) 处都具有时间可逆性, 模型 B 在分位数对 $(q_{0.05}, q_{0.50})$ 和 $(q_{0.50}, q_{0.95})$ 处是不具有时间可逆性的. 也就是说, 对于模型 A, $P(X_t \leq q_{\tau_1}, X_{t+u} \leq q_{\tau_2}) = P(X_t \leq q_{\tau_2}, X_{t+u} \leq q_{\tau_1})$ 对所有的概率对 (τ_1, τ_2) 都是成立的; 而对于模型 B, $P(X_t \leq q_{\tau_1}, X_{t+u} \leq q_{\tau_2}) = P(X_t \leq q_{\tau_2}, X_{t+u} \leq q_{\tau_1})$ 仅在满足 $\tau_1 = \tau_2$ 和 $\tau_1 + \tau_2 = 1$ 的概率对 (τ_1, τ_2) 处成立, 在其他概率对处未必成立. 利用自协方差函数平稳随机过程的这些时间可逆性特征是无法展示的.

2.2.2 时间相依性强弱比较

定义 2.2 设 $\{X_{1t}, t \in \mathbb{R}\}$ 和 $\{X_{2t}, t \in \mathbb{R}\}$ 为两平稳随机过程. 若对于固定的 $(\tau_1, \tau_2) \in (0, 1)^2$, 存在 $u_0 \geq 0$, 使得当 $u \geq u_0$ 时, 两随机过程的分位数 - 耦合交叉协方差函数满足

$$\frac{\gamma_2^{(\tau_1, \tau_2)}(u)}{\gamma_1^{(\tau_1, \tau_2)}(u)} \leq 1 \quad (\text{或} \geq 1) \quad (2.10)$$

恒成立, 则称随机过程 $\{X_{1t}, t \in \mathbb{R}\}$ 较 $\{X_{2t}, t \in \mathbb{R}\}$ 在其各自平稳分布分位数对 (q_{τ_1}, q_{τ_2}) 处的 u_0 - 前历史相依性更强 (或更弱). 若 (2.10) 中不等号严格成立, 则称 $\{X_{1t}, t \in \mathbb{R}\}$ 较 $\{X_{2t}, t \in \mathbb{R}\}$ 在其各自平稳分布分位数对 (q_{τ_1}, q_{τ_2}) 处的 u_0 - 前历史相依性严格更强 (或更弱). 若对于固定的 $(\tau_1, \tau_2) \in (0, 1)^2$, 存在 $u_0 \geq 0$, 使得当 $u \leq -u_0$ 时, 两随机过程的分位数 - 耦合交叉协方差函数满足

$$\frac{\gamma_2^{(\tau_1, \tau_2)}(u)}{\gamma_1^{(\tau_1, \tau_2)}(u)} \leq 1 \quad (\text{或} \geq 1) \quad (2.11)$$

恒成立, 则称随机过程 $\{X_{1t}, t \in \mathbb{R}\}$ 较 $\{X_{2t}, t \in \mathbb{R}\}$ 在其各自平稳分布分位数对 (q_{τ_1}, q_{τ_2}) 处的 u_0 - 后未来相依性更强 (或更弱). 若 (2.11) 中不等号严格成立, 则称 $\{X_{1t}, t \in \mathbb{R}\}$ 较 $\{X_{2t}, t \in \mathbb{R}\}$ 在其各自平稳分布分位数对 (q_{τ_1}, q_{τ_2}) 处的 u_0 - 后未来相依性严格更强 (或更弱).

注 2.2 随机过程 $\{X_{1t}, t \in \mathbb{R}\}$ 较 $\{X_{2t}, t \in \mathbb{R}\}$ 在其各自平稳分布分位数对 (q_{τ_1}, q_{τ_2}) 处的 u_0 - 前历史相依性更强意味着状态 q_{τ_1} 对 u_0 时间段之前状态 q_{τ_2} 的相依性随机过程 $\{X_{1t}, t \in \mathbb{R}\}$ 比 $\{X_{2t}, t \in \mathbb{R}\}$ 更强. 在概率对 (τ_1, τ_2) 处, 任意随机过程所对应的示性过程 $\{\mathbb{I}_{(0, \tau_i]}(U_t)\}$ 的一维平稳边际分布的方差都相同, 即 $\text{Var}(\mathbb{I}_{(0, \tau_i]}(U_t)) = \tau_i(1 - \tau_i)$, $i = 1, 2$. 因此, 不同随机过程在同一概率对处的分位数 - 耦合交叉协方差函数具有可比性.

图 2 给出了模型 A 和 B 分别在概率对 $(0.50, 0.50)$ 和 $(0.50, 0.95)$ 处的分位数 - 耦合交叉协方差函数的比值图像. 由图 2 可见, 模型 A 较模型 B 在其各自平稳分布分位数对 $(q_{0.50}, q_{0.50})$ 处的 0- 前历史相依性和 0- 后未来相依性都更弱; 在各自平稳分布分位数对 $(q_{0.50}, q_{0.95})$ 处, 模型 A 较模型 B 的 0- 前历史相依性更弱. 从图像的趋势可以看出, 存在 $u_0 > 0$, 使得在各自平稳分布分位数对 $(q_{0.50}, q_{0.95})$ 处, 模型 A 较模型 B 的 u_0 - 后未来相依性更弱.

2.2.3 状态相依性

定义 2.3 设 $\{X_t, t \in \mathbb{R}\}$ 为平稳随机过程. 若对于固定的 $(\tau_1, \tau_2) \in (0, 1)^2$, 当 $u \neq 0$ 时, $\gamma^{(\tau_1, \tau_2)}(u) \equiv 0$, 则称该随机过程在平稳分布分位数对 (q_{τ_1}, q_{τ_2}) 处具有状态独立性.

