



M - 结构: 模型不确定性下的一种具有对称独立性的非线性概率结构

献给林正炎教授 85 寿辰

李亦凡*, 于浩, Reg Kulperger

Department of Statistical and Actuarial Sciences, University of Western Ontario, London, ON N6A 5B7, Canada

E-mail: yifan.li@uwo.ca, hyu@stats.uwo.ca, rjk@stats.uwo.ca

收稿日期: 2024-10-26; 接受日期: 2025-04-15; 网络出版日期: 2025-08-29; * 通信作者

摘要 模型不确定性问题是在诸多领域中长期存在的一类基本问题. 非线性期望框架为模型不确定性的讨论提供了一个深刻的理论视角, 并且对经典概率体系进行了非平凡的系统性的扩展. 然而, 在讨论与实际问题的关联时, 非线性期望下的很多基本概念的含义需要仔细斟酌, 如正态分布和独立性. 提出 M - 结构的初衷是想设计一个具备统计意义上的灵活性的结构, 从而能够搭建一个桥梁, 既为非线性期望框架本身提供了一个适合更加广泛的读者群体在理解过程中的切入点, 同时也为模型不确定性问题的讨论提供一个新的视角. 这种结构的灵活性可以从以下角度进行阐述. 从独立性来说, M - 独立性保留了对称性, 能够刻画具有时空对称性的模型不确定性, 并且也与 G - 独立性有关联, 而后者主要刻画具备时间不对称性的模型不确定性. 进而得到 M - 结构下的一个中心极限定理, 描述了具备时空对称性的模型不确定性下的渐近分布, 即 M - 正态分布. 而 M - 正态分布保留了在独立性下一元和多元分布的直接联系, 这是 G - 正态分布所不具备的. 更重要的是, M - 结构揭示了经典与非线性期望意义下的概念之间的隐含联系及差异的来源, 这为后续的理论 and 实践层面的发展带来了一个意义深远的优势.

关键词 非线性数学期望 模型不确定性 中心极限定理 独立性 正态分布 对称性 可交换性

MSC (2020) 主题分类 60A05, 60E05, 60E07, 60F05, 62H05

1 引言

本文介绍我们近年来提出的 M - 结构 (M -structure)¹⁾ 的基本理论框架. 关于 M - 结构的进一步理论细节及应用方面的讨论, 读者可以参见文献 [19]²⁾ 和 [20].

1) M - 结构的技术名称为半 G - 结构 (semi- G -structure), 其中, M 代表最大 (maximum) 或非线性期望框架中的最大分布 (maximal distribution).

2) 文献 [19, 第 2.1 小节] 从状态空间波动率模型的不确定性角度介绍了 M - 结构, 并通过一个思想实验 (第 2.3 小节) 阐释了 M - 结构的内涵与角色.

英文引用格式: Li Y F, Yu H, Kulperger R. M -structure: A nonlinear probabilistic structure with symmetric independence under model uncertainty (in Chinese). Sci Sin Math, 2026, 56: 1–28, doi: [10.1360/SSM-2024-0328](https://doi.org/10.1360/SSM-2024-0328)

经典的概率体系擅长用一个单一的概率模型 $\mathbb{P}_\theta$ (模型本身可能很复杂) 来刻画随机现象. 然而, 在实践中, 由于客观信息的缺乏、建模过程的主观性或主观认知的模糊性 (ambiguity), 经常很难精确地决定出一个唯一固定的概率模型来刻画随机现象, 在这种情形下, 我们不能忽略关于概率模型本身的不确定性. 这种不确定性被称为模型不确定性 (model uncertainty). 模型不确定性是在多个领域中长期存在的一类基本问题, 例如, 这种不确定性在经济学中也被称为 Knight 不确定性^[17], 在统计学中也被称为认知不确定性 (epistemic uncertainty)^[16] 并作为一个核心议题被深入讨论^[5], 以及在稳健型风险度量领域 (robust risk measures) 中作为一个基础概念^[3,11] 等. 从非线性期望角度来对待模型不确定性问题有很长的历史, 详见综述文章^[23]. 非线性期望理论发展到目前阶段, 已经展现了一种能够在底层的公理化体系层面, 对模型不确定性进行深入和系统性探讨的潜能, 并且能够在动态意义下处理一般的模型不确定性集合. 对于模型不确定性问题和非线性期望视角的发展, 现在已经到了一个关键的节点: 需要仔细思考这个系统性的框架和其他领域之间的接口问题, 此时需要更多的灵活性, 甚至跳出已有的框架进行思考. 并且应深入理解清楚非线性期望框架下的底层概念的直觉, 而这里的底层概念包括其中的正态分布和独立性 (即 G - 正态分布和 G - 独立性).

然而, 非线性期望下的 G - 正态分布和经典正态分布在很多层面上都是有本质区别的. 一个显然的概念区别是, G - 正态分布描述的是分布不确定性, 而经典正态分布是确定性的分布. 这里主要关注的是一个基本的理论性质的差别: 从一元到多元的联系. 多个独立的一元经典正态分布作成随机向量仍服从多元正态分布. 这看似是一个很自然的性质, 但是这个性质对于 G - 正态分布不成立 (参见文献^[1]): 多个 G - 独立的一元 G - 正态分布组成的随机向量并不服从多元 G - 正态分布. 因为这个特殊性质, G -Brown 运动不再是一个 Gauss 过程. 实际上, G -Gauss 过程和 G -Brown 运动是两个概念 (参见文献^[14,25]). 虽然这是讨论模型不确定性问题时一种技术上的权衡, 但此区别对后续多维问题的讨论 (如多元统计和随机场理论) 还是会带来一些理解和技术上的困难. 我们应该如何在模型不确定性意义下, 更好地去建立起一元和多元分布之间的关联呢? 本文提出一种新的具有方差不确定性的正态分布, 称作 M - 正态分布, 它在 M - 独立性下仍然满足一元和多元的直接联系, 这也是我们第一次在次线性期望框架内搭建起这种关联. M - 正态分布同时与经典正态分布和 G - 正态分布都有着紧密联系, 这为这个图景构建起了一个桥梁.

在给出了 M - 正态分布和 M - 独立性后, 我们进而证明了 M - 中心极限定理 (定理 5.1), 揭示了 M - 正态分布在模型不确定性意义下具有对称独立性的场景中的中心角色, 我们将这一系列结果统称为 M - 结构. 我们提出 M - 结构的一个愿景, 就是在此基础上, 进一步揭示经典概率与非线性期望框架之间的区别和联系, 从而更好地将实践中已有的模型不确定性的问题和非线性期望框架下 (及其相关理论) 进行关联, 进而为两个方面都带来一种新的视角.

本文余下内容的结构如下. 第 2 节首先介绍非线性期望框架的基本概念与设置³⁾, 为后续讨论 M - 结构进行铺垫. 第 3 节给出 M - 正态分布的定义, 讨论其在一维和多维情形下的基本性质, 并以此为基础引入一种新的对称独立性, 即 M - 独立性. 第 4 节系统地给出一般的 M - 分布族及 M - 独立性的定义. 在此基础上, 第 5 节进一步证明 M - 中心极限定理 (M -central limit theorem, M -CLT), 揭示了 M - 正态分布在具有对称独立性的模型不确定性场景下的核心地位. 第 6 节总结全文, 并讨论 M - 结构未来的发展方向及其与实际应用层面的联系. 此外, 附录给出详细的证明和其他相关理论的细节.

3) 具体而言, 这里以简化方式介绍次线性期望空间的核心内容, 采用的是从一族概率测度出发的视角, 更多细节参见文献^[19, 第 1.6 小节] 或^[24, 第 6 章]; 另一种等价方式是从次线性泛函出发的视角, 参见文献^[24, 第 1 和 2 章].

2 预备知识

考虑在可测空间 $(\Omega, \mathcal{F})$ 上的一族概率测度 $\mathcal{P}$ ⁴⁾. 设 $L^0(\Omega)$ 为所有随机变量 $X: \Omega \rightarrow \mathbb{R}$ 所构成的空间. 令 $E_{\mathbb{P}}$ 代表在 $\mathbb{P} \in \mathcal{P}$ 下的线性期望. 定义 $\mathcal{H} := \{X \in L^0(\Omega) : \sup_{\mathbb{P} \in \mathcal{P}} E_{\mathbb{P}}[|X|] < \infty\}$, 可验证 $\mathcal{H}$ 是一个线性空间. 对任意 $X \in \mathcal{H}$, 定义它的次线性期望为

$$\hat{\mathbb{E}}[X] := \sup_{\mathbb{P} \in \mathcal{P}} E_{\mathbb{P}}[X].$$

之所以称 $\hat{\mathbb{E}}$ 为次线性期望, 是因为它可以被视为一个 $\mathcal{H}$ 上的一个次线性泛函, 即它满足如下性质:

- (i) 单调性 (若 $X \geq Y$, 则 $\hat{\mathbb{E}}[X] \geq \hat{\mathbb{E}}[Y]$);
- (ii) 保常性 (对于任意实数 c , 有 $\hat{\mathbb{E}}[c] = c$);
- (iii) 次可加性 ($\hat{\mathbb{E}}[X + Y] \leq \hat{\mathbb{E}}[X] + \hat{\mathbb{E}}[Y]$);
- (iv) 正齐次性 (对于任意 $\lambda \geq 0$, 有 $\hat{\mathbb{E}}[\lambda X] = \lambda \hat{\mathbb{E}}[X]$).

进而称 $(\Omega, \mathcal{H}, \hat{\mathbb{E}})$ 是一个次线性期望空间.

令 $\mathcal{H}^d := \{(X_1, X_2, \dots, X_d) : X_i \in \mathcal{H}, i = 1, 2, \dots, d\}$ 为随机向量构成的空间. 对于任意随机向量 $X \in \mathcal{H}^d$, 后面将经常提到 X 的一个变换 $\varphi(X)$, 其中 $\varphi: (\mathbb{R}^d, |\cdot|) \rightarrow (\mathbb{R}, |\cdot|)$ 可以视为一个测试函数. 令 $\mathbb{N}_+$ 表示所有正整数的集合. 考虑以下两个函数空间:

- $C_{b, \text{Lip}}(\mathbb{R}^d)$: 所有有界且 Lipschitz 连续的函数所构成的线性空间;
- $C_{l, \text{Lip}}(\mathbb{R}^d)$: 所有满足如下局部 Lipschitz 条件的函数所构成的线性空间:

$$|\varphi(x) - \varphi(y)| \leq C_{\varphi}(1 + |x|^k + |y|^k)|x - y|, \quad \text{对于 } x, y \in \mathbb{R}^d, \text{ 其中 } C_{\varphi} > 0, k \in \mathbb{N}_+ \text{ 取决于 } \varphi.$$

如果 φ 的定义域的维数在上下文中清楚, 将简记为 $\varphi \in C_{b, \text{Lip}}$ 或 $\varphi \in C_{l, \text{Lip}}$.

注意到 $\mathcal{H}$ 满足, 对于任意 $\varphi \in C_{b, \text{Lip}}$, 若 $X \in \mathcal{H}^d$, 则 $\varphi(X) \in \mathcal{H}$. 然而, 这一属性未必适用于任意 $\varphi \in C_{l, \text{Lip}}$. 因此, 当我们在此框架中讨论分布和独立性的定义时, 将默认考虑 $\varphi \in C_{b, \text{Lip}}$. 在后面将看到, 对于特定的分布 (如最大分布和 M -正态分布), 由于它们的特殊性质, 使得 $\varphi(X) \in \mathcal{H}$ 对更大范围的 φ 仍成立, 从而这个函数空间可以扩展到 $C_{l, \text{Lip}}$.

将一个随机向量 $X \in \mathcal{H}^d$ 在某个的概率测度 $\mathbb{P} \in \mathcal{P}$ 下的分布记作 $\mathbb{P}_X := \mathbb{P} \circ X^{-1}$. 由于 $\mathcal{P}$ 描述了概率本身的不确定性, 所以集合

$$\mathcal{P}_X := \{\mathbb{P}_X : \mathbb{P} \in \mathcal{P}\}$$

刻画了 X 的分布不确定性 (distributional uncertainty). 当 $\mathcal{P}_X$ 是一个单点集, 即 X 没有分布不确定性时, 称 X 的分布是确定的 (certain) 或者经典的 (classical).

定义 2.1 以下是次线性期望 $\hat{\mathbb{E}}$ 下的与分布相关的两个概念:

- (1) 称 X 和 Y 是同分布的, 记作 $X \stackrel{d}{=} Y$, 如果对于任意 $\varphi \in C_{b, \text{Lip}}$, 有

$$\hat{\mathbb{E}}[\varphi(X)] = \hat{\mathbb{E}}[\varphi(Y)];$$

- (2) 一个序列 $\{X_n\}_{n=1}^{\infty}$ 依分布收敛于 X , 记作 $X_n \xrightarrow{d} X$, 如果对于任意 $\varphi \in C_{b, \text{Lip}}$, 有

$$\lim_{n \rightarrow \infty} \hat{\mathbb{E}}[\varphi(X_n)] = \hat{\mathbb{E}}[\varphi(X)].$$

4) 这里暂未对 $\mathcal{P}$ 作特殊假设, 后续为确保次线性期望具有正则性, 需要对 $\mathcal{P}$ 引入额外的性质, 如弱紧性, 详见文献 [10, 定理 12]. 同时, 存在弱紧的族 $\mathcal{P}$, 使得典型的分布 (最大分布和 G -正态分布) 和 G -Brown 运动得以存在, 详见文献 [24, 第 6.2 小节].

由于次线性期望框架也常被称为 G - 期望框架⁵⁾, 为与后续提出的 M - 结构中的相关概念明确区分, 本文中次线性期望框架下的对应概念统一添加前缀 G .

定义 2.2 (G - 独立性) 称一个随机向量 Y 是 G - 独立于 X , 记作 $X \dashrightarrow Y$, 如果对于任意 $\varphi \in C_{b,\text{Lip}}$, 有

$$\hat{\mathbb{E}}[\varphi(X, Y)] = \hat{\mathbb{E}}[\hat{\mathbb{E}}[\varphi(x, Y)]_{x=X}].$$

关于 G - 独立性的一个重要事实是, 它是非对称的: $X \dashrightarrow Y$ (Y 独立于 X) 不一定意味着 $Y \dashrightarrow X$ (X 独立于 Y), 详见例 2.1. 这就是为什么 G - 独立性也被称为序列独立性 (sequential independence), 并且使用符号 $\dashrightarrow$ 来代表两者之间独立性的顺序关系.

例 2.1 (参见文献 [24, 例 1.3.15]) 考虑两个同分布的 $X, Y \in \mathcal{H}$, 满足 $\hat{\mathbb{E}}[-X] = \hat{\mathbb{E}}[X] = 0$ 和 $\bar{\sigma}^2 = \hat{\mathbb{E}}[X^2] > -\hat{\mathbb{E}}[-X^2] = \underline{\sigma}^2$. 同时假设 $\hat{\mathbb{E}}[|X|] > 0$ 使得 $\hat{\mathbb{E}}[X^+] = \frac{1}{2}\hat{\mathbb{E}}[|X| + X] = \frac{1}{2}\hat{\mathbb{E}}[|X|] > 0$, 则有

$$\hat{\mathbb{E}}[XY^2] = \begin{cases} (\bar{\sigma}^2 - \underline{\sigma}^2)\hat{\mathbb{E}}[X^+], & \text{如果 } X \dashrightarrow Y, \\ 0, & \text{如果 } Y \dashrightarrow X. \end{cases}$$

称 $\bar{X}$ 是 X 的一个独立拷贝, 如果 $\bar{X} \stackrel{d}{=} X$ 且 $X \dashrightarrow \bar{X}$. 称一个序列 $\{X_i\}_{i=1}^n$ 是 G - 独立或序列独立的, 如果对于 $i = 1, 2, \dots, n-1$, 有

$$(X_1, X_2, \dots, X_i) \dashrightarrow X_{i+1}.$$

记作

$$X_1 \dashrightarrow X_2 \dashrightarrow \dots \dashrightarrow X_n. \quad (2.1)$$

如果 $\{X_i\}_{i=1}^n$ 进而满足同分布, 即 $X_{i+1} \stackrel{d}{=} X_i$ 对任意 $i = 1, 2, \dots, n-1$ 都成立, 则称 $\{X_i\}_{i=1}^n$ 满足 G - 独立同分布, 简称 G - 独立同分布 (G -independent and identically distributed, i.i.d.).