注 2.3 平稳分布的分位数 q_{τ_1} 和 q_{τ_2} 实际上是随机过程状态空间中的两点, 随机过程在平稳分布分位数对 (q_{τ_1}, q_{τ_2}) 处具有状态独立性实际上是这两个状态关于时间没有相依性. 若当 $u \neq 0$ 时, $\gamma^{(\tau_1, \tau_2)}(u) \neq 0$, 但 $\gamma^{(\tau_1, \tau_2)}(u)$ 的值都非常接近于 0, 这说明随机过程平稳分布分位数 q_{τ_1} 与 q_{τ_2} 关于时间的相依性较弱.

图 2 模型 A 与 B 在不同概率对下历史与未来相依性比较

由图 1 可见, 在概率对 $(0.05, 0.95)$ 处, 模型 A 与 B 的分位数 - 耦合交叉协方差函数的值都非常接近于 0, 这说明模型 A 与 B 状态空间中的正极端值与负极端值的时间相依性都非常弱, 接近独立.

3 分位数 - 耦合交叉协方差函数的估计

3.1 估计量

设 $\{t_k = kh, k = 1, 2, \dots, n\}$ 为平稳随机过程 $\{X_t\}$ 的等距时间观测点, 其中 $h > 0$ 为观测的时间间距. 平稳随机过程的一维平稳边际分布函数 $F(x)$ 用其经验版本

$$F_{n,h}(x) = \frac{1}{n} \sum_{k=1}^n \mathbb{I}_{(-\infty, x]}(X_{t_k})$$

来估计. 在引入 $F_{n,h}(x)$ 后, $U_{t_k} = F(X_{t_k})$ 可以用 $\hat{U}_{t_k} = F_{n,h}(X_{t_k})$ 进行估计, $k = 1, \dots, n$. 由于 (1.1) 定义的分位数 - 耦合交叉协方差函数也可以写作

$$\gamma^{(\tau_1, \tau_2)}(u) = P(U_t \leq \tau_1, U_{t-u} \leq \tau_2) - \tau_1 \tau_2, \quad (3.1)$$

因此, 对任意的概率对 $(\tau_1, \tau_2) \in (0, 1)^2$, $\gamma^{(\tau_1, \tau_2)}(kh)$ 可由其经验版本

$$\hat{\gamma}^{(\tau_1, \tau_2)}(kh) = \begin{cases} \frac{1}{n-k} \sum_{i=1}^{n-k} \mathbb{I}_{(0, \tau_1]}(\hat{U}_{t_{i+k}}) \mathbb{I}_{(0, \tau_2]}(\hat{U}_{t_i}) - \tau_1 \tau_2, & \text{若 } k \geq 0, \\ \frac{1}{n+k} \sum_{i=1}^{n+k} \mathbb{I}_{(0, \tau_1]}(\hat{U}_{t_i}) \mathbb{I}_{(0, \tau_2]}(\hat{U}_{t_{i-k}}) - \tau_1 \tau_2, & \text{若 } k < 0 \end{cases} \quad (3.2)$$

来估计. 对任意的 $u \in \mathbb{R}$, 将 $\hat{\gamma}^{(\tau_1, \tau_2)}(u)$ 定义为阶梯型函数

$$\hat{\gamma}^{(\tau_1, \tau_2)}(u) = \begin{cases} \hat{\gamma}^{(\tau_1, \tau_2)}(kh), & \text{若 } u \in [-(n-1)h, (n-1)h] \text{ 且 } kh - \frac{h}{2} \leq u < kh + \frac{h}{2}, \\ 0, & \text{其他,} \end{cases} \quad (3.3)$$

其中 $\hat{\gamma}^{(\tau_1, \tau_2)}(kh)$ 由 (3.2) 定义, k 为满足 $|k| \leq n-1$ 且使 $kh - h/2 \leq u < kh + h/2$ 成立的整数 (即 $k = [u/h + 1/2]$, 其中 $[x]$ 为不超过 x 的最大整数). 本文用由 (3.3) 定义的 $\hat{\gamma}^{(\tau_1, \tau_2)}(u)$ 作为 $\gamma^{(\tau_1, \tau_2)}(u)$ 的估计量, 以下讨论 $\hat{\gamma}^{(\tau_1, \tau_2)}(u)$ 的渐近性质.

3.2 渐近性质

设 $\mathcal{G}_0^- = \sigma(X_t, t \leq 0)$ 和 $\mathcal{G}_u^+ = \sigma(X_t, t \geq u)$, 即 $\mathcal{G}_0^-$ 和 $\mathcal{G}_u^+$ 分别为由 $\{X_t, t \leq 0\}$ 和 $\{X_t, t \geq u\}$ 生成的 σ -代数. 对任意的 $u > 0$, 记

$$\alpha_u = \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|; A \in \mathcal{G}_0^-, B \in \mathcal{G}_u^+\}.$$

为得到 $\hat{\gamma}^{(\tau_1, \tau_2)}(u)$ 的渐近结论, 需要平稳随机过程的如下假设.

假设 3.1 存在 $a > 1$ 使得 $\alpha_u = O(u^{-a})$.

假设 3.2 平稳随机过程 $\{X_t\}$ 是随机连续的, 即对任意的 $\varepsilon > 0$ 和任意的 $t \in \mathbb{R}$, 有

$$\lim_{s \rightarrow t} \mathbb{P}(|X_s - X_t| > \varepsilon) = 0.$$

假设 3.3 平稳随机过程的一维平稳边际分布函数 $F(x)$ 的导数在状态空间上有界.