2.1 最大分布

定义 2.3 (一元最大分布) 一个随机变量 V 服从最大分布 $\mathcal{M}[\underline{\sigma}, \bar{\sigma}]$, 其中 $\underline{\sigma} \leq \bar{\sigma}$, 如果对于任意的 $\varphi \in C_{1,\text{Lip}}(\mathbb{R})$, 有

$$\hat{\mathbb{E}}[\varphi(V)] = \max_{v \in [\underline{\sigma}, \bar{\sigma}]} \varphi(v).$$

注 2.1 这里可以考虑任意 $\varphi \in C_{1,\text{Lip}}$ (而不仅是 $\varphi \in C_{b,\text{Lip}}$) 的原因是, $\hat{\mathbb{E}}[\varphi(V)] < \infty$ 对任意 $\varphi \in C_{1,\text{Lip}}$ 都成立, 所以 $\varphi(V) \in \mathcal{H}$.

注 2.2 最大分布的分布不确定性可以被进一步表示为

$$\hat{\mathbb{E}}[\varphi(V)] = \sup_{\sigma \in \mathcal{A}[\underline{\sigma}, \bar{\sigma}]} \mathbb{E}[\varphi(\sigma)], \quad (2.2)$$

其中 $\mathcal{A}[\underline{\sigma}, \bar{\sigma}]$ 是那些在区间 $[\underline{\sigma}, \bar{\sigma}]$ 服从所有 (离散, 绝对连续或任意) 分布的随机变量作成的集合.

⁵⁾ 具体而言, 是在引入 G -Brown 运动之后, 对应的次线性期望会被称为 G - 期望 (参见文献 [24, 第 3.1 小节]), 其中 G 是指背后的偏微分方程中的一个生成函数, 当然也可以从概念上理解为 general (一般) 的首字母.

定义 2.4 (多元最大分布) 一个随机向量 $V: \Omega \rightarrow \mathbb{R}^d$ 服从多元最大分布 $\mathcal{M}(\mathcal{V})$, 如果存在一个紧且凸的子集 $\mathcal{V} \subset \mathbb{R}^d$ 满足: 对于任意 $\varphi \in C_{1,\text{Lip}}(\mathbb{R}^d)$, 有

$$\hat{\mathbb{E}}[\varphi(V)] = \max_{\sigma \in \mathcal{V}} \varphi(\sigma).$$

注 2.3 多元最大分布也具备类似于注 2.2 的表示.

如例 2.1 所述, 分布之间的 G - 独立性一般具有非对称性. 当然, 在退化情形下, 即 $\mathcal{P}$ 退化为单点集 (等价于 $\hat{\mathbb{E}}$ 退化为线性期望) 时, X 和 Y 的分布变为经典分布, G - 独立性也退化为经典的独立性, 自然就出现了对称性. 在非退化情形中, 如果 X 和 Y 服从最大分布, G - 独立性也会出现对称性. 这里提供一个简短论证, 考虑 $X \sim \mathcal{M}(A)$, $Y \sim \mathcal{M}(B)$, 其中 A, B 是紧且凸的集合. 如果 $X \dashrightarrow Y$, 根据 G - 独立性定义, 有

$$\begin{aligned} \hat{\mathbb{E}}[\varphi(X, Y)] &= \hat{\mathbb{E}}[\hat{\mathbb{E}}[\varphi(x, Y)]_{x=X}] = \hat{\mathbb{E}}[\max_{y \in B} \varphi(x, y)]_{x=X} \\ &= \max_{x \in A} \max_{y \in B} \varphi(x, y) = \max_{(x, y) \in A \times B} \varphi(x, y) \\ &= \max_{x \in B} \max_{y \in A} \varphi(x, y) = \hat{\mathbb{E}}[\hat{\mathbb{E}}[\varphi(X, y)]_{y=Y}]. \end{aligned}$$

从而有 $Y \dashrightarrow X$. 这个性质当然对 n 个最大分布之间也成立. 文献 [13, 定理 15] 进一步证明了, 对称的 G - 独立性仅在经典分布和最大分布这两种情形中出现.

定理 2.1 [13] 对于两个非常数随机变量 $X, Y \in \mathcal{H}$, 如果 X 和 Y 是相互独立的 ($X \dashrightarrow Y$ 且 $Y \dashrightarrow X$), 则它们属于以下两种情形之一:

- (1) X 和 Y 的分布是经典的 (没有分布不确定性);
- (2) X 和 Y 都服从最大分布.

如果在一个对称场景下, 将经典分布和最大分布结合在一起, 出来的分布之间的独立性是什么样的呢? (是否只能是非对称的 G - 独立性?) 此现象引发的思考将在第 3.3 小节进行细致讨论.

2.2 G - 正态分布和 G - 中心极限定理

定义 2.5 (G - 正态分布) 一个 d - 维随机向量 X 服从 G - 正态分布, 如果对于任意独立拷贝 $\bar{X}$, 都有

$$X + \bar{X} \stackrel{d}{=} \sqrt{2}X.$$

注 2.4 当 $d = 1$ 时, 将 X 的分布记作 $\mathcal{N}_G(0, [\underline{\sigma}^2, \bar{\sigma}^2])$, 其中 $\underline{\sigma}^2 := -\hat{\mathbb{E}}[-X^2]$ 和 $\bar{\sigma}^2 := \hat{\mathbb{E}}[X^2]$.

命题 2.1 (G - 正态分布通过 G - 热方程的特征化) 一个 d - 维随机向量 X 服从 G - 正态分布, 当且仅当

$$v(t, x) := \hat{\mathbb{E}}[\varphi(x + \sqrt{t}X)]$$

是以下 G - 热方程在 $(t, x) \in [0, 1] \times \mathbb{R}^d$ 上定义的唯一 (黏性) 解:

$$v_t - G(D_x^2 v) = 0, \quad v|_{t=0} = \varphi,$$

其中 $G(A) := \frac{1}{2}\hat{\mathbb{E}}[\langle AX, X \rangle]: \mathbb{S}_d \rightarrow \mathbb{R}$ 是一个描述 X 分布的次线性函数.

注 2.5 对于 $d = 1$ 的情形, 有 $G(a) = \frac{1}{2}(\bar{\sigma}^2 a^+ - \underline{\sigma}^2 a^-)$, 当 $\underline{\sigma}^2 > 0$ 时, 该方程也称为带有波动率不确定性的 Black-Scholes-Barenblatt 方程. 当 $\underline{\sigma} = \bar{\sigma} = \sigma$ 时, 因为 G - 热方程退化为经典热方程, 所以此时 G - 正态分布 $\mathcal{N}_G(0, [\underline{\sigma}^2, \bar{\sigma}^2])$ 等价于经典正态 $N(0, \sigma^2)$.

定理 2.2 (G - 中心极限定理) 设 $\{X_i\}_{i=1}^\infty$ 为一列 G - 独立且同分布的随机变量, 满足均值确定性 $\hat{\mathbb{E}}[X_1] = -\hat{\mathbb{E}}[-X_1] = 0$ 和方差不确定性

$$\underline{\sigma}^2 = -\hat{\mathbb{E}}[-X_1^2] \leq \hat{\mathbb{E}}[X_1^2] = \bar{\sigma}^2,$$

并假设

$$\lim_{\lambda \rightarrow \infty} \hat{\mathbb{E}}[(|X_1|^2 - \lambda)^+] = 0,$$

则当 $n \rightarrow \infty$ 时, 有

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \xrightarrow{d} W^G,$$

其中 $W^G \sim \mathcal{N}_G(0, [\underline{\sigma}^2, \bar{\sigma}^2])$. 进一步地, 对任意 $\varphi \in C(\mathbb{R})$ 且满足线性增长条件 $|\varphi(x)| \leq C(1 + |x|)$, 有

$$\lim_{n \rightarrow \infty} \hat{\mathbb{E}} \left[\varphi \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \right) \right] = \hat{\mathbb{E}}[\varphi(W^G)].$$

2.3 符号和术语

本文中的经典概率空间特指 Kolmogorov 概率空间^[18]. 在经典概率空间 $(\Omega, \mathcal{F}, \mathbb{P})$ 中, 常见的随机变量包括 $\sigma : \Omega \rightarrow [\underline{\sigma}, \bar{\sigma}]$ 和 $\epsilon \sim N(0, 1)$. 在某些情形下, σ 也可表示 $[\underline{\sigma}, \bar{\sigma}]$ 上的一个常数. 记 $\mathbb{E}_{\mathbb{P}}$ 为对应于概率测度 $\mathbb{P}$ 的线性期望, 若上下文中 $\mathbb{P}$ 已明确, 可简记为 $\mathbb{E}$. 在次线性期望空间 $(\Omega, \mathcal{H}, \hat{\mathbb{E}})$ 中, 常见的随机变量包括 $V \sim \mathcal{M}[\underline{\sigma}, \bar{\sigma}]$, $\epsilon \sim \mathcal{N}_G(0, [1, 1])$, $W \sim \mathcal{N}_M(0, [\underline{\sigma}^2, \bar{\sigma}^2])$ 和 $W^G \sim \mathcal{N}_G(0, [\underline{\sigma}^2, \bar{\sigma}^2])$, 注意 $\mathcal{N}_G(0, [1, 1])$ 可被视为经典分布 $N(0, 1)$ (见注 2.5), 为表述方便, 仍使用同样的 ϵ 作为相应随机变量的符号. 同理, 在多元情形下, 经典的标准多元正态分布 $N(0, I_d)$ 也可被视为多元 G - 正态分布的退化版本.

我们在两个空间中使用两套不同的随机变量符号主要是为了讨论的清晰和简便⁶⁾, 尤其是在涉及一个非线性期望概念的表示或者背后的分布不确定性时. 除非特别声明, 本文后续的结论默认均在次线性期望空间中.

3 M - 正态分布

在介绍一般的 M - 分布和独立性之前, 本节首先讨论具有代表性的 M - 正态分布及其相关的独立性, 以此阐明 M - 结构的核心思想.

3.1 一维情形

定义 3.1 (M - 正态分布) 在次线性期望空间中, 一个随机变量 W 服从 M - 正态分布, 记作 $W \sim \mathcal{N}_M(0, [\underline{\sigma}^2, \bar{\sigma}^2])$, 如果对于任意 $\varphi \in C_{1,\text{Lip}}(\mathbb{R})$, 有

$$\hat{\mathbb{E}}[\varphi(W)] = \max_{\sigma \in [\underline{\sigma}, \bar{\sigma}]} \mathbb{E}[\varphi(\sigma\epsilon)], \quad (3.1)$$

其中 $\epsilon \sim N(0, 1)$ 且 σ 是一个常数.

6) 由于两者的理解有一些差异, 这样恰恰可以避免可能的困惑.

注 3.1 有两种方式来描述 M -正态分布的分布不确定性. 根据定义 3.1, 其可以表示为一族具有不同方差的经典正态分布. 另一种描述是将其视为一族 Gauss 混合分布, 其中 σ 的先验分布可以在 $[\underline{\sigma}, \bar{\sigma}]$ 上取任意形式 (如离散或绝对连续的分布)⁷⁾. 例如, 令 $\mathcal{A}[\underline{\sigma}, \bar{\sigma}]$ 代表 $[\underline{\sigma}, \bar{\sigma}]$ 上的所有服从绝对连续分布的随机变量, 则 (3.1) 可以转换为

$$\hat{\mathbb{E}}[\varphi(W)] = \sup_{\sigma \in \mathcal{A}[\underline{\sigma}, \bar{\sigma}]} \mathbb{E}[\varphi(\sigma\epsilon)]. \quad (3.2)$$

在实践中, 可以根据需要来选择 $\mathcal{A}[\underline{\sigma}, \bar{\sigma}]$ 的描述形式⁸⁾.

注 3.2 在此对比 M -正态分布与 G -正态分布的次线性期望. 设 $W^G \sim \mathcal{N}_G(0, [\underline{\sigma}^2, \bar{\sigma}^2])$ 和 $W \sim \mathcal{N}_M(0, [\underline{\sigma}^2, \bar{\sigma}^2])$, 对于任意 $\varphi \in C_{1,\text{Lip}}(\mathbb{R})$, 有

$$\hat{\mathbb{E}}[\varphi(W^G)] \geq \max_{v \in [\underline{\sigma}, \bar{\sigma}]} \mathbb{E}[\varphi(v\epsilon)] = \hat{\mathbb{E}}[\varphi(W)]. \quad (3.3)$$

这可以通过将抛物型偏微分方程的比较定理 (参见文献 [9]) 应用于与 G -热方程和经典热方程相关的初始条件 φ 来证明. 当 φ 是非凸且非凹的函数时 (如 $\varphi(x) = x^3$), 不等号严格成立. 详见文献 [12, 引理 2]. 实际上, 与 M -正态分布对应的不确定性集合 (uncertainty set) 是 G -正态分布的一个真子集 (文献 [10] 给出了 G -正态分布的不确定性集合的一个表示).

以下命题 3.1 给出一个例子, 展示了 M -正态分布在次线性期望空间中的存在性.

命题 3.1 考虑一个服从最大分布的随机变量 $V \sim \mathcal{M}[\underline{\sigma}, \bar{\sigma}]$ 和一个服从经典正态分布的随机变量 $\epsilon \sim N(0, 1)$, 使得 ϵ 独立于 V , 记作 $V \dashrightarrow \epsilon$. 令

$$W := V\epsilon, \quad (3.4)$$

则有 $W \sim \mathcal{N}_M(0, [\underline{\sigma}^2, \bar{\sigma}^2])$.

注 3.3 满足上述条件的 V 和 ϵ 的存在性由两个独立的 G -分布随机向量的存在性保证⁹⁾:

$$(V, \eta) \dashrightarrow (K, \epsilon),$$

其中 $V \stackrel{d}{=} K \stackrel{d}{=} \mathcal{M}[\underline{\sigma}, \bar{\sigma}]$ 且 $\eta \stackrel{d}{=} \epsilon \stackrel{d}{=} \mathcal{N}_G(0, [1, 1])$. 通过 $\varphi_1(x, y) = y$ 和 $\varphi_2(x, y) = x$, 直接有

$$\varphi_1(V, \eta) = V \dashrightarrow \varphi_2(K, \epsilon) = \epsilon.$$

我们将主要关注 (3.4) 形式的 M -正态分布. 称这种类型的 M -正态分布为可分解的 M -正态分布. 这里给出正式定义.

定义 3.2 (可分解的 M -正态分布) 假设 $W \sim \mathcal{N}_M(0, [\underline{\sigma}^2, \bar{\sigma}^2])$, 如果存在一个最大分布随机变量 $V \sim \mathcal{M}[\underline{\sigma}, \bar{\sigma}]$ 和一个服从经典正态分布的随机变量 $\epsilon \sim N(0, 1)$, 并且 ϵ 独立于 V (记为 $V \dashrightarrow \epsilon$), 使得

$$W = V\epsilon,$$

则称 W 是可分解的.

7) 更多细节可参见文献 [19, 定理 2.1.11].

8) 例如, 考虑 $\sigma = \underline{\sigma} + (\bar{\sigma} - \underline{\sigma})\xi$, 其中 ξ 的分布可以是具备任意参数 $\alpha, \beta > 0$ 的 $\text{Beta}(\alpha, \beta)$. 这样一个分布族的期望上确界也会与 M -正态分布的期望相等.

9) 参见文献 [24, 第 2.3 小节], 通过构造一个乘积空间.

注 3.4 在定义 3.2 中, 需注意 ϵ 与 V 之间的独立性的方向不可反转. 有如下两个原因:

(1) 如果反转方向, 导出的次线性期望下的分布或不确定性集合将发生本质的改变. 具体而言, 若 V 独立于 ϵ (记为 $\epsilon \dashrightarrow V$), 考虑 $\tilde{W} := V\epsilon$, 有

$$\begin{aligned}\hat{\mathbb{E}}[\tilde{W}] &= \hat{\mathbb{E}}[\hat{\mathbb{E}}[Vx]_{x=\epsilon}] = \mathbb{E}[\bar{\sigma}\epsilon^+ - \underline{\sigma}\epsilon^-] \\ &= \mathbb{E}[\bar{\sigma}\epsilon + (\bar{\sigma} - \underline{\sigma})\epsilon^-] = (\bar{\sigma} - \underline{\sigma})\mathbb{E}[\epsilon^-] \\ &= \frac{1}{2}(\bar{\sigma} - \underline{\sigma})\mathbb{E}[|\epsilon|] > 0,\end{aligned}$$

其中 $\epsilon^+ = \max\{\epsilon, 0\}$ 且 $\epsilon^- = \max\{-\epsilon, 0\}$. 同样,

$$-\hat{\mathbb{E}}[-\tilde{W}] = -\frac{1}{2}(\bar{\sigma} - \underline{\sigma})\mathbb{E}[|\epsilon|] < 0.$$

可以看到 $\tilde{W}$ 和 W 在均值意义上展现出了它们的不同: W 具有确定的零均值, 但 $\tilde{W}$ 具有均值不确定性.