假设 3.1 为平稳随机过程 $\{X_t\}$ 的混合性 (mixing) 条件. 当假设 3.1 成立时, 由弱相依过程的经验分布函数在 Skorohod 空间上的泛函收敛定理 (参见文献 [12, 定理 5.3]) 可得, 当 $n \rightarrow \infty$, $h \rightarrow 0$ 且 $nh^a \rightarrow \infty$ 时,

$$\sup_{x \in \mathbb{R}} |F_{n,h}(x) - F(x)| = O_P\left(\frac{1}{\sqrt{nh^a}}\right) \quad (3.4)$$

成立. 还存在其他多个使得 (3.4) 成立的弱相依性条件, 但为了使所提供的条件不依赖于二阶矩的存在性且在证明中易于技术操作, 这里仅给出了假设 3.1 中的条件. 假设 3.2 是证明过程中的技术性条件. 假设 3.3 是为了保证当 h 较小时 (3.3) 提供的估计量可以用于近似 $\gamma^{(\tau_1, \tau_2)}(u)$.

定理 3.1 设 $\{X_t, t \in \mathbb{R}\}$ 为严平稳的随机过程, 且假设 3.1–3.3 都成立, 则对任意的 $u \in \mathbb{R}$, 当 $n \rightarrow \infty$, $h \rightarrow 0$ 且 $nh^a \rightarrow \infty$ 时, 有 $\hat{\gamma}^{(\tau_1, \tau_2)}(u) \xrightarrow{P} \gamma^{(\tau_1, \tau_2)}(u)$, 其中 “ $\xrightarrow{P}$ ” 代表依概率收敛.

证明 对任意的 $u \in \mathbb{R}$, 由 (3.3) 可知, 当 n 充分大时, 一定存在整数 k 满足 $kh - h/2 \leq u < kh + h/2$, 且使得

$$\hat{\gamma}^{(\tau_1, \tau_2)}(u) - \gamma^{(\tau_1, \tau_2)}(u) = \hat{\gamma}^{(\tau_1, \tau_2)}(kh) - \gamma^{(\tau_1, \tau_2)}(u), \quad (3.5)$$

这里, k 实际上是 u 和 h 的函数, 为简练起见省略了自变量.

作为代表, 仅证 $u \geq 0$ 的情形. 由 (3.5)、(3.1) 和 (3.2) 可知,

$$|\hat{\gamma}^{(\tau_1, \tau_2)}(u) - \gamma^{(\tau_1, \tau_2)}(u)| = \left| \frac{1}{n-k} \sum_{i=1}^{n-k} \mathbb{I}_{(0, \tau_1]}(\hat{U}_{t_i+k}) \mathbb{I}_{(0, \tau_2]}(\hat{U}_{t_i}) - \mathbb{P}(U_{t+u} \leq \tau_1, U_t \leq \tau_2) \right|. \quad (3.6)$$

对 (3.6) 右端使用三角不等式放大, 可得

$$|\hat{\gamma}^{(\tau_1, \tau_2)}(u) - \gamma^{(\tau_1, \tau_2)}(u)| \leq \left| \frac{1}{n-k} \sum_{i=1}^{n-k} (\mathbb{I}_{(0, \tau_1]}(\hat{U}_{t_i+k}) - \mathbb{I}_{(0, \tau_1]}(U_{t_i+u})) \mathbb{I}_{(0, \tau_2]}(\hat{U}_{t_i}) \right|$$

$$\begin{aligned}
& + \left| \frac{1}{n-k} \sum_{i=1}^{n-k} \mathbb{I}_{(0,\tau_1]}(U_{t_i+u}) (\mathbb{I}_{(0,\tau_2]}(\hat{U}_{t_i}) - \mathbb{I}_{(0,\tau_2]}(U_{t_i})) \right| \\
& + \left| \frac{1}{n-k} \sum_{i=1}^{n-k} \mathbb{I}_{(0,\tau_1]}(U_{t_i+u}) \mathbb{I}_{(0,\tau_2]}(U_{t_i}) - \mathbb{P}(U_{t+u} \leq \tau_1, U_t \leq \tau_2) \right| \\
& \leq \frac{1}{n-k} \sum_{i=1}^{n-k} |\mathbb{I}_{(0,\tau_1]}(\hat{U}_{t_{i+k}}) - \mathbb{I}_{(0,\tau_1]}(U_{t_i+u})| + \frac{1}{n-k} \sum_{i=1}^{n-k} |\mathbb{I}_{(0,\tau_2]}(\hat{U}_{t_i}) - \mathbb{I}_{(0,\tau_2]}(U_{t_i})| \\
& + \left| \frac{1}{n-k} \sum_{i=1}^{n-k} \mathbb{I}_{(0,\tau_1]}(U_{t_i+u}) \mathbb{I}_{(0,\tau_2]}(U_{t_i}) - \mathbb{P}(U_{t+u} \leq \tau_1, U_t \leq \tau_2) \right| \\
& =: I_{n,h}^{(1)} + I_{n,h}^{(2)} + I_{n,h}^{(3)}. \tag{3.7}
\end{aligned}$$

由 $kh - h/2 \leq u < kh + h/2$ 可得 $n - u/h - 1/2 \leq n - k < n - u/h + 1/2$. 又由 $nh^a \rightarrow \infty$ 得 $n \geq C \frac{1}{h^a}$, 其中 $C > 0$ 为常数. 因 $a > 1$, 故在定理条件下, $n - k \rightarrow \infty$.

由 (3.4), 可知 $F_{n,h}(x) = F(x) + o_{\mathbb{P}}(\sqrt{\log(nh^a)/(nh^a)})$, 其中 $o_{\mathbb{P}}(\sqrt{\log(nh^a)/(nh^a)})$ 取为与 x 无关的量. 因此,

$$\begin{aligned}
\mathbb{E}|\mathbb{I}_{(0,\tau_2]}(\hat{U}_{t_i}) - \mathbb{I}_{(0,\tau_2]}(U_{t_i})| &= \mathbb{E}|\mathbb{I}_{(0,\tau_2]}(F_{n,h}(X_{t_i})) - \mathbb{I}_{(0,\tau_2]}(F(X_{t_i}))| \\
&= \mathbb{E} \left| \mathbb{I}_{(0,\tau_2]} \left(U_{t_i} + o_{\mathbb{P}} \left(\sqrt{\frac{\log(nh^a)}{nh^a}} \right) \right) - \mathbb{I}_{(0,\tau_2]}(U_{t_i}) \right| \\
&\leq o \left(\sqrt{\frac{\log(nh^a)}{nh^a}} \right), \tag{3.8}
\end{aligned}$$

其中 $U_{t_i} \sim U[0, 1]$. 于是, 对任意的 $\varepsilon > 0$, 由 Markov 不等式和三角不等式, 得

$$\mathbb{P}(I_{n,h}^{(2)} > \varepsilon) \leq \frac{\mathbb{E}(I_{n,h}^{(2)})}{\varepsilon} \leq \frac{\mathbb{E}|\mathbb{I}_{(0,\tau_2]}(\hat{U}_{t_i}) - \mathbb{I}_{(0,\tau_2]}(U_{t_i})|}{\varepsilon} \leq \varepsilon^{-1} o \left(\sqrt{\frac{\log(nh^a)}{nh^a}} \right) \rightarrow 0,$$

即

$$I_{n,h}^{(2)} \xrightarrow{\mathbb{P}} 0. \tag{3.9}$$

由三角不等式, 得

$$\begin{aligned}
& \mathbb{E}|\mathbb{I}_{(0,\tau_1]}(\hat{U}_{t_{i+k}}) - \mathbb{I}_{(0,\tau_1]}(U_{t_i+u})| \\
& \leq \mathbb{E}|\mathbb{I}_{(0,\tau_1]}(\hat{U}_{t_{i+k}}) - \mathbb{I}_{(0,\tau_1]}(U_{t_{i+k}})| + \mathbb{E}|\mathbb{I}_{(0,\tau_1]}(U_{t_{i+k}}) - \mathbb{I}_{(0,\tau_1]}(U_{t_i+u})|. \tag{3.10}
\end{aligned}$$

用证 (3.8) 相同的方法, 可得

$$\mathbb{E}|\mathbb{I}_{(0,\tau_1]}(\hat{U}_{t_{i+k}}) - \mathbb{I}_{(0,\tau_1]}(U_{t_{i+k}})| \leq o \left(\sqrt{\frac{\log(nh^a)}{nh^a}} \right). \tag{3.11}$$

注意到

$$\begin{aligned}
\mathbb{E}|\mathbb{I}_{(0,\tau_1]}(U_{t_{i+k}}) - \mathbb{I}_{(0,\tau_1]}(U_{t_i+u})| &= \mathbb{E}|\mathbb{I}_{(0,\tau_1]}(U_{kh}) - \mathbb{I}_{(0,\tau_1]}(U_u)| \\
&\leq \mathbb{P}(X_{kh} \leq q_{\tau_1}, X_u > q_{\tau_1}) + \mathbb{P}(X_u \leq q_{\tau_1}, X_{kh} > q_{\tau_1}). \tag{3.12}
\end{aligned}$$

而对任意的 $\varepsilon > 0$, 有

$$\begin{aligned}
 & P(X_{kh} \leq q_{\tau_1}, X_u > q_{\tau_1}) \\
 &= P(X_{kh} \leq q_{\tau_1}, X_u > q_{\tau_1}, |X_u - X_{kh}| > \varepsilon) + P(X_{kh} \leq q_{\tau_1}, X_u > q_{\tau_1}, |X_u - X_{kh}| \leq \varepsilon) \\
 &\leq P(|X_u - X_{kh}| > \varepsilon) + P(q_{\tau_1} < X_u \leq q_{\tau_1} + \varepsilon) \\
 &\leq P(|X_u - X_{kh}| > \varepsilon) + P(\tau_1 < U_u \leq \tau_1 + C\varepsilon),
 \end{aligned} \tag{3.13}$$

其中 $C > 0$ 为 $F(x)$ 导数的上界. 在 (3.13) 中, 先令 $n \rightarrow \infty$, 这时 $h \rightarrow 0$, 由假设 3.2 得 $P(|X_u - X_{kh}| > \varepsilon) \rightarrow 0$; 再令 $\varepsilon \rightarrow 0$, 得 $P(\tau_1 < U_u \leq \tau_1 + C\varepsilon) = C\varepsilon \rightarrow 0$. 这样便证明了 $P(X_{kh} \leq q_{\tau_1}, X_u > q_{\tau_1}) \rightarrow 0$, 用同样方法可以证明 $P(X_u \leq q_{\tau_1}, X_{kh} > q_{\tau_1}) \rightarrow 0$. 由 (3.12) 知,

$$E|\mathbb{I}_{(0, \tau_1]}(U_{t_i+k}) - \mathbb{I}_{(0, \tau_1]}(U_{t_i+u})| \rightarrow 0. \tag{3.14}$$