(2) 由于 V 服从最大分布, ϵ 服从非退化的经典分布, 如果它们相互独立, 这将与定理 2.1 矛盾. 因此, 没有 V 与 ϵ 之间的相互独立性.

注 3.5 一个比较有趣的问题是, 是否存在不可分解的 M - 正态分布? 目前我们还没有完整的答案, 值得未来进一步探索.

接下来的命题 3.2 给出了 M - 正态分布与 G - 正态分布之间的一种特殊联系¹⁰⁾.

命题 3.2 设 $W^G \sim \mathcal{N}_G(0, [\underline{\sigma}^2, \bar{\sigma}^2])$ 且 $W \sim \mathcal{N}_M(0, [\underline{\sigma}^2, \bar{\sigma}^2])$. 对于 $\varphi \in C_{1,\text{Lip}}(\mathbb{R})$, 当 φ 是凸函数或凹函数时, 有

$$\hat{\mathbb{E}}[\varphi(W^G)] = \hat{\mathbb{E}}[\varphi(W)] = \begin{cases} \mathbb{E}_{\mathbb{P}}[\varphi(N(0, \bar{\sigma}^2))], & \varphi \text{ 是凸函数,} \\ \mathbb{E}_{\mathbb{P}}[\varphi(N(0, \underline{\sigma}^2))], & \varphi \text{ 是凹函数.} \end{cases}$$

3.2 多维情形

一元 M - 正态分布的定义可以自然扩展到多维情形. 直观来讲, 多元 M - 正态分布可以被视为经典多元正态分布的类比, 而后者可写为

$$N(0, \Sigma) = \Sigma^{1/2}N(0, I_d), \quad (3.5)$$

其中 I_d 是一个 $d \times d$ 的单位矩阵, Σ 是协方差矩阵且它有一个对称平方根矩阵 $\Sigma^{1/2}$.

这里看起来是从一元到多元的常规扩展. 然而, 其实多元 M - 正态分布带来的好处远不止定义本身. 在后续讨论中, 我们可以直接使用两个 M - 独立的 (将在第 3.3 小节中定义) M - 正态分布形成一个二元 M - 正态分布. 这似乎是正态分布的一个自然属性 (就像经典正态分布所具有的), 但对于现有的 G - 正态分布却不成立 (具体参见文献 [1, 引理 4.3]). 从多元统计和随机场等领域的讨论视角来看, M - 正态分布的这个性质会是一个优势, 能带来很多便利.

定义 3.3 (多元 M - 正态分布) 考虑 $\mathcal{C}$ 为一个由协方差矩阵组成的集合. 称一个 d - 维随机向量 W 服从 M - 正态分布 $\mathcal{N}_M(0, \mathcal{C})$, 如果对于任意 $\varphi \in C_{1,\text{Lip}}(\mathbb{R}^d)$, 有

$$\hat{\mathbb{E}}[\varphi(W)] = \sup_{\Sigma \in \mathcal{C}} \mathbb{E}[\varphi(\Sigma^{1/2}\epsilon)],$$

10) 该结果可直接由 G - 正态分布的性质 (参见文献 [24, 命题 2.2.15]) 推导得到, 也可从 M - 正态分布的期望出发, 考虑对应的优化问题 (参见文献 [19, 第 103 页]).

其中 $\epsilon \sim N(0, I_d)$.

注 3.6 与注 3.1 类似, 多元 M -正态分布的分布不确定性既可以视为一族具有协方差不确定性的经典多元正态分布, 也可以描述为一族多元的 Gauss 混合分布, 其中的随机协方差矩阵 Σ 可以服从 $\mathcal{C}$ 上的各种形式的分布 (如离散或绝对连续的分布).

定义 3.4 (可分解的多元 M -正态分布) 假设 $W \sim \mathcal{N}_M(0, \mathcal{C})$, 令 $\mathcal{V} := \{\Sigma^{1/2} : \Sigma \in \mathcal{C}\}$, 即对称平方根矩阵作成的集合. 如果存在随机向量 $\epsilon \sim N(0, I_d)$ 和随机矩阵 $V : \Omega \rightarrow \mathbb{R}^{d \times d}$ 服从 $\mathcal{M}(\mathcal{V})$, 且满足 ϵ 独立于 V (记为 $V \dashrightarrow \epsilon$), 使得 $W = V\epsilon$, 则称 W 是可分解的.

注 3.7 在定义 3.4 中, ϵ 与 V 之间的独立性的方向不可反转 (原因与注 3.4 类似). 矩阵值最大分布 $\mathcal{M}(\mathcal{V})$ 可以从定义 2.4 扩展¹¹⁾ 得到. 关系 $V \dashrightarrow \epsilon$ 是 G -独立性的多元版本.

从定义 3.3 出发, 可以方便地研究随机向量 $W = (W_1, W_2, \dots, W_d)$ 的分量之间的协方差不确定性. 首先, 可以定义 W 的分量之间的上下协方差为 (注意, W_i 具有确定的均值且为零)

$$\bar{\gamma}(i, j) := \hat{\mathbb{E}}[W_i W_j] \quad \text{和} \quad \underline{\gamma}(i, j) := -\hat{\mathbb{E}}[-W_i W_j].$$

这两个量与 $\mathcal{C}$ 密切相关, 如下所示.

命题 3.3 (多元 M -正态分布的协方差不确定性) 对于每个 $\Sigma \in \mathcal{C}$, 令 Σ_{ij} 表示 Σ 的第 (i, j) 项, 并定义

$$[\underline{\Sigma}_{ij}, \bar{\Sigma}_{ij}] := \left[\min_{\Sigma \in \mathcal{C}} \Sigma_{ij}, \max_{\Sigma \in \mathcal{C}} \Sigma_{ij} \right].$$

则 $\underline{\gamma}(i, j) = \underline{\Sigma}_{ij}$, $\bar{\gamma}(i, j) = \bar{\Sigma}_{ij}$. 特别地, 有

$$\bar{\sigma}_i^2 := \hat{\mathbb{E}}[W_i^2] = \bar{\Sigma}_{ii}, \quad \underline{\sigma}_i^2 := -\hat{\mathbb{E}}[-W_i^2] = \underline{\Sigma}_{ii}.$$

证明 对于每个 $(i, j) \in \{1, 2, \dots, d\}^2$, 设 $f_{ij}(W) = W_i W_j$. 显然, $f_{ij} \in C_{1,\text{Lip}}(\mathbb{R}^d)$. 对于每个 Σ , 设 $(Y_1, Y_2, \dots, Y_d) := \Sigma^{1/2}\epsilon$. 根据定义 3.3, 可得

$$\bar{\gamma}(i, j) = \hat{\mathbb{E}}[W_i W_j] = \max_{\Sigma^{1/2} \in \mathcal{V}} \mathbb{E}[f_{ij}(\Sigma^{1/2}\epsilon)] = \max_{\Sigma^{1/2} \in \mathcal{V}} \mathbb{E}[Y_i Y_j] = \max_{\Sigma^{1/2} \in \mathcal{V}} \Sigma_{ij} = \bar{\Sigma}_{ij}.$$

类似可得 $\underline{\gamma}(i, j) = \underline{\Sigma}_{ij}$. □

这里仅给出了多元 M -正态分布的定义和部分性质 (从而服务于本文后面的讨论), 关于更加细致的多维乃至高维情形的讨论, 将在未来的其他工作进行展示.

3.3 一种新的独立性

本小节以服从 M -正态分布的随机变量为代表, 通过探讨它们之间独立性, 来带领读者从源头出发来重新发现一种新的独立性, 即后面将引入的 M -独立性. 这里以 M -正态分布为主体进行一个思想实验. 旨在简要阐述 M -独立性的源头和思想, 更多理论细节参见文献 [19, 第 2.1.5 和 2.1.6 小节].

考虑在两个不同空间位置上的互不干扰的噪声源 (为了更具体一些, 可以将这个噪声源想象为一个音叉, 上面有一个旋钮来调节声音的波动程度, 每次旋转旋钮, 再敲击音叉发出声音. 而我们只知道旋钮的范围, 而不知道旋钮的取值). 两个位置的噪声源分别服从 $N(0, \sigma_1^2)$ 和 $N(0, \sigma_2^2)$, 其中 σ_1 和 σ_2 都是落在 $[\underline{\sigma}, \bar{\sigma}]$ 上的未知参数, 从而它们的分布不确定性都可以视为 M -正态分布 $\mathcal{N}_M(0, [\underline{\sigma}^2, \bar{\sigma}^2])$, 为方便起见, 分别记为可分解的形式 $W_1 = V_1 \epsilon_1$ 和 $W_2 = V_2 \epsilon_2$, 其中 V_1 和 V_2 都服从最大分布 $\mathcal{M}[\underline{\sigma}, \bar{\sigma}]$,

¹¹⁾ 严格来讲, 为了符合最大分布的定义, 我们要求集合 $\mathcal{V}$ 是紧且凸的, 具体可以用相应的矩阵空间内的概念来定义.

ϵ_1 和 ϵ_2 都服从 $N(0, 1)$. 因为它们在空间上互不干扰, 则 V_1 和 V_2 以及 ϵ_1 和 ϵ_2 之间, 具有一定的“独立性”. 它们在空间上对称, 具体而言, 在空间上交换位置后, 如果只是观测两个位置的噪声, 则无法区分两者, 换言之, (W_1, W_2) 和 (W_2, W_1) 具有同样的分布不确定性¹²⁾.

则可以问, 在次线性期望意义下, W_1 和 W_2 之间是什么样的“独立性”关系呢? 这种具有空间对称性的关系并不是 G - 独立性, 因为对称的 G - 独立性只发生于两个最大分布或者两个经典分布之间 (定理 2.1). 即使内部的 V_1 和 V_2 , 以及 ϵ_1 和 ϵ_2 之间, 是满足 (对称) G - 独立性的, 但是 W_1 和 W_2 并不 G - 独立. 但是从设定上 W_1 和 W_2 之间并没有任何相依关系 (dependence). 而现实世界中也普遍存在这种具有空间对称性的随机变量, 它们所对应的分布不确定性多种多样 (并不限于最大分布), 这就要求我们引入一种新的独立性关系, 来刻画这种具有空间对称性的“独立性”.

首先, (W_1, W_2) 的分布不确定性可以写成

$$\{P_{(\sigma_1 \epsilon_1, \sigma_2 \epsilon_2)} : (\sigma_1, \sigma_2) \in [\underline{\sigma}, \bar{\sigma}]^2, \epsilon_1 \perp \epsilon_2, \epsilon_1 \stackrel{d}{=} \epsilon_2 \stackrel{d}{=} N(0, 1)\}.$$

进而 (W_1, W_2) 的次线性期望可以表示为, 对于任意 $\varphi \in C_{1,\text{Lip}}$, 有

$$\hat{\mathbb{E}}[\varphi(W_1, W_2)] = \sup_{(\sigma_1, \sigma_2) \in [\underline{\sigma}, \bar{\sigma}]^2} \mathbb{E}[\varphi(\sigma_1 \epsilon_1, \sigma_2 \epsilon_2)], \quad (3.6)$$

其中 $\epsilon_1 \perp \epsilon_2$. 从中也可以进一步看出 (σ_1, σ_2) 及 (ϵ_1, ϵ_2) 内部的对称性. 实际上 (3.6) 正是两个 M - 正态分布之间的 M - 独立性的定义 (将在第 4 节给出).

在 M - 独立性下, M - 正态分布保留了很多与经典正态分布类似的性质 (而且由于有 (3.6) 这样的表示, 证明也非常简洁直观), 例如, M - 正态分布是一个稳定分布 (命题 3.4).

命题 3.4 如果 W_1 和 W_2 都服从 $\mathcal{N}_M(0, [\underline{\sigma}^2, \bar{\sigma}^2])$, 且它们是 M - 独立的, 则对于任意 $a, b \geq 0$, 有

$$aW_1 + bW_2 \stackrel{d}{=} \sqrt{a^2 + b^2} W_1.$$

证明 对于任意 $\psi \in C_{1,\text{Lip}}$, 不难验证 $\varphi : (x_1, x_2) \mapsto \psi(ax_1 + bx_2)$ 仍属于 $C_{1,\text{Lip}}$, 进而由 (3.6) 可得

$$\hat{\mathbb{E}}[\psi(aW_1 + bW_2)] = \sup_{(\sigma_1, \sigma_2) \in [\underline{\sigma}, \bar{\sigma}]^2} \mathbb{E}[\psi(a\sigma_1 \epsilon_1 + b\sigma_2 \epsilon_2)] = \sup_{(\sigma_1, \sigma_2) \in [\underline{\sigma}, \bar{\sigma}]^2} \mathbb{E}[\psi(\sqrt{a^2 \sigma_1^2 + b^2 \sigma_2^2} \epsilon_1)].$$

注意到

$$\{\sqrt{a^2 \sigma_1^2 + b^2 \sigma_2^2} : (\sigma_1, \sigma_2) \in [\underline{\sigma}, \bar{\sigma}]^2\} = [\sqrt{a^2 + b^2} \underline{\sigma}, \sqrt{a^2 + b^2} \bar{\sigma}] = \{\sqrt{a^2 + b^2} \sigma : \sigma \in [\underline{\sigma}, \bar{\sigma}]\},$$

因此,

$$\hat{\mathbb{E}}[\psi(aW_1 + bW_2)] = \sup_{\sigma_1 \in [\underline{\sigma}, \bar{\sigma}]} \mathbb{E}[\psi(\sqrt{a^2 + b^2} \sigma_1 \epsilon_1)] = \hat{\mathbb{E}}[\psi(\sqrt{a^2 + b^2} W_1)].$$

最后, 由同分布的定义 (定义 2.1, 注意到 $C_{b,\text{Lip}} \subset C_{1,\text{Lip}}$) 可得结论成立. $\square$

更重要的是, 在 M - 独立性下, M - 正态分布仍然保留了一元和多元的紧密联系 (定理 3.1).

¹²⁾ 这里涉及一个更深层的概念, 称作可交换性 (exchangeability), σ_1 和 σ_2 可能取不同的值, 两个随机变量如果在经典意义下可能是不可交换的, 但是正因为对这两个参数有同样的不确定性且互不干扰 (我们并不知道像 $\sigma_1 > \sigma_2$ 这样的关系), 所以二者其实具备一种在分布不确定性意义下的可交换性.

定理 3.1 对于两个服从 M - 正态分布 $\mathcal{N}_M(0, [\underline{\sigma}^2, \bar{\sigma}^2])$ 的随机变量 W_1 和 W_2 , 它们是 M - 独立的, 当且仅当

$$(W_1, W_2) \sim \mathcal{N}_M(0, [\underline{\sigma}^2, \bar{\sigma}^2]I_2),$$

其中 I_2 是 2×2 的单位矩阵, $[\underline{\sigma}^2, \bar{\sigma}^2]I_2$ 代表集合

$$\{\text{diag}(\sigma_1^2, \sigma_2^2), \sigma_i \in [\underline{\sigma}, \bar{\sigma}], i = 1, 2\}.$$

证明 通过对比 (3.6) 和定义 3.4 后直接得到结论成立. $\square$

注 3.8 注意到定理 3.1 所描述的性质是 G - 正态分布不具备的^[1]: 两个 G - 独立的 G - 正态分布组成的随机向量并不服从二元 G - 正态分布¹³⁾, 二元 G - 正态分布 $\mathcal{N}_G(0, [\underline{\sigma}^2, \bar{\sigma}^2]I_2)$ 的分量之间并不满足 G - 独立.

注 3.9 定理 3.1 体现了 M - 正态分布的一个优势, 可以通过单变量对象构造多变量正态分布. 这也是现代概率论与统计学中的重要基础, 如在多元统计和随机场理论的讨论中. 同样, 这对模型不确定性下的空间白噪声的发展至关重要 (Brown 运动是 Gauss 过程).

这种独立性理论上 (在已有的次线性期望空间中) 是否存在呢? 答案是肯定的. 虽然 M - 独立性本身与 G - 独立性在对称性上有本质区别, 但是有趣的是, M - 独立性可以出现在满足特定顺序的 G - 独立序列 $(V_1, \epsilon_1, V_2, \epsilon_2)$ 当中.