结合 (3.10)、(3.11) 和 (3.14), 得

$$E|\mathbb{I}_{(0, \tau_1]}(\hat{U}_{t_i+k}) - \mathbb{I}_{(0, \tau_1]}(U_{t_i+u})| \rightarrow 0.$$

采用与证 (3.9) 相同的方法, 可得

$$I_{n,h}^{(1)} \xrightarrow{P} 0. \tag{3.15}$$

由 Birkhoff 遍历定理 (参见文献 [13, 第 337 页]), 可得

$$I_{n,h}^{(3)} \xrightarrow{P} 0. \tag{3.16}$$

融合 (3.7)、(3.9)、(3.15) 和 (3.16), 可得 $\hat{\gamma}^{(\tau_1, \tau_2)}(u) \xrightarrow{P} \gamma^{(\tau_1, \tau_2)}(u)$. 证毕. $\square$

4 数值分析

在得到分位数 - 耦合交叉协方差函数的估计量 (3.3) 后, 本节首先通过模拟实验验证估计量 (3.3) 的小样本性质; 然后, 探讨具有相同一维平稳边际分布和相同自协方差函数但类型不同的两类平稳随机过程 (OU 型过程与扩散过程) 自相依性质的差异; 最后, 利用估计量 (3.3) 研究船体振动实测数据的自相依性质.

4.1 估计量小样本性质

由于由不同 Lévy 过程驱动的随机微分方程 (2.8) 的两实例 (模型 A 与 B) 在不同概率对 $(\tau_1, \tau_2) \in (0, 1)^2$ 下的分位数 - 耦合交叉协方差函数都可以通过数值方式求解, 因此, 这里仍用模型 A 与 B 来评估估计量 (3.3) 的小样本性质, 其中参数 λ 和 $(\tau_1, \tau_2) \in (0, 1)^2$ 的设定与图 1 中的情形完全相同.

对每个模型考虑四种不同的抽样方案, 即 n 取 128、256、512 和 1,024 四种情形, 且设定 $h = 10n^{-1/3}$, 这样的设定与定理 3.1 中的条件 $n \rightarrow \infty$ 、 $h \rightarrow 0$ 且 $nh^a \rightarrow \infty$ (可取 $a = 2$) 相吻合. 对每个模型及每种抽样方案, 各重复抽样 1,000 次. 对每个概率对 (τ_1, τ_2) 以及每次重复抽样, 皆计算 RMSE (root mean square error) 如下:

$$RMSE = \sqrt{\frac{1}{K} \sum_{k=1}^K (\hat{\gamma}^{(\tau_1, \tau_2)}(u_k) - \gamma^{(\tau_1, \tau_2)}(u_k))^2}, \tag{4.1}$$

其中取

$$K = 50, \quad u_k = -10 + \frac{2}{5} \frac{k-1}{K-1}, \quad k = 1, \dots, K,$$

即在 $u \in [-10, 10]$ 区间内等距取 50 个点评价 $\hat{\gamma}^{(\tau_1, \tau_2)}(u)$ 与 $\gamma^{(\tau_1, \tau_2)}(u)$ 的差别. 最后, 计算 RMSE 1,000 次实现值的均值和标准偏差 (standard deviation) 如表 1 所示, 其中括号内的数字为标准偏差.

由表 1 可见, (1) 随着抽样区间的扩大和抽样频率的增加, 由 (4.1) 定义的 RMSE 的均值不断变小, 并且其值也更趋于稳定 (标准偏差也变小); (2) 对于由不同矩性质的 Lévy 过程驱动的 OU 过程而言, 估计量 (3.3) 与真值的偏差和稳定性都非常类似; (3) 在概率对 (0.50, 0.50) 处, 估计量 (3.3) 的 RMSE 的均值较大, 而在概率对 (0.05, 0.95) 处, RMSE 的均值较小, 这些都与随机过程在不同状态处的时间相依性强弱相吻合 (见图 1).

4.2 过程相依特征比较

考虑两类一维平稳边际分布与自协方差函数完全相同的模型—模型 C 与 D. 模型 C 为 OU 型随机微分方程

$$dX_t = -\lambda X_t dt + dZ_{\lambda t} \quad (4.2)$$

的解, 其中 $Z_t = \sum_{i=1}^{N_t} Y_i$ 为带有均值为 $1/\alpha$ 的指数分布跳 $\{Y_i\}$ 的复合 Poisson 分布, $\{N_t, t \geq 0\}$ 为强度参数为 ν 的 Poisson 过程. 模型 C 的平稳分布为 Gamma 分布, 因此称为 Gamma-OU 过程 (参见文献 [14, 15]), 主要用于随机波动率模型中波动率的刻画. 模型 D 为扩散型随机微分方程

$$dX_t = \lambda \left(\frac{\nu}{\alpha} - X_t \right) dt + \sqrt{\frac{2\lambda X_t}{\alpha}} dW_t \quad (4.3)$$

的解, 其中 $\{W_t, t \geq 0\}$ 为标准 Brown 运动. 模型 D 称为 CIR 模型 (参见文献 [16]), 常被用于刻画短期利率等. 模型 C 与 D 中的参数 λ 、 ν 和 α 都为正实数, (4.2) 与 (4.3) 的参数化使得两模型具有相同的一维平稳边际分布 $\Gamma(\nu, \alpha)$ 分布 (其密度为 $\frac{\alpha^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\alpha x} \mathbb{I}_{(0, \infty)}(x)$) 与相同的自协方差函数

$$C(u) = \frac{\nu}{\alpha^2} e^{-\lambda|u|}.$$

并且, 模型 C 与 D 都满足假设 3.1 的混合性条件, 可分别参见文献 [17, 定理 4.3] 和 [18, 定理 2.5].