为了引导读者思考这里的 G - 独立性的顺序, 我们尝试将此思想实验中噪声源的解释和理论相结合. 首先考虑一个服从 M - 正态分布的随机变量 W , 并可分解为 $W = V\epsilon$, 其内部有额外的 G - 独立性要求, 即 ϵ 独立于 V , 记为 $V \dashrightarrow \epsilon$. 这里“隐含”的 G - 独立性顺序可以解释为一种事件发生顺序, 这里将噪声源视为带有旋钮的音叉, V 可以视为旋钮, ϵ 对应于敲击“取样”的步骤: 先旋转旋钮, 决定了参数 V 的取值后, 再进行敲击抽样 (或者是随机噪声的生成), 所以出现了这种独立性顺序. 这里有两个音叉, 按照事件发生顺序, 应当是先旋转了两个旋钮, 再敲击两个音叉 (发出声音, 我们得到观测)¹⁴⁾. 从这个事件发生顺序来看, 应当将 (V_1, V_2) 放在前面, (ϵ_1, ϵ_2) 放在后面. 从这个思路出发, 可以尝试列举并验证, 在如下两个顺序的 G - 独立性关系中:

$$(S1) \quad V_1 \dashrightarrow V_2 \dashrightarrow \epsilon_1 \dashrightarrow \epsilon_2;$$

$$(S2) \quad V_2 \dashrightarrow V_1 \dashrightarrow \epsilon_2 \dashrightarrow \epsilon_1.$$

随机向量 (W_1, W_2) 的次线性期望都满足 (3.6), 换言之, 它们是 M - 独立的. 而且这两种顺序都等价于¹⁵⁾ 如下关系:

(1) V_1 和 V_2 是相互 G - 独立的;

(2) ϵ_1 和 ϵ_2 是相互经典独立的;

(3) $(V_1, V_2) \dashrightarrow (\epsilon_1, \epsilon_2)$.

在上述关系中, V_1 和 V_2 , 以及 ϵ_1 和 ϵ_2 之间, 具有对称性. 为方便讨论, 对于可分解的 M - 正态分布, 将顺序 (S1) 称为 W_2 半序列独立于 W_1 , 即半序列独立性 (semi-sequential independence), 记为 $W_1 \xrightarrow{S} W_2$. 从上面的讨论可以看出, $W_1 \xrightarrow{S} W_2$ 等价于 $W_2 \xrightarrow{S} W_1$, 都可以推导出二者之间的 M - 独

13) 这个随机向量的任何可逆线性变换也都不服从 G - 正态分布.

14) 这里旋转两个旋钮 (或者敲击两个音叉), 可能同时或者不同时, 其实没有关系 (并不影响观测者感知到的分布不确定性), 两个旋钮 (或两个音叉) 之间的关系是对称的. 换言之, (V_1, V_2) 或 (ϵ_1, ϵ_2) 的内部, 应当是满足对称性的 (这个直觉将在后面得到理论验证).

15) 由于这种等价关系, 读者可以验证, 其实还有另外两种顺序 (只是将 V 的部分内部, 或 ϵ 的部分内部进行互换), 都是等价的.

立性. 而且在可分解的情形下, 也可以得到一个类似于定理 3.1 的结论: W_1 和 W_2 之间的半序列独立性, 等价于 (W_1, W_2) 服从二元 M - 正态分布 $\mathcal{N}_M(0, [\sigma^2, \bar{\sigma}^2]I_2)$.

注意到这里 $(V_1, \epsilon_1, V_2, \epsilon_2)$ 的 G - 独立性顺序很重要, 如果修改为以下两种顺序:

(F1) $V_1 \dashrightarrow \epsilon_1 \dashrightarrow V_2 \dashrightarrow \epsilon_2$;

(F2) $V_2 \dashrightarrow \epsilon_2 \dashrightarrow V_1 \dashrightarrow \epsilon_1$.

将顺序 (F1) 称为 W_2 全序列独立于 W_1 , 记为 $W_1 \xrightarrow{F} W_2$. 全序列独立性 (fully-sequential independence) 是比序列独立性 (sequential independence) 或 G - 独立性更强的一种关系. 其中顺序 (F1) 可以推导出 $W_1 \dashrightarrow W_2$, (F2) 意味着 $W_2 \dashrightarrow W_1$, 这两者不等价 (它们之间是一种非对称的 G - 独立性). 这种理论上的非对称性, 有以下直观解释. 例如, (F1) 可以视为, 先旋转第一个音叉的旋钮 (V_1), 敲击音叉生成噪声 (由 ϵ_1 主导); 再去旋转第二个音叉的旋钮 (V_2), 敲击音叉产生噪声 (ϵ_2). 由于在敲击第二个音叉之前, 已经观测到了前一个音叉的噪声, 所以在旋转第二个旋钮时, 有可能受到前一个噪声的影响. 因此, 它们之间的关系不再对称 (因为第一个音叉的旋钮位置不可能受到第二个音叉的噪声的影响). 总结来说, W_1 和 W_2 之间具备对称还是非对称的独立性, 完全取决于 $(V_1, \epsilon_1, V_2, \epsilon_2)$ 内部的 G - 独立性顺序.

为了描述 M - 正态分布在半序列独立性 (或 M - 独立性)、序列独立性 (或 G - 独立性) 和全序列独立性下的分布不确定性, 可以得到对应的一系列表示 (参见文献 [19, 第 2.1.7 小节]), 并且在此基础上, 可以得到 M - 独立性下的 M - 正态分布与 G - 独立性下的 G - 正态分布之间的特殊联系 (参见文献 [19, 定理 2.4.12]), 是对命题 3.2 的一个扩展.

4 M - 分布族和独立性

在介绍了 M - 正态分布及其相关的独立性的思想后, 本节给出 M - 分布和 M - 独立性的一般定义, 更多理论细节见附录 B.

定义 4.1 在次线性期望空间 $(\Omega, \mathcal{H}, \hat{\mathbb{E}})$ 中, 称一个 d - 维随机向量 W 服从 M - 分布 (或称其分布属于 M - 分布族), 如果存在一个紧且凸¹⁶⁾ 的集合 $\Theta \subset \mathbb{R}^m$ 、一个服从经典分布的随机向量 $\eta : \Omega \rightarrow \mathbb{R}^n$ 和一个实值的 Borel 可测函数 $f : \Theta \times \mathbb{R}^n \rightarrow \mathbb{R}^d$, 使得对于任意 $\varphi \in C_{b, \text{Lip}}$, 有

$$\hat{\mathbb{E}}[\varphi(W)] = \sup_{\theta \in \Theta} \mathbb{E}[\varphi(f(\theta, \eta))].$$

此时记

$$W \sim f(\Theta, P_\eta) := \{P_{f(\theta, \eta)}, \theta \in \Theta\}.$$

定义 4.2 称一个服从 M - 分布的随机向量 $W \sim f(\Theta, P_\eta)$ 是可分解的, 如果存在一个服从最大分布的 $K \sim \mathcal{M}(\Theta)$ 和一个服从经典分布的 η , 满足

$$K \dashrightarrow \eta,$$

且

$$W = f(K, \eta).$$

¹⁶⁾ 这个性质只是后面考虑最大分布时的技术要求 (参见文献 [24, 第 2.2 小节]), 而 M - 分布定义本身对 Θ 的性质要求可以放宽.

注 4.1 分解的存在性来自 G -独立性的存在性 (与注 3.3 类似). 实际上, 所有可分解的服从 M -分布的随机变量构成了一个乘积空间 $\bar{\mathcal{H}}_s$ (在附录 B.1 中定义).

注 4.2 大家所熟知的很多经典概率分布在次线性期望意义下仍然存在¹⁷⁾. 一种简单直观的验证思路是, 首先有 $\epsilon \sim \mathcal{N}_G(0, [1, 1])$, 它服从标准正态分布 $N(0, 1)$, 其经典意义下的累积分布函数 Φ 定义如下:

$$\Phi(x) := \hat{\mathbb{E}}[\mathbb{1}_{\epsilon \leq x}] = \mathbb{E}[\mathbb{1}_{\epsilon \leq x}].$$

设 $U := \Phi(\epsilon)$. 注意 Φ 是一个有界且连续的函数, 因此 $U \in \mathcal{H}$. 接着可以验证 U 服从 $[0, 1]$ 上的 (经典) 均匀分布. 对于具有累积分布函数 F 的经典分布, 考虑 F 的广义逆

$$F^{-1}(y) := \inf\{x : F(x) \geq y\}.$$

令 $X := F^{-1}(U)$. 只需要对 F 添加适当条件以确保 $\hat{\mathbb{E}}[|X|] < \infty$, 从而使 $X \in \mathcal{H}$. 因此, 获得一个服从具有分布函数为 F 的随机变量 X .

定义 4.3 (M -独立性) 对于 $X \sim f(\Theta_x, P_\xi)$ 和 $Y \sim g(\Theta_y, P_\eta)$, 称它们是 M -独立的, 如果

$$\hat{\mathbb{E}}[\varphi(X, Y)] = \sup_{(\theta_x, \theta_y) \in \Theta_x \times \Theta_y} \mathbb{E}[\varphi(f(\theta_x, \xi), g(\theta_y, \eta))],$$

其中 ξ 和 η 是 (经典意义下) 独立的.

注 4.3 对于任意正整数 N , 类似地可以得到关于 $\{X_i\}_{i=1}^N$ 的 M -独立性的定义. 称 $\{X_i\}_{i=1}^\infty$ 是 M -独立的, 如果对于任意 N , 序列 $\{X_i\}_{i=1}^N$ 是 M -独立的.

接下来的讨论将主要集中在可分解的 M -版本对象上, 因为如果有分解 $X = f(K, \eta)$, 这将为深入讨论带来更多便利和丰富的结果. 因此, 在接下来的内容中, 如果没有引起任何混淆, 我们不会过多区分半序列独立性和 M -独立性 (前者只是后者的略强版本, 并且前者是为可分解的 M -分布设计的). 简而言之, 当两个可分解的 M -版本对象 M -独立时, 实际上是指它们是半序列独立的 (在组内有更多的独立性).

定义 4.4 称序列 $\{X_i\}_{i=1}^\infty = \{f_i(K_i, \eta_i)\}_{i=1}^\infty \subset \bar{\mathcal{H}}_s$ 是 M -独立的, 如果对任意正整数 N , $\{X_i\}_{i=1}^N$ 是 M -独立的, 即

- (1) $\{K_i\}_{i=1}^N$ 是 G -独立的;
- (2) $\{\eta_i\}_{i=1}^N$ 是经典独立的;
- (3) $(K_1, K_2, \dots, K_N) \dashrightarrow (\eta_1, \eta_2, \dots, \eta_N)$.

称 $\{X_i\}_{i=1}^\infty$ 是 M -独立同分布 (或 M -i.i.d.), 如果它们满足 M -独立性且 $X_i \stackrel{d}{=} X_{i+1}$ 对任意 i 都成立.

注 4.4 通过第 3.3 小节中关于对称性的讨论, 可以看出 X_i 之间的顺序是可以任意打乱的. 关于 N 个以及可列个满足 M -独立性的随机变量的存在性, 见附录 B.2.

例 4.1 列举了若干具有代表性的 M -分布的示例, 从中可以看出 M -分布在描述分布不确定性时的结构灵活性和广泛性.

例 4.1 在以下例子中, 都统一使用定义 4.2 中的符号和设定, 其中有独立性 $K \dashrightarrow \eta$.

(1) 考虑 $K \sim \mathcal{M}[\underline{\sigma}, \bar{\sigma}]$, $\eta \sim N(0, 1)$ 且 $f(x, y) = xy$, 则 W 服从 M -正态分布, 其分布不确定性可以描述为

$$\{N(0, \sigma^2) : \sigma \in [\underline{\sigma}, \bar{\sigma}]\}.$$

¹⁷⁾ 一种理论解读是, 将经典的 (允许常见的概率分布存在的) 线性期望看成退化的次线性期望, 进而考虑乘积空间 (详见附录 B.1). 这里给出一种 (非正式但较直观的) 思路, 直接用退化的 G -正态分布来进行构造.

(2) 考虑 $K = (U, V) \sim \mathcal{M}([\underline{\mu}, \bar{\mu}] \times [\underline{\sigma}, \bar{\sigma}])$, $\eta \sim N(0, 1)$ 并且 $W = U + V\eta$, 则 W 的分布不确定性可以描述为

$$\{N(\mu, \sigma^2) : \mu \in [\underline{\mu}, \bar{\mu}], \sigma \in [\underline{\sigma}, \bar{\sigma}]\},$$

其中也包含了一系列 Gauss 混合模型 (均值和方差服从任意在对应区间上的先验分布).

(3) (M - 指数分布) 考虑 $K \sim \mathcal{M}[\underline{\lambda}, \bar{\lambda}]$, $\eta \sim \exp(1)$ 并且 $W = K\eta$, 则 W 的分布不确定性可以写为

$$\{\exp(\lambda) : \lambda \in [\underline{\lambda}, \bar{\lambda}]\},$$

其中每个 $\exp(\lambda)$ 有密度函数 $f(x) = \frac{1}{\lambda} e^{-x/\lambda} \mathbb{1}_{\{x \geq 0\}}$.

(4) (M -Bernoulli 分布) 对于 $0 \leq \underline{p} \leq \bar{p} \leq 1$, 考虑 $\eta \sim \text{Unif}[0, 1]$, $K \sim \mathcal{M}[\underline{p}, \bar{p}]$, 并且

$$W = \mathbb{1}_{\{\eta - K < 0\}},$$

则 W 具有分布不确定性

$$\{\text{Bern}(p) : p \in [\underline{p}, \bar{p}]\}.$$

(5) (M - 二项分布) 考虑与上一示例中同样的 K . 设 $\eta = (\eta_i)_{i=1}^n$, 其中 $\eta_i \sim \text{Unif}[0, 1]$ 且满足经典的独立同分布条件. 令 $W_i := \mathbb{1}_{\{\eta_i - K < 0\}}$, 则每个 W_i 服从具有参数不确定性 $[\underline{p}, \bar{p}]$ 的 M -Bernoulli 分布. 考虑它们的求和

$$S_n = \sum_{i=1}^n W_i,$$

则 S_n 具有分布不确定性

$$\{\text{Bin}(n, p) : p \in [\underline{p}, \bar{p}]\}.$$

(6) (M -Poisson 二项分布) 在上一个示例中, W_i 共享同一个 K . 如果改变设置为 $W_i := \mathbb{1}_{\{\eta_i - K_i < 0\}}$, 其中 $K_i \sim \mathcal{M}[\underline{p}, \bar{p}]$ 是 G - 独立的, 且满足

$$(K_1, K_2, \dots, K_n) \dashrightarrow (\eta_1, \eta_2, \dots, \eta_n),$$

换言之, $\{W_i\}_{i=1}^n$ 是 M - 独立的, 则 S_n 具有分布不确定性

$$\{\text{PB}(p_1, \dots, p_n) : p_i \in [\underline{p}_i, \bar{p}_i]\},$$

其中 $\text{PB}(p_1, \dots, p_n)$ 代表经典的 Poisson 二项分布.

(7) (回归模型) 对于 $i = 1, 2, \dots, N$, 设 $K_i = (U_i, B_i, V_i) \sim \mathcal{M}([\underline{\mu}, \bar{\mu}] \times [\underline{\beta}, \bar{\beta}] \times [\underline{\sigma}, \bar{\sigma}])$, 以及 η_i 服从均值为 0 且方差为 1 的经典分布. 它们满足

$$K_1 \dashrightarrow K_2 \dashrightarrow \dots \dashrightarrow K_N \dashrightarrow \eta_1 \dashrightarrow \eta_2 \dashrightarrow \dots \dashrightarrow \eta_N.$$

对于观测数据 $\{(x_i, y_i)\}_{i=1}^N$, 定义误差项 $W_i = V_i \eta_i$. 设因变量为随机变量 Y_i , 考虑回归模型

$$Y_i = U_i + B_i x_i + W_i. \quad (4.1)$$

根据上述设置, Y_i 服从 M - 分布且 $\{Y_i\}_{i=1}^N$ 为一列 M - 独立的随机变量. 同时, (4.1) 给出了一类考虑到分布不确定性 (包括截距、斜率及方差不确定性) 的线性回归模型的一般理论形式. 关于分布不确定性下的回归问题 (主要针对确定性的斜率即 $\underline{\beta} = \bar{\beta} = \beta$ 的情形) 已有一系列探索和讨论 [21, 22, 29, 30],

而 (4.1) 能够给出一个统一的理论框架, 尤其是通过 M - 结构实现了理论与数据层面尤其是数据模拟 (simulation) 的紧密结合. 例如, 文献 [30] 中的数据生成部分 (第 3.1 小节) 可视为对 (4.1) 的一个特例 (假定斜率确定性以及 W_i 服从 $\mathcal{N}_M(0, [\underline{\sigma}^2, \bar{\sigma}^2])$ 分布) 的模拟¹⁸⁾. 由于这种理论和数据的对应关系, 我们能够用 M - 结构来对类似于 max-mean 方法 [15] 的估计量的行为给出更加精细的刻画. 特别地, 可以考虑其在有限样本下的 M - 分布和渐近性质 (基于将在第 5 节中给出的 M - 中心极限定理), 进而能够和已有的统计理论和方法进行更加密切的结合.