表 1 不同 (τ_1, τ_2) 下模型 A 和 B RMSE 1,000 次实现的均值和标准偏差

模型	n	h	nh	(τ_1, τ_2)					
				(0.05, 0.05)	(0.05, 0.50)	(0.05, 0.95)	(0.50, 0.50)	(0.50, 0.95)	(0.95, 0.95)
A	128	1.98	254	0.008 (0.001)	0.010 (0.002)	0.005 (0.002)	0.037 (0.006)	0.015 (0.005)	0.013 (0.003)
	256	1.57	403	0.007 (0.001)	0.007 (0.002)	0.004 (0.001)	0.033 (0.003)	0.009 (0.003)	0.008 (0.001)
	512	1.25	640	0.007 (0.000)	0.005 (0.001)	0.002 (0.001)	0.030 (0.002)	0.006 (0.002)	0.007 (0.000)
	1,024	0.99	1,016	0.004 (0.000)	0.004 (0.001)	0.001 (0.000)	0.018 (0.002)	0.004 (0.001)	0.005 (0.000)
B	128	1.98	254	0.007 (0.002)	0.010 (0.003)	0.005 (0.003)	0.033 (0.009)	0.015 (0.005)	0.013 (0.004)
	256	1.57	403	0.006 (0.002)	0.007 (0.002)	0.004 (0.002)	0.026 (0.005)	0.009 (0.003)	0.007 (0.002)
	512	1.25	640	0.005 (0.001)	0.005 (0.002)	0.002 (0.001)	0.022 (0.003)	0.006 (0.002)	0.005 (0.001)
	1,024	0.99	1,016	0.004 (0.001)	0.004 (0.001)	0.002 (0.001)	0.015 (0.003)	0.005 (0.002)	0.004 (0.001)

从一维平稳边际分布与自协方差函数两个方面很难对模型 C 与 D 的自相依性特征进行区分, 但事实上它们的相依性是有差异的, 因为模型 C 与 D 的转移分布截然不同 (参见文献 [15, 16]). 由于 (X_t, X_{t-u}) 的联合分布函数没有闭形式 (closed-form), 直接计算模型 C 与 D 的分位数 - 耦合交叉协方差函数是不方便的. 因此, 通过模拟产生两模型的轨道, 利用分位数 - 耦合交叉协方差函数的估计对其自相依性特征进行比较研究, 其中 Gamma-OU 过程的模拟算法参见文献 [19], CIR 模型的模拟算法参见文献 [20]. 参数设定为 $\lambda = 0.5$, $\nu = 3$, $\alpha = 6$, 取 $n = 5 \times 10^5$, $h = 0.2$, 各产生两模型的一条轨道, 然后利用估计量 (3.3) 对分位数 - 耦合交叉协方差函数 (2.9) 进行估计, 如图 3 所示, 其中 “—” 为模型 C 一条模拟轨道的分位数 - 耦合交叉自协方差函数估计, “...” 为模型 D 一条模拟轨道的分位数 - 耦合交叉协方差函数估计.

由图 3 可见, (1) 模型 D 在其平稳分布各分位数对 (q_{τ_1}, q_{τ_2}) 处都具有时间可逆性, 而模型 C 在其平稳分布分位数对 $(q_{0.05}, q_{0.50})$ 、 $(q_{0.05}, q_{0.95})$ 和 $(q_{0.50}, q_{0.95})$ 处都不具有时间可逆性; (2) 在有限滞后时间内, 模型 C 平稳分布分位数 $q_{0.05}$ 对 $q_{0.50}$ 和 $q_{0.95}$ 的历史相依性都强于模型 D, 模型 D 平稳分布分位数 $q_{0.50}$ 和 $q_{0.95}$ 对其自身的 u_0 - 前历史相依性与 u_0 - 后未来相依性都强于模型 C (其中 $u_0 \geq 0$ 为某常数), 但两模型分位数 $q_{0.05}$ 对其自身的历史相依性与未来相依性较难判断 (见图 4, 其中给出了 γ_D/γ_C 在不同分位数对处的图像); (3) 模型 C 与 D 下极端状态与上极端状态的相依性都较小, 接近独立 (见图 3 中 γ_{13} 的图像).

4.3 船体振动数据分析

船体主梁的弯矩一般由船舶的振动所驱动, 二者经常存在线性或非线性的过滤 (filter) 关系 (参见

图 3 模型 C 与 D 在不同概率对下分位数 - 耦合交叉协方差函数比较

图 4 模型 C 与 D 在不同概率对下历史与未来相依性比较

文献 [21]). 因此, 分析船体的振动性质对研究航行中船舶的主梁弯矩具有重要的实践意义. 船体垂向加速度 (vertical acceleration, VAC) 是船舶振动的一个重要指标. 通过最大经验似然比检验, 文献 [22] 建议用具有特殊核函数的一类 Gauss 滑动平均过程 $X_t = \sigma \int_{\mathbb{R}} K(t-s) dW_s$ 对 VAC 序列进行建模, 其中 $\sigma > 0$ 为标准差参数, $\{W_s\}$ 为标准 Brown 运动, 核函数为

$$K(s) = \cos(\omega s) e^{-\lambda s} \mathbb{I}_{(0, \infty)}(s).$$

这里, 将用分位数 - 耦合交叉协方差函数对文献 [22] 中的 VAC 序列 (单位: 米/秒²) 与 Gauss 滑动平均过程的自相依性进行比较. 所分析的 VAC 序列取自安装在上海海事大学教学实习船“育锋轮”上的船体强度实时监测系统在船首一测点处的测量值. 序列 VAC 覆盖一个小时的测量值, 2009 年 12 月 13 日下午 3 至 4 时, 共计 1,801 个测量值, 即每 2 秒钟取得一个测量值 (见图 5).