在上述的回归模型中, 不仅因变量 Y_i 具有分布不确定性¹⁹⁾, 在非退化情形即 $\underline{\beta} < \bar{\beta}$ 时, 还涉及两个变量间相依关系 (dependence) 的不确定性. 在实践中, 如果观测到的两个变量均服从经典分布, 是否意味着它们之间的相依性也可由单一确定性模型刻画? 简而言之, 两个服从经典分布的随机变量之间的相依性一定是确定的吗? 其实未必, 借助 M - 结构, 例 4.2 给出了一个简洁反例: 两个随机变量均服从经典的 $N(0, 1)$, 但是它们之间的协方差是不确定的. 换言之, 即使两个随机变量都服从经典的分布 (可以被线性期望刻画), 它们的联合分布可能仍然具备分布不确定性 (需要用次线性期望来刻画). 此现象也被称为相依不确定性 (dependence uncertainty)²⁰⁾.

在以下例 4.2 中, 借助 M - 结构在次线性期望意义下给出一个关于相依不确定性的简单例子.

例 4.2 考虑一个服从 (经典的) 标准二元正态分布的随机向量 $\eta = (\eta_1, \eta_2)$ 和服从最大分布的 $K \sim \mathcal{M}[\underline{\rho}, \bar{\rho}]$, 其中 $[\underline{\rho}, \bar{\rho}] \subset [-1, 1]$, 并假设 $K \dashrightarrow \eta$. 令 $X = \eta_1$, $Y = K\eta_1 + \sqrt{1 - K^2}\eta_2$. 由于经典正态分布的稳定性, 对于任意测试函数 $\varphi \in C_{b, \text{Lip}}(\mathbb{R})$, 有

$$\hat{\mathbb{E}}[\varphi(Y)] = \max_{k \in [\underline{\rho}, \bar{\rho}]} \mathbb{E}[\varphi(k\eta_1 + \sqrt{1 - k^2}\eta_2)] \equiv \mathbb{E}[\varphi(\eta_1)].$$

虽然 Y 依赖于服从最大分布的 K , 但其分布始终为经典的 $N(0, 1)$, 而 K 真正带来的是 X 与 Y 之间的相依不确定性. 例如, 它们具备协方差不确定性:

$$\hat{\mathbb{E}}[XY] = \max_{k \in [\underline{\rho}, \bar{\rho}]} \mathbb{E}[k\eta_1^2 + \sqrt{1 - k^2}\eta_1\eta_2] = \max_{k \in [\underline{\rho}, \bar{\rho}]} k = \bar{\rho}.$$

类似可得 $-\hat{\mathbb{E}}[-XY] = \underline{\rho}$.

5 M - 中心极限定理

在明确了 M - 分布族和 M - 独立性之后, 本节进一步给出 M - 中心极限定理 (M -CLT). 它所刻画模型不确定性的程度介于经典中心极限定理和 G - 中心极限定理 (G -CLT) 之间, 也体现了 M - 正态分布在整个 M - 结构中的中心地位.

称一个经典的 ϵ 服从标准化的分布, 如果它的均值为 0 且方差为 1. 考虑 $\bar{\mathcal{H}}_s$ 的一个子集:

$$\mathcal{H}_s = \{X \in \bar{\mathcal{H}}_s : X = V\epsilon, \{V, \epsilon\} \subset \bar{\mathcal{H}}_s, V \sim \mathcal{M}[\underline{\sigma}, \bar{\sigma}], \epsilon \sim \text{标准化}, V \dashrightarrow \epsilon\}.$$

这里 $\mathcal{H}_s$ 可以被视为一个具有零均值和方差不确定性的 M - 分布类. 当前版本的 M - 中心极限定理围绕这样一个分布类进行展开.

18) 注意这里不能看成是对 G - 正态分布 $\mathcal{N}_G(0, [\underline{\sigma}^2, \bar{\sigma}^2])$ 和 G - 独立性的模拟, 后者的模拟是困难的, 详见文献 [26, 第 1 节] 和 [19, 第 3.5 小节].

19) 当然也可将自变量视为随机变量 X_i , 并允许其具备分布不确定性.

20) 这是稳健型风险度量领域里的一个常见概念, 可参见文献 [2].

定理 5.1 (M - 中心极限定理) 考虑一列 M - 独立同分布的序列 $\{X_i\}_{i=1}^\infty \subset \mathcal{H}_s$, 具有确定的零均值和不确定方差:

$$\underline{\sigma}^2 = -\hat{\mathbb{E}}[-X_1^2] \leq \hat{\mathbb{E}}[X_1^2] = \bar{\sigma}^2.$$

当 $n \rightarrow \infty$ 时, 有 $\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \xrightarrow{d} W$, 其中 $W \sim \mathcal{N}_M(0, [\underline{\sigma}^2, \bar{\sigma}^2])$. 进一步地, 对于任意 $\varphi \in C(\mathbb{R})$ 且满足线性增长条件 $|\varphi(x)| \leq C(1 + |x|)$, 有

$$\lim_{n \rightarrow \infty} \hat{\mathbb{E}} \left[\varphi \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \right) \right] = \hat{\mathbb{E}}[\varphi(W)].$$

注 5.1 依分布收敛 (定义 2.1) 只需要考虑测试函数 $\varphi \in C_{b,\text{Lip}}(\mathbb{R})$ 即可, 这里给出了在更广泛意义下的测试函数空间. 注意到 M - 中心极限定理并不需要在 G - 中心极限定理中的条件

$$\lim_{\lambda \rightarrow \infty} \hat{\mathbb{E}}[(|X_1|^2 - \lambda)^+] = 0.$$

当考虑退化情形即 $\underline{\sigma} = \bar{\sigma} = \sigma$ 时, M - 中心极限定理将退化为经典的中心极限定理, 且不需要添加额外的条件. 这与经典的中心极限定理一致.

在定理 5.1 的证明中, 我们将 Lindeberg 的留一法²¹⁾ 与次线性期望在 M - 结构的设定下相结合. 我们能够这样做的一个原因是, M - 独立性的对称性: X_i 与 $\{X_j, j \neq i\}$ 是 M - 独立的 (注意, 不能在序列独立性或 G - 独立性下进行这样的结合, 这是因为它的非对称性). 此外, 定理 5.1 为展示 M - 正态分布的稳定性提供了一种直接方式 (推论 5.1).

考虑 $W \in \mathcal{H}_s$, 称 $\bar{W} \in \mathcal{H}_s$ 是 W 的一个 M - 独立拷贝, 如果 $W \stackrel{d}{=} \bar{W}$ 且它们是 M - 独立的.

推论 5.1 (稳定分布) 考虑 $W, \bar{W} \in \mathcal{H}_s$ 并且 $\hat{\mathbb{E}}[W^2] < \infty$. 以下 3 个陈述等价:

- (1) $W + \bar{W} \stackrel{d}{=} \sqrt{2}W$ 对任意 W 的 M - 独立拷贝 $\bar{W}$;
- (2) $aW + b\bar{W} \stackrel{d}{=} \sqrt{a^2 + b^2}W, \forall a, b \geq 0$ 对任意 W 的 M - 独立拷贝 $\bar{W}$;
- (3) W 服从 M - 正态分布.

注 5.2 可以用推论 5.1(1) 或 5.1(2) 作为 M - 正态分布的一个等价定义.

如果将 M -CLT (定理 5.1) 中 X_i 之间的独立性换成 G - 独立性, 则根据 G - 中心极限定理 (G -CLT), 标准化和就会依分布收敛到 G - 正态 $\mathcal{N}_G(0, [\underline{\sigma}^2, \bar{\sigma}^2])$. 下面借助第 3.3 小节中的设置, 对 M -CLT 和 G -CLT 作一个直观的比较和解释 (详见文献 [19, 第 2.3 小节]). 将 $\{X_i\}_{i=1}^n = \{V_i \epsilon_i\}_{i=1}^n$ 视为一列带有旋钮的音叉作为噪声源: V_i 可以看成旋钮的位置 (范围是 $[\underline{\sigma}, \bar{\sigma}]$), 用于调节噪声 $V_i \epsilon_i$ 的波动程度, 其中 ϵ_i 服从任意经典的标准化的分布. 正如第 3.3 小节中提到的关于 M - 独立性和 G - 独立性的对比, 当 n 很大时, 考虑以下两种情形.

(1) 如果这一系列旋钮的位置从一开始就是未知但固定的 (且各个旋钮的位置可能不同), 然后再敲击音叉发出声音, 噪声 X_i 之间的独立性可以被视为 M - 独立性 (定义 4.4), 从而根据 M -CLT, 这一系列噪声的标准化和的分布不确定性²²⁾ 会接近于 M - 正态分布 $\mathcal{N}_M(0, [\underline{\sigma}^2, \bar{\sigma}^2])$.

(2) 如果旋钮的位置是随着敲击音叉的过程中逐个调整的 (即第 i 个旋钮的位置可能受到来自之前 $i-1$ 个音叉的噪声的影响), X_i 之间的独立性可以被视为 G - 独立性, 则这一系列噪声的标准化和

21) 即 “leave one out” 方法, 参见文献 [4, 第 8.6 小节].

22) 注意, M - 正态分布描述一种分布不确定性. 例如, 考虑一种特殊情形, 这一系列旋钮的位置中有比例 $p \in [0, 1]$ 是 $\underline{\sigma}$, 余下的是 $\bar{\sigma}$, 而给定旋钮位置后噪声之间 (或 ϵ_i 之间) 满足经典的独立性, 则对于给定的 p , 可以验证标准化和的渐近分布是一个经典正态, 而它的方差 $\sigma^2 p + \bar{\sigma}^2 (1-p)$ 依赖于这个未知的比例 p . 由于 p 是不确定的, 所以这个渐近分布的分布不确定性集合就可以被 M - 正态分布刻画.

的分布不确定性会接近于 G -正态分布 $\mathcal{N}_G(0, [\underline{\sigma}^2, \bar{\sigma}^2])$, 这也可以看成是对 G -CLT 以及文献 [28] 中的结果的一种直观解释.

6 结论和展望

本文的主要作用是呈现 M -结构的理论基础, 这个结构仍处于初期的发展阶段, 但是此结构已经能够允许关于模型不确定性意义下的基本概念 (分布、独立性和中心极限定理) 的更加细致的讨论, 尤其是可以解释非线性期望框架下的概念与经典概念的“差异”的来源, 例如, G -独立性的非对称性来自于哪里以及如何理解 (第 3.3 小节). 同时也为经典框架和非线性期望框架在理论和直觉上提供了一个连接, 进而在底层结构的意义上, 为各种模型不确定性问题的讨论带来更多便利和灵活性.

总体来说, M -结构可以从以下视角来解读.

(1) 在直觉上, 可以从一个服从经典分布的随机变量 η 出发, 考虑变换 $f(k, \eta)$, 若其中的变量或参数 k 取值于 Θ , 但其分布是不确定的, 则可以将其换成服从最大分布 $\mathcal{M}(\Theta)$ 的 K , 从而得到一个服从 M -分布的随机变量 $f(K, \eta)$ (详见定义 4.2). 这种设定自然涵盖了各种参数化分布在参数不确定性意义下的情形 (例 4.1). 更进一步地, 在一个统计模型 (或统计建模的过程) 中, 可以将有确定分布的部分归结到 η 以及 f 本身的形式中, 而将分布不确定的部分归结到 K 的部分中. 进而可以从底层上考虑分布不确定性并进行理论分析 (例如, 可以讨论在这个意义下模型仍然具备的一些确定性的性质). 随着数据和信息的积累, 这两个部分也可相应进行动态更新.

(2) 在理论上, 可视为由非线性期望下的最大分布借助经典分布衍生出来的一大类 (非线性期望意义下的) 分布, 以及由最大分布之间的独立性衍生出来一种新的独立性 (即 M -独立性), 从而保留了对称性. 更进一步地, M -结构可以看成是唯二的两类允许对称的 (非线性) 独立性的分布的结合 (即最大分布和经典分布, 详见定理 2.1). 如果出现多个在空间上对称的随机变量, 都有分布不确定性, 它们之间如果具备某种“独立性” (如空间上分割开来的两个随机变量), 由于对称性, 这种独立性其实只能是 M -独立性, 而非 G -独立性. 从这一视角来看, M -结构扩充并丰富了能够 (在非线性期望意义下) 讨论的模型不确定性问题的范畴.

除了独立性部分的新的洞察, M -结构带来的另一个理论优势是结构上的灵活性, 如一元和多元正态分布之间的在独立性下的直接关联. 注意到多个 M -独立的 M -正态分布组成的随机向量仍然服从多元 M -正态分布, 这个性质是 G -正态分布所不具备的. 更进一步地, 如果去定义 M -正态分布对应的 M -Brown 运动, 则它仍然是一个 Gauss 过程, 而 G -Brown 运动并不是一个 Gauss 过程 [25]. 总之, M -结构在多元意义上的灵活性为后面统计意义上的讨论及相关的随机场理论的发展都提供了更多的可能性.

在 M -结构的基础上, 我们将会在未来的研究工作中展现关于 M -结构与不同实践领域的关联, 尤其是关于具有对称独立性场景模型不确定性问题的讨论 (如截面数据场景). 我们目前已经在发展的关联包含计算和数据层面, 涉及统计和金融领域, 具体如下.

(1) 计算层面包括 G -EM (expectation-maximization) 算法, 对于给定的测试函数 φ , 通过选取特定的 M -分布族, 以及结合 M -独立性和 G -独立性, 来进行各种时空结构下的分布不确定性所对应的次线性期望的计算, 以及通过存储其中的转换机制 (switching scheme) 来为后续的取样作准备.

(2) 数据层面上, 数据类型可分为 (常见的) 数值型数据 (与 max-mean 估计方法 [15] 以及基于分组思想的数据模拟方法紧密相关) 和区间型数据 (即每个观测用一个区间表示, 与 M -结构有直接对

表 1 正态分布、 M - 正态分布和 G - 正态分布的比较

	正态分布 $N(0, \sigma^2)$	M - 正态分布 $\mathcal{N}_M(0, [\underline{\sigma}^2, \bar{\sigma}^2])$	G - 正态分布 $\mathcal{N}_G(0, [\underline{\sigma}^2, \bar{\sigma}^2])$
期望	线性	次线性	次线性
一阶矩	确定的 (0)	确定的 (0)	确定的 (0)
二阶矩	确定的 (σ^2)	不确定的 ($[\underline{\sigma}^2, \bar{\sigma}^2]$)	不确定的 ($[\underline{\sigma}^2, \bar{\sigma}^2]$)
三阶矩	确定的 (0)	确定的 (0)	不确定的
独立性	经典独立($\perp$)	M - 独立或半序列($\overset{S}{\dashrightarrow}$)	G - 独立或序列($\dashrightarrow$)
	对称的	对称的	非对称的
(设置)	$X \stackrel{d}{=} \bar{X} \stackrel{d}{=} N(0, \sigma^2)$	$X \stackrel{d}{=} \bar{X} \stackrel{d}{=} \mathcal{N}_M(0, [\underline{\sigma}^2, \bar{\sigma}^2])$	$X \stackrel{d}{=} \bar{X} \stackrel{d}{=} \mathcal{N}_G(0, [\underline{\sigma}^2, \bar{\sigma}^2])$
	$X \perp \bar{X}$	$X \overset{S}{\dashrightarrow} \bar{X}$	$X \dashrightarrow \bar{X}$
稳定性	$X + \bar{X} \stackrel{d}{=} \sqrt{2}X$	$X + \bar{X} \stackrel{d}{=} \sqrt{2}X$	$X + \bar{X} \stackrel{d}{=} \sqrt{2}X$
多元分布	$(X, \bar{X}) \stackrel{d}{=} N(0, \sigma^2 I_2)$	$(X, \bar{X}) \stackrel{d}{=} \mathcal{N}_M(0, [\underline{\sigma}^2, \bar{\sigma}^2] I_2)$	$(X, \bar{X}) \stackrel{d}{\neq} \mathcal{N}_G(0, [\underline{\sigma}^2, \bar{\sigma}^2] I_2)$
中心极限定理	经典中心极限定理	M - 中心极限定理	G - 中心极限定理

应关系). 特别地, 数据模拟对于一个统计模型和方法的长远发展至关重要, 正如在讨论回归模型 (4.1) 时提及的, M - 结构在数据模拟上具有优势, 从而能够实现理论和数据实践更加精细的结合. 具体而言, 包括以下两个紧密相关的板块.

(i) 数据搜集和生成 (也称为 Alice's side): 如何在分布不确定性意义下及特殊的信息设置下来进行采样, 生成的数据可以看成关于一种 M - 分布意义下的分布不确定性的模拟. 在结合了 G -EM 算法的计算后, 从而实现一些既定的目标函数的优化. 这个过程其实也与最近的关于强化学习中的多臂老虎机问题和取样策略的讨论 [6–8] 的背景不谋而合.

(ii) 数据分析 (也称为 Bob's side): 给定一个实际数据, 如何检验数据本身是否具备显著的分布不确定性²³⁾, 这将决定是否需要引入次线性期望的相关方法进行讨论, 我们给出了一个关于分布不确定性的可视化 (visualization) 和假设检验. 如果具备显著的分布不确定性, 可以借助 M - 结构来更加系统化地去构建背后的分布不确定性集合, 进而在此基础上考虑统计模型.

(3) 在金融决策领域里, 有一大类问题是, 如何在模型不确定意义性下进行稳健的资产投资 (robust portfolio optimization), 而 M - 结构在这里可以提供一个关于资产之间的协方差不确定性的衡量方式, 进而给出一个如何最小化资产的方差不确定性程度的方法, 并为方差不确定性进行一个类似于有效前沿 (efficient frontier) 的可视化.

而上述的关联之所以能够产生, 很大程度上是得益于 M - 结构的灵活性和可扩展性. 关于目前这些发展的更多细节可以参见文献 [19], 我们也会在未来的研究工作中进行更加完整的展示. 最后用表 1 来总结 M - 结构的核心内容与角色.

致谢 衷心感谢审稿人在本文形成过程中提出的宝贵修改建议. 特别感谢彭实戈教授于 2017 年 5 月应邀访问我系期间, 以及随后在 Fields 研究所的深入交流, 这些讨论促使我们深入思考次线性期望下的一类估计方法背后的分布结构, 这成为促成本文提出 M - 结构的重要启发之一.

参考文献

- 1 Bayraktar E, Munk A. Comparing the G-normal distribution to its classical counterpart. *Commun Stoch Anal*, 2015, 9: 1–18

²³⁾ 一个比较典型的例子是时序型的数据, 在不同的时间段内, 数据背后的分布模型可能会发生系统性的改变.

- 2 Bernard C, Jiang X, Wang R D. Risk aggregation with dependence uncertainty. *Insurance Math Econom*, 2014, 54: 93–108
- 3 Bion-Nadal J, Kervarec M. Risk measuring under model uncertainty. *Ann Appl Probab*, 2012, 22: 213–238
- 4 Breiman L. Probability. Philadelphia: SIAM, 1992
- 5 Chatfield C. Model uncertainty, data mining and statistical inference. *J Roy Statist Soc Ser A*, 1995, 158: 419–466
- 6 Chen Z J, Epstein L G, Zhang G D. A central limit theorem, loss aversion and multi-armed bandits. *J Econom Theory*, 2023, 209: 105645
- 7 Chen Z J, Epstein L G, Zhang G D. Approximate optimality and the risk/reward tradeoff given repeated gambles. *Econom Theory*, 2025, 79: 1351–1381
- 8 Chen Z J, Yan X D, Zhang G D. Strategic two-sample test via the two-armed bandit process. *J R Stat Soc Ser B Stat Methodol*, 2023, 85: 1271–1298
- 9 Crandall M G, Ishii H, Lions P L. User's guide to viscosity solutions of second order partial differential equations. *Bull Amer Math Soc (NS)*, 1992, 27: 1–67
- 10 Denis L, Hu M S, Peng S G. Function spaces and capacity related to a sublinear expectation: Application to G -Brownian motion paths. *Potential Anal*, 2011, 34: 139–161
- 11 Föllmer H, Schied A. Stochastic Finance: An Introduction in Discrete Time, 5th ed. Berlin: De Gruyter, 2025
- 12 Hu M S. Explicit solutions of the G -heat equation for a class of initial conditions. *Nonlinear Anal*, 2012, 75: 6588–6595
- 13 Hu M S, Li X J. Independence under the G -expectation framework. *J Theoret Probab*, 2014, 27: 1011–1020
- 14 Ji X J, Peng S G. Spatial and temporal white noises under sublinear G -expectation. *Sci China Math*, 2020, 63: 61–82
- 15 Jin H Q, Peng S G. Optimal unbiased estimation for maximal distribution. *Probab Uncertain Quant Risk*, 2021, 6: 189–198
- 16 Kiureghian A D, Ditlevsen O. Aleatory or epistemic? Does it matter? *Struct Saf*, 2009, 31: 105–112
- 17 Knight F H. Risk, Uncertainty and Profit. Boston-New York: Houghton Mifflin Company, 1921
- 18 Kolmogorov A N. Foundations of the Theory of Probability. New York: Chelsea, 1956
- 19 Li Y F. Statistical roles of the G -expectation framework in model uncertainty: The semi- G -structure as a stepping stone. PhD Thesis. London: The University of Western Ontario, 2022
- 20 Li Y F, Kulperger R, Yu H. Semi- G -normal: A hybrid between normal and G -normal (full version). arXiv:2104.04910, 2021
- 21 Lin L, Dong P, Song Y Q, et al. Upper expectation parametric regression. *Statist Sinica*, 2017, 27: 1265–1280
- 22 Lin L, Shi Y F, Wang X, et al. k -sample upper expectation linear regression—modeling, identifiability, estimation and prediction. *J Statist Plann Inference*, 2016, 170: 15–26
- 23 Peng S G. Theory, methods and meaning of nonlinear expectation theory (in Chinese). *Sci Sin Math*, 2017, 47: 1223–1254 [彭实戈. 非线性期望的理论、方法及意义. 中国科学: 数学, 2017, 47: 1223–1254]
- 24 Peng S G. Nonlinear Expectations and Stochastic Calculus under Uncertainty: With Robust CLT and G -Brownian motion. Heidelberg-Berlin: Springer-Verlag, 2019
- 25 Peng S G. G -Gaussian processes under sublinear expectations and q -Brownian motion in quantum mechanics. *Numer Algebra Control Optim*, 2023, 13: 583–603
- 26 Peng S G, Yang S Z. Distributional uncertainty of the financial time series measured by G -expectation. *Theory Probab Appl*, 2022, 66: 729–741
- 27 Pursell L E. Uniform approximation of real continuous functions on the real line by infinitely differentiable functions. *Math Mag*, 1967, 40: 263–265
- 28 Rokhlin D B. Central limit theorem under uncertain linear transformations. *Statist Probab Lett*, 2015, 107: 191–198
- 29 Xu Q, Xuan X H. Nonlinear regression without i.i.d. assumption. *Probab Uncertain Quant Risk*, 2019, 4: 8
- 30 Yang S Z, Yao J F. Linear regression under model uncertainty. *Probab Uncertain Quant Risk*, 2023, 8: 523–546

附录 A 证明

附录 A.1 第 3.1 小节中的证明

命题 3.1 的证明与最大分布的定义紧密相关. 首先, 需要准备以下引理.

引理 A.1 考虑 $\varphi \in C_{1,\text{Lip}}(\mathbb{R})$. 对于任意固定的 $v \in [\underline{\sigma}, \bar{\sigma}]$, 定义 $\varphi_\epsilon(v) := \mathbb{E}[\varphi(v\epsilon)]$, 其中 ϵ 服从 $\mathcal{N}_G(0, [1, 1])$, 则有 $\varphi_\epsilon \in C_{1,\text{Lip}}(\mathbb{R})$.

证明 注意到 $\epsilon \stackrel{d}{=} \mathcal{N}_G(0, [1, 1]) \stackrel{d}{=} N(0, 1)$, 即退化的 G - 正态分布等价于经典正态分布. 因此 $\varphi_\epsilon(v) := \hat{\mathbb{E}}[\varphi(v\epsilon)] = \mathbb{E}[\varphi(v\epsilon)]$. 接下来可以通过定义证明 $\varphi_\epsilon \in C_{1,\text{Lip}}(\mathbb{R})$:

$$\begin{aligned} |\varphi_\epsilon(x) - \varphi_\epsilon(y)| &= |\mathbb{E}[\varphi(x\epsilon) - \varphi(y\epsilon)]| \\ &\leq \mathbb{E}[C_\varphi(1 + |x\epsilon|^k + |y\epsilon|^k) |\epsilon| \cdot |x - y|] \\ &= C_\varphi(\mathbb{E}[|\epsilon|] + \mathbb{E}[|\epsilon|^{k+1}] |x|^k + \mathbb{E}[|\epsilon|^{k+1}] |y|^k) |x - y| \\ &\leq C(1 + |x|^k + |y|^k) |x - y|. \end{aligned} \quad \square$$

命题 3.1 的证明 可以直接通过 G - 独立性和最大分布的定义得到. 根据 $V \dashrightarrow \epsilon$ 的定义, 对于任意 $\varphi \in C_{1,\text{Lip}}(\mathbb{R})$, 有

$$\hat{\mathbb{E}}[\varphi(W)] = \hat{\mathbb{E}}[\varphi(V\epsilon)] = \hat{\mathbb{E}}[\hat{\mathbb{E}}[\varphi(v\epsilon)]_{v=V}] = \hat{\mathbb{E}}[\varphi_\epsilon(V)].$$

根据引理 A.1, 有 $\varphi_\epsilon \in C_{1,\text{Lip}}(\mathbb{R})$, 从而得到 $\hat{\mathbb{E}}[|\varphi(W)|] = \max_{v \in [\underline{\sigma}, \bar{\sigma}]} |\varphi_\epsilon(v)| < \infty$. 再运用最大分布的定义, 有

$$\hat{\mathbb{E}}[\varphi(W)] = \hat{\mathbb{E}}[\varphi_\epsilon(V)] = \max_{\sigma \in [\underline{\sigma}, \bar{\sigma}]} \varphi_\epsilon(\sigma) = \max_{\sigma \in [\underline{\sigma}, \bar{\sigma}]} \mathbb{E}[\varphi(\sigma\epsilon)].$$

因此, $W \sim \mathcal{N}_M(0, [\underline{\sigma}^2, \bar{\sigma}^2])$. $\square$

附录 A.2 第 5 节中的证明

为了准备证明, 考虑以下函数空间:

- (1) $C^k(\mathbb{R})$: 定义在 $\mathbb{R}$ 上的 k 次连续可微函数空间;
- (2) $C_b(\mathbb{R})$: 定义在 $\mathbb{R}$ 上的有界连续函数空间;
- (3) $C_l(\mathbb{R})$: 定义在 $\mathbb{R}$ 上的所有满足线性增长条件 $|\varphi(x)| \leq C(1 + |x|)$ 的连续函数所构成的空间;
- (4) $C^*(\mathbb{R}) = \{\varphi \in C^2(\mathbb{R}) : \varphi'' \text{ 是有界且一致连续的}\}$.

对于任意的 $\varphi \in C^*(\mathbb{R})$, 由于 φ'' 是有界的, 所以有 $M := \sup_{x \in \mathbb{R}} |\varphi''(x)| < \infty$.

下面主要讨论 $\varphi \in C^*(\mathbb{R})$. 为了扩展测试函数 φ 的空间, 我们准备了引理 A.3 和 A.2. 其中引理 A.2 是文献 [24, 引理 2.4.12] 在 $p = 2$ 时的一个特殊情形. 在以下讨论中, 考虑次线性期望空间中的一个序列 $\{Z_n\}_{n=1}^\infty$ 和 Z 的在测试函数 φ 下的收敛性问题:

$$\lim_{n \rightarrow \infty} \hat{\mathbb{E}}[\varphi(Z_n)] = \hat{\mathbb{E}}[\varphi(Z)]. \quad (\text{A.1})$$

引理 A.2 假设

$$\sup_n \hat{\mathbb{E}}[|Z_n|^2] + \hat{\mathbb{E}}[|Z|^2] < \infty$$

成立. 如果收敛性 (A.1) 对任意 $\varphi \in C_{b,\text{Lip}}(\mathbb{R})$ 成立, 则它也对任意 $\varphi \in C_l(\mathbb{R})$ 也成立.

引理 A.3 假设 $\sup_n \hat{\mathbb{E}}[|Z_n|] < \infty$ 且 $\hat{\mathbb{E}}[|Z|] < \infty$. 如果对任意 $\varphi \in C^*(\mathbb{R})$, 收敛性 (A.1) 成立, 则对 $\varphi \in C_b(\mathbb{R})$ 也成立.

引理 A.3 的证明 首先考虑有紧支集 S 的 $\varphi \in C_b(\mathbb{R})$. 通过文献 [27] 提供的一致逼近, 对于任意 $a > 0$, 存在一个在 S 上有支集的 $\varphi_a \in C^3(\mathbb{R})$ 使得

$$\sup_{x \in \mathbb{R}} |\varphi(x) - \varphi_a(x)| < \frac{a}{2}.$$

对于 $k = 1, 2, 3$, 由于 $\varphi_a^{(k)}$ 是连续的且定义在紧支集上, 所以它必然是有界的, 记为 M_k . 根据均值定理, 对于 $\delta > 0$ 和某个 $\beta \in [0, 1]$, 有 $|\varphi_a^{(2)}(x) - \varphi_a^{(2)}(x + \delta)| \leq |\varphi^{(3)}(x + \beta\delta)|\delta \leq M_3\delta$, 因此 $\varphi_a^{(2)}$ 是一致连续且有界的, 这意味着 $\varphi_a \in C^*(\mathbb{R})$. 由此得到

$$\begin{aligned} |\hat{\mathbb{E}}[\varphi(Z_n)] - \hat{\mathbb{E}}[\varphi(Z)]| &\leq |\hat{\mathbb{E}}[\varphi(Z_n)] - \hat{\mathbb{E}}[\varphi_a(Z_n)]| + |\hat{\mathbb{E}}[\varphi(Z)] - \hat{\mathbb{E}}[\varphi_a(Z)]| \\ &\quad + |\hat{\mathbb{E}}[\varphi_a(Z_n)] - \hat{\mathbb{E}}[\varphi_a(Z)]| \\ &\leq a + |\hat{\mathbb{E}}[\varphi_a(Z_n)] - \hat{\mathbb{E}}[\varphi_a(Z)]|. \end{aligned}$$

因此, $\limsup_{n \rightarrow \infty} |\hat{\mathbb{E}}[\varphi(Z_n)] - \hat{\mathbb{E}}[\varphi(Z)]| \leq a$. 这意味着

$$0 \leq \liminf_{n \rightarrow \infty} |\hat{\mathbb{E}}[\varphi(Z_n)] - \hat{\mathbb{E}}[\varphi(Z)]| \leq \limsup_{n \rightarrow \infty} |\hat{\mathbb{E}}[\varphi(Z_n)] - \hat{\mathbb{E}}[\varphi(Z)]| \leq a.$$

由于 a 可以任意小, 所以收敛性 (A.1) 成立.

接下来考虑任意的 $\varphi \in C_b(\mathbb{R})$, 它由 B 界定. 对于任意的 $K > 0$, 可以将 φ 分解为 $\varphi = \varphi_1 + \varphi_2$, 其中 $\varphi_1(x) := \varphi(x)\mathbb{1}_{\{|x| \leq K\}}$, $\varphi_2 := \varphi - \varphi_1$, 从而有

$$|\hat{\mathbb{E}}[\varphi(Z_n)] - \hat{\mathbb{E}}[\varphi(Z)]| \leq |\hat{\mathbb{E}}[\varphi_1(Z_n)] - \hat{\mathbb{E}}[\varphi_1(Z)]| + |\hat{\mathbb{E}}[\varphi_2(Z_n)] - \hat{\mathbb{E}}[\varphi_2(Z)]|,$$

其中第 1 项通过之前的论证必然收敛到 0. 只需考虑第 2 项, 注意到 φ_2 满足在 $[-K, K]$ 上取值为 0, 而当 $|x| > K$ 时, 有

$$|\varphi_2(x)| \leq B \leq \frac{B|x|}{K},$$

所以有

$$|\hat{\mathbb{E}}[\varphi_2(Z_n)] - \hat{\mathbb{E}}[\varphi_2(Z)]| \leq \frac{B}{K}(\hat{\mathbb{E}}[|Z_n|] + \hat{\mathbb{E}}[|Z|]).$$

注意到 $L := \sup_n \hat{\mathbb{E}}[|Z_n|] + \hat{\mathbb{E}}[|Z|] < \infty$, 从而有 $\limsup_{n \rightarrow \infty} |\hat{\mathbb{E}}[\varphi(Z_n)] - \hat{\mathbb{E}}[\varphi(Z)]| \leq \frac{BL}{K}$, 由于 K 可以取任意大, 因此得到收敛性 (A.1). $\square$

进而准备一个与 $\varphi \in C^*(\mathbb{R})$ 紧密相关的一个函数的性质 (引理 A.4).