利用经验似然方法, 文献 [22] 中给出的用 Gauss 滑动平均过程拟合 VAC 序列的参数值为 $\lambda = 0.260$, $\sigma^2 = 0.149$ 和 $\omega = 2.054$. 借助估计量 (3.3), 对 VAC 序列和文献 [22] 中拟合参数 VAC 序列的 Gauss 滑动平均过程模拟轨道 (模拟方法参见文献 [22, 23], 其中设 $h = 1$ 代表 2 秒, $n = 1,801$) 的分位数 - 耦合交叉协方差函数进行估计. 对估计的结论进行观察和比较发现, (1) VAC 序列是具有时间可逆性的过程; (2) 在平稳分布中位数对 $(q_{0.50}, q_{0.50})$ 处, 文献 [22] 中推荐的 Gauss 滑动平均过程与 VAC 序列的分位数 - 耦合交叉协方差函数十分一致, 但在平稳分布分位数对 $(q_{0.50}, q_{0.95})$ 处, Gauss 滑动平均过程与 VAC 序列的分位数 - 耦合交叉协方差函数出现较大的差异 (如图 6 所示, 其中 “—” 为 VAC

图 5 船体垂向加速度序列 VAC

图 6 VAC 序列与 Gauss 滑动平均在不同概率对下的分位数 - 耦合交叉自协方差函数比较

序列的分位数 - 耦合交叉协方差函数的估计, “...” 为文献 [22] 推荐的 Gauss 滑动平均过程一条模拟轨道的分位数 - 耦合交叉协方差函数的估计). 这说明, 文献 [22] 推荐的 Gauss 滑动平均过程并不能完全拟合 VAC 序列的自相依特征, 该结果与文献 [9] 中对该序列进行 Bayes 耦合谱分析的结论一致. 主要原因在于, 文献 [22] 中的经验似然方法主要建立在二阶矩条件基础上, 不能全面捕获 VAC 序列的自相依特征.

5 结论

同时基于分位数和耦合的概念, 本文提出了描述平稳随机过程自相依性的一个概念: 分位数 - 耦合交叉协方差函数. 分位数 - 耦合交叉协方差函数将对 X_t 和 X_{t-u} 的相依性讨论转化为 $U_t = F(X_t)$ 和 $U_{t-u} = F(X_{t-u})$ 在其平稳分布不同分位点处的相依性讨论, 这一转化将随机过程的状态空间规范化, 从而弱化了估计量 (3.3) 对随机过程 $\{X_t\}$ 有限维分布性质 (如矩存在性) 的依赖性. 例如, 矩性质差异较大的两模型 (如模型 A 与 B) 的分位数 - 耦合交叉协方差函数估计量的小样本性质 (如均方误差) 差别较小. 相对于自协方差函数, 该函数可以描述平稳过程更丰富的自相依特征, 如时间可逆性、状态相依性、二阶矩不存在时的长程相依性等.

以下给出借助分位数 - 耦合交叉协方差函数分析随机过程自相依性时的几点应用和研究建议.

(1) 构造具有渐近正态性的估计量. 基于离散抽样, 本文给出了分位数 - 耦合交叉协方差函数的一个具有相合性的估计量 (3.3). 一方面, 估计量 (3.3) 的计算量小, 使用较为方便. 但另一方面, 在估计量 (3.3) 的构造过程中使用了二重估计, 即 $\hat{\gamma}^{(\tau_1, \tau_2)}(u)$ 为过程 $\mathbb{I}_{(0, \tau_1]}(\hat{U}_t)$ 与 $\mathbb{I}_{(0, \tau_2]}(\hat{U}_t)$ 的经验交叉协方差函数, 而 $\{\hat{U}_t\}$ 又是 $\{U_t\}$ 通过 $F_{n,h}(x)$ 得到的估计量. 这使得估计量 $\hat{\gamma}^{(\tau_1, \tau_2)}(u)$ 相对于 $\gamma^{(\tau_1, \tau_2)}(u)$ 的收敛速度依赖于 $F_{n,h}(x)$ 收敛到 $F(x)$ 的收敛速度, 而 $F_{n,h}(x)$ 收敛到 $F(x)$ 的收敛速度由 (3.4) 给出, 且无法进一步提高, 致使讨论 $\hat{\gamma}^{(\tau_1, \tau_2)}(u)$ 关于 $\gamma^{(\tau_1, \tau_2)}(u)$ 的渐近正态性十分困难. 因此, 构造分位数 - 耦合交叉协方差函数具有渐近正态性的估计量是一个值得进一步研究的问题.

(2) 时间可逆性检验. 借助分位数 - 耦合交叉协方差函数可以检验随机过程的时间可逆性, 即检验

$$H_0: \text{对所有的 } (\tau_1, \tau_2) \in (0, 1)^2, \gamma^{(\tau_1, \tau_2)}(u) \text{ 关于 } u \text{ 都为偶函数.}$$

(3) 关于时间两两独立性的检验. 借助分位数 - 耦合交叉协方差函数可以检验平稳随机过程本身是否关于时间具有两两独立性, 也可以通过随机过程增量分位数 - 耦合交叉协方差函数的研究检验一个随机过程的平稳增量是否具有两两独立性. 由命题 2.1, 这类问题实际上是检验

$$H_0: \gamma^{(\tau_1, \tau_2)}(u) = 0 \text{ 对所有的 } u \neq 0 \text{ 和 } (\tau_1, \tau_2) \in (0, 1)^2 \text{ 都成立.}$$

(4) 平稳 Gauss 过程之外的时间可逆过程. 因两变量 Gauss 分布完全由其协方差阵决定, 故平稳的 Gauss 过程一定是时间可逆的过程. 例如, Brown 运动驱动的 OU 型过程、CIR 模型和 Gauss 滑动平均等都具有时间可逆性. 但是, 还有哪些其他典型过程也一定具有时间可逆性尚待进一步考证.

(5) 时域研究的优势. 本文所提出的分位数 - 耦合交叉协方差函数为随机过程时域研究的方法. 相对于频域研究的方法 (如文献 [7, 9] 等), 时域研究可以探讨随机过程状态相依性的强弱、不同随机过程时间相依性的强弱等, 而这些工作在频域内研究都较为困难.

致谢 作者对审稿人提出的宝贵意见表示感谢.

参考文献

- Beran J, Feng Y, Ghosh S, et al. Long-Memory Processes: Probabilistic Properties and Statistical Methods. London: Springer, 2013
- Ibragimov I A, Linnik Y V. Independent and Stationary Sequences of Random Variables. Netherlands: Wolters-Noordhoff Publishing Groningen, 1971
- Sato K. Lévy Processes and Infinitely Divisible Distributions. Cambridge: Cambridge University Press, 1999
- Nelsen R B. An Introduction to Copulas, 2nd ed. New York: Springer, 2006
- Hong Y. Generalized spectral tests for serial dependence. J R Stat Soc Ser B Stat Methodol, 2000, 62: 557–574
- Li T H. Quantile periodograms. J Amer Statist Assoc, 2012, 107: 765–776
- Dette H, Hallin M, Kley T, et al. Of copulas, quantiles, ranks and spectra: An L_1 -approach to spectral analysis. Bernoulli, 2015, 21: 781–831
- Kley T, Volgushev S, Dette H, et al. Quantile spectral processes: Asymptotic analysis and inference. Bernoulli, 2016, 22: 1770–1807
- Zhang S. Bayesian copula spectral analysis for stationary time series. Comput Statist Data Anal, 2018, doi: 10.1016/j.csda.2018.10.001
- Birr S, Volgushev S, Kley T, et al. Quantile spectral analysis for locally stationary time series. J R Stat Soc Ser B Stat Methodol, 2017, 79: 1619–1643
- Zhang S, Zhang X. A least squares estimator for discretely observed Ornstein-Uhlenbeck processes driven by symmetric α -stable motions. Ann Inst Statist Math, 2013, 65: 89–103
- Ango Nze P, Doukhan P. Weak dependence: Models and applications to econometrics. Econometric Theory, 2004, 20: 995–1045
- Durrett R. Probability: Theory and Examples, 3rd ed. Belmont: Thomson, 2004
- Barndorff-Nielsen O E, Shephard N. Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. J R Stat Soc Ser B Stat Methodol, 2001, 63: 167–241
- Zhang S, Zhang X, Sun S. Parametric estimation of discretely sampled Gamma-OU processes. Sci China Ser A, 2006, 49: 1231–1257
- Cox J C, Ingersoll J E Jr, Ross S A. A theory of the term structure of interest rates. Econometrica, 1985, 53: 385–407
- Masuda H. On multidimensional Ornstein-Uhlenbeck processes driven by a general Lévy process. Bernoulli, 2004, 10: 97–120
- Li Z, Ma C. Asymptotic properties of estimators in a stable Cox-Ingersoll-Ross model. Stochastic Process Appl, 2015, 125: 3196–3233
- Zhang S. Transition law-based simulation of generalized inverse Gaussian Ornstein-Uhlenbeck processes. Methodol Comput Appl Probab, 2011, 13: 619–656
- Iacus S M. Simulation and Inference for Stochastic Differential Equations. New York: Springer, 2008
- Pedersen P T, Jensen J J. Estimation of hull girder vertical bending moments including non-linear and flexibility effects using closed form expressions. Proc Inst Mech Engrg, 2009, 223: 377–390

- 22 Zhang S. Empirical likelihood methods for discretely observed Gaussian moving averages. *J Stat Comput Simul*, 2016, 86: 942–958
- 23 Wood A T A, Chan G. Simulation of stationary Gaussian processes in $[0, 1]^d$. *J Comput Graph Statist*, 1994, 3: 409–432

Dependence diagnosis for stationary stochastic processes based on both quantiles and copulas

Shibin Zhang & Xinsheng Zhang

Abstract Based on both quantiles and copulas, a new concept, the quantile-copula cross-covariance function, for stationary stochastic processes, is given. The function can capture many dependence characteristics that the traditional auto-covariance function cannot cover, such as time reversibility, state dependence as well as long-range dependence when the auto-covariance function does not exist. A consistent estimator of the quantile-copula cross-covariance function is proposed based on discretely sampled stationary processes. By comparing the estimated quantile-copula cross-covariance functions, the dependence differences are distinguished for two types of processes, Gamma-OU processes and CIR models, which have the common one-dimensional stationary marginal distribution and the common auto-covariance function. By analyzing the estimated quantile-copula cross-covariance function of a ship vibration series, some suggestions for modeling the vibration process are provided.

Keywords copula, time reversibility, long-range dependence, state dependence, stationary process, Ornstein-Uhlenbeck process, diffusion

MSC(2010) 60G10, 62M10

doi: 10.1360/N012018-00043