引理 A.4 对于任意的 $\varphi \in C^*(\mathbb{R})$, 函数 $\delta: \mathbb{R}_+ \rightarrow \mathbb{R}_+$, 定义为

$$\delta(a) := \sup_{|x-y| \leq a} |\varphi''(x) - \varphi''(y)|,$$

必定是一个有界且递增的函数. 它还满足 $\lim_{a \downarrow 0} \delta(a) = 0$.

引理 A.4 的证明 δ 的有界性可以直接从 φ'' 的有界性得出. 其极限性质来自 φ'' 的一致连续性. 对于单调性, 对任意 $0 < a \leq b$, 由于 $\{(x, y) : |x - y| \leq a\} \subset \{(x, y) : |x - y| \leq b\}$, 所以有 $\delta(a) \leq \delta(b)$. $\square$

定理 5.1 的证明 只需要考虑任意测试函数 $\varphi \in C^*(\mathbb{R})$ 即可. 因为根据 M -独立性和 X_i 的性质, 有 $\hat{\mathbb{E}}[X_i X_j] = -\hat{\mathbb{E}}[-X_i X_j] = 0$. 令 $Z_n := \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$ 和 $Z := W$, 进而有

$$\hat{\mathbb{E}}[Z_n^2] = \frac{1}{n} \hat{\mathbb{E}} \left[\left(\sum_{i=1}^n X_i \right)^2 \right] = \frac{1}{n} \hat{\mathbb{E}} \left[\sum_{i=1}^n X_i^2 + \sum_{i \neq j} X_i X_j \right] = \frac{1}{n} \hat{\mathbb{E}} \left[\sum_{i=1}^n X_i^2 \right] = \sigma^2.$$

类似可得 $\sup_n \hat{\mathbb{E}}[|Z_n|] < \infty$. 从而根据引理 A.3, 满足收敛性 (A.1) 的函数空间可以扩展到 $C_b(\mathbb{R})$. 这个空间是包含 $C_{b, \text{Lip}}(\mathbb{R})$, 即依分布收敛所要求的函数空间. 进而根据引理 A.2, 测试函数空间可以扩展到 $C_l(\mathbb{R})$.

由于 $X_i \in \mathcal{H}_s$, $i = 1, 2, \dots, n$, 令 $X_i := V_i \eta_i$, 其中 V_i 服从最大分布, η_i 服从经典分布. 序列 $\{X_i\}_{i=1}^n$ 具有 M - 独立性, 这意味着

$$(V_1, \dots, V_n) \dashrightarrow (\eta_1, \dots, \eta_n).$$

然后考虑一系列经典独立同分布的 $\{\epsilon_i\}_{i=1}^n$, 它们满足 $\epsilon_1 \sim N(0, 1)$ 并且

$$(V_1, \dots, V_n) \dashrightarrow (\eta_1, \dots, \eta_n) \dashrightarrow (\epsilon_1, \dots, \epsilon_n).$$

对于每个 n , 定义一个三角阵列 $\{e_{i,n}\}_{i=1}^n$, 其中

$$e_{i,n} := \frac{X_i}{\sqrt{n}},$$

并设

$$S_n := e_{1,n} + \dots + e_{n,n}.$$

对于这个 n , 考虑另一个三角阵列 $\{W_{i,n}\}_{i=1}^n := \{(V_i \epsilon_i)/\sqrt{n}\}_{i=1}^n$, 它们是 M - 独立同分布的, 服从 M - 正态分布并满足

$$W_{i,n} \stackrel{d}{=} W_{1,n} \stackrel{d}{=} \frac{W}{\sqrt{n}}.$$

注意, 这里在 W_i 中使用了相同的 V_i 序列. 这种设置对于我们克服 $\hat{\mathbb{E}}$ 的次线性属性带来的困难很重要²⁴⁾. 令

$$W_n := W_{1,n} + \dots + W_{n,n},$$

通过 M - 正态分布的稳定性 (命题 3.4), 有 $W_n \sim \mathcal{N}_M(0, [\sigma^2, \bar{\sigma}^2])$.

我们的目标是, 展示对于任意 $\varphi \in C^*(\mathbb{R})$, 当 $n \rightarrow \infty$ 时, 有

$$|\hat{\mathbb{E}}[\varphi(S_n)] - \hat{\mathbb{E}}[\varphi(W)]| = |\hat{\mathbb{E}}[\varphi(S_n)] - \hat{\mathbb{E}}[\varphi(W_n)]| \rightarrow 0. \quad (\text{A.2})$$

考虑以下求和:

$$M_{i,n} = \sum_{j=1}^i e_{j,n} + \sum_{j=i+1}^n W_{j,n} \quad (\text{A.3})$$

和

$$U_{i,n} = \sum_{j=1}^{i-1} e_{j,n} + \sum_{j=i+1}^n W_{j,n}, \quad (\text{A.4})$$

其中空和定义为零. 注意 $M_{0,n} = W_n$ 和 $M_{n,n} = S_n$, 然后可以将 (A.2) 中的差转化为裂项求和

$$\begin{aligned} \hat{\mathbb{E}}[\varphi(S_n)] - \hat{\mathbb{E}}[\varphi(W_n)] &\leq \hat{\mathbb{E}}[\varphi(S_n) - \varphi(W_n)] \\ &= \hat{\mathbb{E}}\left[\sum_{i=1}^n (\varphi(M_{i,n}) - \varphi(M_{i-1,n}))\right] \\ &\leq \sum_{i=1}^n \hat{\mathbb{E}}[\varphi(M_{i,n}) - \varphi(M_{i-1,n})] \end{aligned} \quad (\text{A.5})$$

²⁴⁾ 这里可以看出经典中心极限定理中 σ^2 与次线性期望意义下的 V^2 的角色之间的一些联系.

以及

$$\hat{\mathbb{E}}[\varphi(W_n)] - \hat{\mathbb{E}}[\varphi(S_n)] \leq \sum_{j=1}^n \hat{\mathbb{E}}[\varphi(M_{n-j,n}) - \varphi(M_{n-j+1,n})]. \quad (\text{A.6})$$

因此, 只需要处理求和项 $\hat{\mathbb{E}}[\varphi(M_{i,n}) - \varphi(M_{i-1,n})]$. 通过 Taylor 展开 (其中 $\alpha, \beta \in [0, 1]$), 有

$$\begin{aligned} \varphi(M_{i,n}) - \varphi(M_{i-1,n}) &= \varphi(U_{i,n} + e_{i,n}) - \varphi(U_{i,n} + W_{i,n}) \\ &= (e_{i,n} - W_{i,n})\varphi'(U_{i,n}) \\ &\quad + \left[\frac{1}{2}e_{i,n}^2\varphi''(U_{i,n} + \alpha e_{i,n}) - \frac{1}{2}W_{i,n}^2\varphi''(U_{i,n} + \beta W_{i,n}) \right] \\ &=: (a) + (b). \end{aligned}$$

对于第 1 项 (a), 其次线性期望一定存在, 因为 φ' 的增长最多是线性的 (由于 φ'' 有界). 另外, $U_{i,n}$ 是 $V_u = (V_1, \dots, V_{i-1}, V_{i+1}, \dots, V_n)$ 和 $\xi_u = (\eta_1, \dots, \eta_{i-1}, \epsilon_{i+1}, \dots, \epsilon_n)$ 的内积, 有独立性 $V_u \dashrightarrow \xi_u$, 所以 $e_{i,n} - W_{i,n} (= n^{-1/2}V_i(\eta_i - \epsilon_i))$ 和 $U_{i,n}$ 是 M - 独立的. 那么

$$\begin{aligned} \hat{\mathbb{E}}[(e_{i,n} - W_{i,n})\varphi'(U_{i,n})] &= \max_{(v_i, v_u)} \mathbb{E}[n^{-1/2}v_i(\eta_i - \epsilon_i)\varphi'(v_u^T \xi_u)] \\ &= \max_{(v_i, v_u)} n^{-1/2}v_i \underbrace{\mathbb{E}[\eta_i - \epsilon_i]}_{=0} \mathbb{E}[\varphi'(v_u^T \xi_u)] \\ &= 0, \end{aligned}$$

其中第二个等式是由于经典独立性. 类似地, 有 $-\hat{\mathbb{E}}[-(e_{i,n} - W_{i,n})\varphi'(U_{i,n})] = 0$. 因此, (a) 有确定的零均值. 所以

$$\hat{\mathbb{E}}[\varphi(M_{i,n}) - \varphi(M_{i-1,n})] = \hat{\mathbb{E}}[(b)].$$

对于第 2 项 (b), 注意到

$$\begin{aligned} 2 \times (b) &= e_{i,n}^2[\varphi''(U_{i,n} + \alpha e_{i,n}) - \varphi''(U_{i,n})] - W_{i,n}^2[\varphi''(U_{i,n} + \beta W_{i,n}) - \varphi''(U_{i,n})] \\ &\quad + (e_{i,n}^2 - W_{i,n}^2)\varphi''(U_{i,n}) \\ &=: (b)_1 - (b)_2 + (b)_3. \end{aligned}$$

对于 (b)₁, 由 $|\alpha e_{i,n}| \leq |e_{i,n}|$, 通过回顾 $\delta(\cdot)$ 的性质 (引理 A.4), 又因为 $e_{i,n} = \frac{X_i}{\sqrt{n}}$ 和 $X_i \stackrel{d}{=} X_1$, 从而有

$$\hat{\mathbb{E}}[(b)_1] \leq \hat{\mathbb{E}}[e_{i,n}^2 \delta(|e_{i,n}|)] = \frac{1}{n} \hat{\mathbb{E}}[X_1^2 \delta(n^{-1/2} |X_1|)].$$

对于 (b)₂, 因为 $W_{i,n} \stackrel{d}{=} \frac{W}{\sqrt{n}}$, 所以可得

$$\hat{\mathbb{E}}[(b)_2] \leq \hat{\mathbb{E}}[W_{i,n}^2 \delta(|W_{i,n}|)] = \frac{1}{n} \hat{\mathbb{E}}[W^2 \delta(n^{-1/2} |W|)].$$

对于 (b)₃, 因为 $(e_{i,n}, W_{i,n})$ 和 $U_{i,n}$ 是 M - 独立的 (注意 $e_{i,n}$ 和 $W_{i,n}$ 依赖于同一个 V_i), 所以

$$\hat{\mathbb{E}}[(b)_3] = \max_{(v_i, v_u)} \mathbb{E}[n^{-1}v_i^2(\eta_i^2 - \epsilon_i^2)\varphi''(v_u^T \xi_u)]$$

$$\begin{aligned}
 &= \max_{(v_i, v_u)} n^{-1} v_i^2 \underbrace{\mathbb{E}[\eta_i^2 - \epsilon_i^2]}_{=0} \mathbb{E}[\varphi''(v_u^T \xi_u)] \\
 &= 0,
 \end{aligned}$$

其中运用了经典独立性和 $\mathbb{E}[\eta_i^2] = \mathbb{E}[\epsilon_i^2] = 1$. 由类似逻辑可得 $-\hat{\mathbb{E}}[-(b)_3] = 0$. 因此 $(b)_3$ 有确定的零均值. 综上可得

$$\begin{aligned}
 \hat{\mathbb{E}}[\varphi(M_{i,n}) - \varphi(M_{i-1,n})] &= \frac{1}{2} \hat{\mathbb{E}}[(b)_1 - (b)_2] \\
 &\leq \frac{1}{2} (\hat{\mathbb{E}}[|b|_1] + \hat{\mathbb{E}}[|b|_2]) \\
 &= \frac{1}{2n} (\hat{\mathbb{E}}[X_1^2 \delta(n^{-1/2} |X_1|)] + \hat{\mathbb{E}}[W^2 \delta(n^{-1/2} |W|)]).
 \end{aligned}$$

同时, 如果交换 $\varphi(M_{i,n})$ 和 $\varphi(M_{i-1,n})$ 的角色, 并令 $i = n - j + 1$, 其中 $j = 1, 2, \dots, n$, 得到

$$\begin{aligned}
 \hat{\mathbb{E}}[\varphi(M_{n-j,n}) - \varphi(M_{n-j+1,n})] &= \hat{\mathbb{E}}[\varphi(M_{i-1,n}) - \varphi(M_{i,n})] \\
 &= \frac{1}{2} \hat{\mathbb{E}}[(b)_2 - (b)_1] \\
 &\leq \frac{1}{2} (\hat{\mathbb{E}}[|b|_2] + \hat{\mathbb{E}}[|b|_1]) \\
 &= \frac{1}{2n} (\hat{\mathbb{E}}[X_1^2 \delta(n^{-1/2} |X_1|)] + \hat{\mathbb{E}}[W^2 \delta(n^{-1/2} |W|)]).
 \end{aligned}$$

因此, 通过 (A.5) 和 (A.6), 有

$$\begin{aligned}
 |\hat{\mathbb{E}}[\varphi(S_n)] - \hat{\mathbb{E}}[\varphi(W)]| &= |\hat{\mathbb{E}}[\varphi(S_n)] - \hat{\mathbb{E}}[\varphi(W_n)]| \\
 &= \max\{\hat{\mathbb{E}}[\varphi(S_n)] - \hat{\mathbb{E}}[\varphi(W_n)], \hat{\mathbb{E}}[\varphi(W_n)] - \hat{\mathbb{E}}[\varphi(S_n)]\} \\
 &\leq \max\left\{\sum_{i=1}^n \hat{\mathbb{E}}[\varphi(M_{i,n}) - \varphi(M_{i-1,n})], \sum_{j=1}^n \hat{\mathbb{E}}[\varphi(M_{n-j,n}) - \varphi(M_{n-j+1,n})]\right\} \\
 &\leq \sum_{i=1}^n \frac{1}{2n} (\hat{\mathbb{E}}[X_1^2 \delta(n^{-1/2} |X_1|)] + \hat{\mathbb{E}}[W^2 \delta(n^{-1/2} |W|)]) \\
 &= \frac{1}{2} (\hat{\mathbb{E}}[X_1^2 \delta(n^{-1/2} |X_1|)] + \hat{\mathbb{E}}[W^2 \delta(n^{-1/2} |W|)]).
 \end{aligned}$$

注意对于任意的 $v_1 \in [\underline{\sigma}, \bar{\sigma}]$, $|v_1 \eta_1| \leq \bar{\sigma} |\eta_1|$, 有

$$\begin{aligned}
 \hat{\mathbb{E}}[X_1^2 \delta(n^{-1/2} |X_1|)] &= \max_{v_1 \in [\underline{\sigma}, \bar{\sigma}]} \mathbb{E}[v_1^2 \eta_1^2 \delta(n^{-1/2} |v_1 \eta_1|)] \\
 &\leq \max_{v_1 \in [\underline{\sigma}, \bar{\sigma}]} v_1^2 \mathbb{E}[\eta_1^2 \delta(n^{-1/2} \bar{\sigma} |\eta_1|)] \\
 &= \bar{\sigma}^2 \mathbb{E}[\eta_1^2 \delta(n^{-1/2} \bar{\sigma} |\eta_1|)],
 \end{aligned}$$

其中 $\leq$ 来自 δ 的单调性. 由引理 A.4 可推导出, 对所有 $a \in \mathbb{R}^+$, 都有 $\delta(a) \leq 2M$. 所以

$$\eta_1^2 \delta(n^{-1/2} \bar{\sigma} |\eta_1|) \leq 2M \eta_1^2.$$

同时, $\eta_1^2 \delta(n^{-1/2} \bar{\sigma} |\eta_1|)$ 在 $n \rightarrow \infty$ 时逐点收敛到零, 则通过经典的控制收敛定理, 可推导出

$$\mathbb{E}[\eta_1^2 \delta(n^{-1/2} \bar{\sigma} |\eta_1|)] \rightarrow 0.$$

这意味着 $\hat{\mathbb{E}}[X_1^2 \delta(n^{-1/2} |X_1|)] \rightarrow 0$. 类似可得 $\hat{\mathbb{E}}[W^2 \delta(n^{-1/2} |W|)] \rightarrow 0$. 最终, 有

$$|\hat{\mathbb{E}}[\varphi(S_n)] - \hat{\mathbb{E}}[\varphi(W)]| \rightarrow 0,$$

或者

$$\lim_{n \rightarrow \infty} \hat{\mathbb{E}} \left[\varphi \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \right) \right] = \hat{\mathbb{E}}[\varphi(W)]. \quad \square$$

推论 5.1 的证明 证明过程可以形式化为 $(1) \Rightarrow (3) \Rightarrow (2) \Rightarrow (1)$. 首先显然有 $(2) \Rightarrow (1)$. 对于 $(3) \Rightarrow (2)$, 这是命题 3.4 的直接结论. 为了验证 $(1) \Rightarrow (3)$, 首先声明, 对于任意 M -i.i.d. 序列 $\{W_i\}_{i=1}^{2^m} \subset \mathcal{H}_s$, $m \in \mathbb{N}_+$ 且 $W_1 := W$, 有

$$\sum_{i=1}^{2^m} W_i \stackrel{d}{=} 2^{m/2} W_1.$$

这个声明可以通过数学归纳法证明, 并使用 (1), 然后有

$$W_1 \stackrel{d}{=} \frac{1}{\sqrt{2^m}} \sum_{i=1}^{2^m} W_i.$$

令 $n = 2^m$, 则有 $\hat{\mathbb{E}}[\varphi(W)] = \hat{\mathbb{E}}[\varphi(\frac{1}{\sqrt{n}} \sum_{i=1}^n W_i)]$. 注意 W_i 是 M -i.i.d. 的, 所以可以使用 M -中心极限定理. 两边同时令 $n \rightarrow \infty$, 则有 $\hat{\mathbb{E}}[\varphi(W)] = \sup_{\sigma \in [\underline{\sigma}, \bar{\sigma}]} \mathbb{E}[\varphi(\sigma \epsilon)]$. 因此, W 服从 M -正态分布. $\square$

附录 B M -分布族的更多理论细节

附录 B.1 乘积空间

本节为定义 4.1 和 4.2 提到的次线性期望空间给出一个具体的构造方式. 考虑一族概率测度 $\mathcal{Q}$ 和一个概率测度 P , 其中 P 不必在 $\mathcal{Q}$ 中. 设 $\hat{\mathbb{E}}_1[\cdot] := \sup_{Q \in \mathcal{Q}} \mathbb{E}_Q[\cdot]$ 和 $\hat{\mathbb{E}}_2[\cdot] := \mathbb{E}_P[\cdot]$, 进而引入两个次线性期望空间²⁵⁾ $(\Omega_1, \mathcal{H}_1, \hat{\mathbb{E}}_1)$ 和 $(\Omega_2, \mathcal{H}_2, \hat{\mathbb{E}}_2)$, 其中 $\hat{\mathbb{E}}_2$ 可以被视为一个退化的次线性期望. 定义一个乘积空间 $(\Omega, \mathcal{H}, \hat{\mathbb{E}}) := (\Omega_1 \times \Omega_2, \mathcal{H}_1 \otimes \mathcal{H}_2, \hat{\mathbb{E}}_1 \otimes \hat{\mathbb{E}}_2)$. 该空间也是一个次线性期望空间 (详见文献 [24, 第 1.3 小节]), 本节仅给出必要细节以确保讨论的自洽性.

(1) 空间 $\mathcal{H}_1 \otimes \mathcal{H}_2$ 定义为

$$\begin{aligned} \mathcal{H}_1 \otimes \mathcal{H}_2 &= \{X(\omega_1, \omega_2) = f(K(\omega_1), \eta(\omega_2)), (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2, \\ &\quad K \in \mathcal{H}_1^m, \eta \in \mathcal{H}_2^n, f \in C_{1,\text{Lip}}(\mathbb{R}^{m+n})\}, \end{aligned}$$

其中 η 可以被视为一个经典随机变量, 因为 $\hat{\mathbb{E}}_2 := \mathbb{E}_P$ 是线性期望; f 的函数空间²⁶⁾ 可以根据实践中的需要适当放宽 (例如, 考虑 Borel 可测函数).

(2) 对于 $X(\omega_1, \omega_2) = f(K(\omega_1), \eta(\omega_2)) \in \mathcal{H}$, 令

$$\begin{aligned} \hat{\mathbb{E}}[X] &= \hat{\mathbb{E}}_1 \otimes \hat{\mathbb{E}}_2[X] := \hat{\mathbb{E}}_1[\hat{\mathbb{E}}_2[f(k, \eta)]_{k=K}] = \sup_{Q \in \mathcal{Q}} \mathbb{E}_Q[\mathbb{E}_P[f(k, \eta)]_{k=K}] \\ &= \sup_{Q \in \mathcal{Q}} \int \int f(k, y) P_\eta(dy) Q_K(dk) = \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[X], \end{aligned}$$

其中 $\mathcal{P} := \{Q \otimes P, Q \in \mathcal{Q}\}$, 这里 $Q \otimes P$ 是 P 和 Q 的乘积测度. 注意 $\hat{\mathbb{E}}_2 \otimes \hat{\mathbb{E}}_1 \neq \hat{\mathbb{E}}_1 \otimes \hat{\mathbb{E}}_2$.

²⁵⁾ 关于如何从一族概率测度出发来构造次线性期望空间, 参见第 2 节.

²⁶⁾ 关于次线性期望空间中的函数空间的可能选择, 详见文献 [24, 第 5 页].

命题 B.1 考虑在 $(\Omega_1, \mathcal{H}_1, \hat{\mathbb{E}}_1)$ 上的随机向量 K 和在 $(\Omega_2, \mathcal{H}_2, \hat{\mathbb{E}}_2)$ 上的 η , 通过令 $\bar{K}(\omega_1, \omega_2) := K(\omega_1)$ 和 $\bar{\eta}(\omega_1, \omega_2) := \eta(\omega_2)$, 有以下结果:

- (1) $\bar{K}, \bar{\eta} \in \mathcal{H}_1 \otimes \mathcal{H}_2$;
- (2) 对于任意 $\omega \in \Omega_1 \times \Omega_2$, 有 $X(\omega) = f(\bar{K}(\omega), \bar{\eta}(\omega))$;
- (3) 对于任意 $\varphi \in C_{b, \text{Lip}}$, 有 $\hat{\mathbb{E}}[\varphi(\bar{K})] = \hat{\mathbb{E}}_1[\varphi(K)]$;
- (4) 对于任意 $\varphi \in C_{b, \text{Lip}}$, 有 $\hat{\mathbb{E}}[\varphi(\bar{\eta})] = \mathbb{E}_P[\varphi(\eta)]$;
- (5) $\bar{K} \dashrightarrow \bar{\eta}$.

命题 B.1 的证明 第 1 项来源于

$$\bar{K}(\omega_1, \omega_2) = f_1(K(\omega_1), \eta(\omega_2)), \quad \bar{\eta}(\omega_1, \omega_2) = f_2(K(\omega_1), \eta(\omega_2)),$$

其中 $f_1(x, y) = x$ 和 $f_2(x, y) = y$. 第 2 项可以从以下看出:

$$X(\omega_1, \omega_2) = f(K(\omega_1), \eta(\omega_2)) = f(\bar{K}(\omega), \bar{\eta}(\omega)).$$

对于第 3 项, 有

$$\begin{aligned} \hat{\mathbb{E}}[\varphi(\bar{K})] &= \hat{\mathbb{E}}_1[\hat{\mathbb{E}}_2[\varphi(f_1(k, \eta))]_{k=\bar{K}}] \\ &= \hat{\mathbb{E}}_1[\hat{\mathbb{E}}_2[\varphi(k)]_{k=K}] = \hat{\mathbb{E}}_1[\varphi(K)]. \end{aligned}$$

类似可得第 4 项.

接下来验证第 5 项, 即 $\hat{\mathbb{E}}[\varphi(\bar{K}, \bar{\eta})] = \hat{\mathbb{E}}[\hat{\mathbb{E}}[\varphi(k, \bar{\eta})]_{k=\bar{K}}]$. 令 $H(k) := \mathbb{E}_P[\varphi(k, \bar{\eta})]$, 从而有

$$\begin{aligned} \text{RHS} &= \hat{\mathbb{E}}[\mathbb{E}_P[\varphi(k, \bar{\eta})]_{k=\bar{K}}] \\ &= \hat{\mathbb{E}}[H(\bar{K})] = \hat{\mathbb{E}}_1[H(K)] \\ &= \hat{\mathbb{E}}_1[\mathbb{E}_P[\varphi(k, \bar{\eta})]_{k=K}] \\ &= \hat{\mathbb{E}}[\varphi(K, \bar{\eta})] = \hat{\mathbb{E}}[\varphi(\bar{K}, \bar{\eta})]. \end{aligned} \quad \square$$

注 B.1 除非另有说明, 我们将不区分 $\bar{K}$ (或 $\bar{\eta}$) 和 K (或 η). 此外, 令 $\eta(\omega_1, \omega_2) := \eta(\omega_2)$ 使得 $\eta \in \mathcal{H}_1 \otimes \mathcal{H}_2$, 从第 4 项, 有

$$\hat{\mathbb{E}}[\varphi(\eta)] = \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[\varphi(\eta)] = \mathbb{E}_P[\varphi(\eta)],$$

其中 $\mathbb{P}$ 是任意乘积测度 $Q \otimes P$, $Q \in \mathcal{Q}$. 通过将 φ 改为 $-\varphi$, 可以展示对于任意 $\mathbb{P} \in \mathcal{P}$, 有

$$\mathbb{E}_P[\varphi(\eta)] = \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[\varphi(\eta)] \leq \mathbb{E}_{\mathbb{P}}[\varphi(\eta)] \leq \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[\varphi(\eta)] = \mathbb{E}_P[\varphi(\eta)],$$

或简单地 $\mathbb{E}_{\mathbb{P}}[\varphi(\eta)] = \mathbb{E}_P[\varphi(\eta)]$. 这意味着在每个乘积测度 $\mathbb{P} \in \mathcal{P}$ 下, 经典 η 的概率分布始终是 P_{η} . 设

$$\bar{\mathcal{H}}_s := \{X \in \mathcal{H} : X(\omega_1, \omega_2) = f(K(\omega_1), \eta(\omega_2)), K \in \mathcal{H}_1^m, K \sim \mathcal{M}(\Theta)\}$$

表示 $\mathcal{H}$ 的一个子空间, 则对于任意 $X \in \bar{\mathcal{H}}_s$, 有

$$\hat{\mathbb{E}}[X] = \sup_{\theta \in \Theta} \mathbb{E}_P[f(\theta, \eta)].$$

注 B.2 除了考虑子空间, 另一种更加直接的构造方式是令

$$\mathcal{Q} := \{\delta_v : v \in \Theta\},$$

其中 δ_v 是 Dirac 测度. 考虑 $K(\omega) := \omega$, 从而 K 在 $\hat{\mathbb{E}}_1$ 下自然服从最大分布 $\mathcal{M}(\Theta)$.

附录 B.2 满足 M -独立性的随机变量的存在性

关于 k 个满足 M -独立性的随机变量的存在性, 注意到命题 B.1 中的 $\bar{K}$ 和 $\bar{\eta}$ 都是多维的, 设 $\bar{K} = (K_i)_{i=1}^k$ 为由 k 个 G -独立的服从最大分布的随机变量组成的向量, $\bar{\eta} = (\eta_i)_{i=1}^k$ 为服从经典的 $N(0, I_k)$ 的随机向量 (其中 I_k 为 k 阶单位矩阵), 从而 η_i 之间满足经典的独立性²⁷⁾. 对于 $i = 1, 2, \dots, k$, 令 $X_i := f(K_i, \eta_i)$. 从而 $\{X_i\}_{i=1}^k$ 满足 M -独立性 (定义 4.4).

为了进一步考虑可列个满足 M -独立性的随机变量的存在性, 只需要将 $(\Omega_1, \mathcal{H}_1, \hat{\mathbb{E}}_1)$ 和 $(\Omega_2, \mathcal{H}_2, \hat{\mathbb{E}}_2)$ 分别设定为无穷维的乘积空间即可 (更多细节参见文献 [24, 注 1.3.18 和第 6.2 小节]). 也可以直接由可列个满足 G -独立性的随机变量²⁸⁾ 来构造可列个满足 M -独立性的序列. 下面给出一种简洁的方式.

令 $\mathbb{Z}$ 表示所有整数的集合, $\mathbb{N}_+$ 为所有正整数的集合. 不失一般性, 考虑一系列满足 G -i.i.d. 的随机变量 $\{Z_t\}_{t \in \mathbb{Z}}$, 其中

$$Z_t = (K_t^*, \eta_t^*), \quad K_t^* \sim \mathcal{M}(\Theta), \quad \eta_t^* \sim \mathcal{N}_G(0, I_n) \stackrel{d}{=} N(0, I_n).$$

序列 $\{Z_t\}_{t \in \mathbb{Z}}$ 的 G -独立性意味着, 对任意 $N \in \mathbb{N}_+$, 有

$$Z_{-N} \dashrightarrow Z_{-(N-1)} \dashrightarrow \cdots \dashrightarrow Z_{-1} \dashrightarrow Z_0 \dashrightarrow Z_1 \dashrightarrow \cdots \dashrightarrow Z_{N-1} \dashrightarrow Z_N.$$

令

$$p(x, y) := x, \quad q(x, y) := y,$$

并对任意 $i \in \mathbb{N}_+$, 定义

$$K_i := p(Z_{-i}), \quad \eta_i := q(Z_i).$$

由此可得如下结论.

(1) 由于 $Z_{-N} \dashrightarrow \cdots \dashrightarrow Z_{-1}$, 故 $K_N \dashrightarrow \cdots \dashrightarrow K_1$, 且每个 K_i 均服从最大分布, 因此 $\{K_i\}_{i=1}^N$ 之间的 G -独立性是对称的 (见第 2.1 小节).

(2) 由于 $Z_1 \dashrightarrow \cdots \dashrightarrow Z_N$, 故 $\eta_1 \dashrightarrow \cdots \dashrightarrow \eta_N$, 且每个 η_i 均为经典分布, 因此 $\{\eta_i\}_{i=1}^N$ 之间的 G -独立性退化为对称的经典独立.

(3) 由 $(Z_{-N}, \dots, Z_{-1}) \dashrightarrow (Z_1, \dots, Z_N)$, 可得 $(K_N, \dots, K_1) \dashrightarrow (\eta_1, \dots, \eta_N)$, 即 $(\eta_i)_{i=1}^N$ (G -) 独立于 $(K_i)_{i=1}^N$.

最后令

$$X_i := f(K_i, \eta_i), \quad i \in \mathbb{N}_+,$$

由于 $K_i \dashrightarrow \eta_i$, 从而 X_i 的分布属于 M -分布族, 具体而言, $\{X_i\}_{i=1}^\infty \subset \bar{\mathcal{H}}_s$. 并且根据上述讨论, 序列 $\{X_i\}_{i=1}^\infty$ 满足 M -独立性.

27) 这里考虑 $N(0, I_k)$ 只是出于讨论的简便, 当然也可将 $\bar{\eta}$ 换成由一系列满足经典独立性的随机变量所构成的向量.

28) 具体而言, 是服从 G -分布 (参见文献 [24, 定义 2.2.8]) 的一系列下标为所有整数的 G -独立的随机变量. 它的存在性来自于无穷维乘积空间, 例如, 将文献 [24, 注 1.3.18] 中的下标换成从 $i = -k$ 到 $i = k$ 即可.

M -structure: A nonlinear probabilistic structure with symmetric independence under model uncertainty

Yifan Li, Hao Yu & Reg Kulperger

Abstract Model uncertainty is a long-existing fundamental concern in various areas. The nonlinear expectation framework offers a profound theoretical view for discussing model uncertainty, presenting a non-trivial generalization of the classical probability system. However, when discussing the connection with real-world problems, we need to be more cautious about the interpretation of many basic concepts under nonlinear expectation, such as normal distribution and independence. Our initial motivation to introduce the M -structure is to have a structure that preserves statistical flexibility, so that it can serve as a bridge to make nonlinear expectation more accessible to a broader audience and also provide a fresh perspective on model uncertainty. The flexibility of this structure can be elaborated as follows. Regarding independence, M -independence preserves symmetry, capturing model uncertainty with spatial or temporal symmetry, and is related to G -independence, which primarily addresses model uncertainty with temporal asymmetry. We establish a central limit theorem under the M -structure, describing the asymptotic distribution under spatiotemporal symmetry, which is the M -normal distribution. Notably, the M -normal distribution preserves a direct link between univariate and multivariate distributions under independence, a property lacking in the G -normal distribution. More importantly, the M -structure uncovers the implicit connection and the source of distinctions between concepts under classical and nonlinear expectations. This foundational insight brings a far-reaching advantage for future developments in both theoretical and practical contexts.

Keywords nonlinear expectation, model uncertainty, central limit theorem, independence, normal distribution, symmetry, exchangeability

MSC(2020) 60A05, 60E05, 60E07, 60F05, 62H05

doi: 10.1360/SSM-2024-0328