# 主导测度、编码与信息

Yifeng Chu 和 Maxim Raginsky

arXiv:2305.02960v2 [cs.IT] 6 May 2023

摘要——Fernique 和 Talagrand 的主导测度定理是随机过程理论中的一项基本结果。该定理将由度量空间元素索引的随机过程的有界性，与源自某些多尺度组合结构（如打包树和覆盖树）的复杂度度量联系起来。本文基于 Andreas Maurer 一篇鲜为人知的预印本中最初提出的思想，从信息论视角阐述主导测度定理；在这一视角下，随机过程的有界性被表述为：索引度量空间中的元素存在高效变长编码。

## I. 引言

设 $(T,d)$ 是一个紧度量空间，并设 $(X_t)_{t\in T}$ 是一个以 $T$ 的元素为索引且满足增量条件的零均值随机过程

$$\mathbf{P}\Big[|X_s - X_t| \geq u d(s,t)\Big] \leq 2e^{-u^2}, \quad s,t \in T,\, u > 0. \tag{1}$$

期望上确界 $\mathrm{E}[\sup_{t\in T} X_t]$ 是否有界这一问题，在许多应用中都会出现，例如高维概率、凸几何、泛函分析、理论计算机科学、数理统计和学习理论 [1]。Fernique [2] 和 Talagrand [3] 的一项基本结果指出，对于任意固定但任取的 $t_0 \in T$，

$$\mathbf{E}\Big[\sup_{t\in T} |X_t - X_{t_0}|\Big] \lesssim \inf_{\mu\in\mathcal{P}(T)} \sup_{t\in T} I_\mu(t), \tag{2}$$

其中 $\mathcal{P}(T)$ 是 $T$ 上所有博雷尔概率测度的集合，

$$I_\mu(t) := \int_0^{\mathrm{diam}(T)} \sqrt{\log \frac{1}{\mu(B(t,\varepsilon))}}\, \mathrm{d}\varepsilon \tag{3}$$

是 Fernique–Talagrand 泛函，而 $a \lesssim b$ 表示对于某个绝对常数 $C > 0$，$a \leq Cb$。这里，$\mathrm{diam}(T) := \sup\{d(s,t) : s,t \in T\}$ 是 $T$的（有限）直径，$B(t,\varepsilon)$表示以$t$为中心、半径为 $\varepsilon$的闭球，且对数以 2 为底。此外，如果 $(X_t)_{t\in T}$是一个高斯过程，并且 $T$ 赋予典范度量

$$d(s,t) := \sqrt{\mathbf{E}[(X_s - X_t)^2]}, \tag{4}$$

那么，界限 (2) 可以反向成立 [3]:

$$\mathbf{E}\Big[\sup_{t\in T} |X_t - X_{t_0}|\Big] \gtrsim \inf_{\mu\in\mathcal{P}(T)} \sup_{t\in T} I_\mu(t). \tag{5}$$

作者现供职于美国伊利诺伊大学厄巴纳分校电气与计算机工程系及协同科学实验室，地址为美国伊利诺伊州厄巴纳市 61801。电子邮箱：ychu26@illinois.edu, maxim@illinois.edu。

本研究由伊利诺伊数据科学与动力系统研究所（iDS²）资助。该研究所为美国国家科学基金会 HDRTRIPODS 研究所，资助编号为 CCF-1934986。

不等式 (2) 表明，只要存在至少一个 $\mu \in \mathcal{P}(T)$，使得 $\sup_{t\in T} I_\mu(t) < \infty$ （此类概率测度称为主导测度），则 $(X_t)$ 的期望上确界是有界的；此外，对于高斯过程，主导测度的存在等价于期望上确界有界。

尽管 Fernique–Talagrand 泛函为 $(X_t)$ 的期望上确界提供了紧确界，但主导测度的理论工具最终被一种等价的组合框架所取代，该框架涉及各种多尺度离散结构，例如覆盖树和打包树（参见例如 [4]或 [1, 第 6] 章）。在一篇鲜为人知的预印本 [5], 中，A. Maurer 从一个观察出发，为主导测度提供了另一种信息论视角：式 (3) 中的被积函数带有信息论意味——事实上，$\log \frac{1}{\mu(B(t,\varepsilon))}$ 大致正比于在给定一个“先验”概率测度 $\mu \in \mathcal{P}(T)$ 的情况下，将点 $t \in T$ 定位到半径为$\varepsilon$ 的球内所需的比特数。基于这一思路，Maurer 利用了离散（有限或可数无限）字母表[6, 上的概率测度与唯一可译码之间众所周知的对应关系（第 5 节），5] 证明 Fernique–Talagrand 泛函可以用关于 $T$ 上一系列变长有损编码的信息论量来给出上界和下界；这些编码允许通过足够长的码字，以任意精度重构点 $t \in T$，其中度量 $d$ 扮演保真度准则 [7]的角色。

本文通过更直接地构造多尺度变长编码，进一步拓展了这一视角，使我们能够无缝地恢复若干关于期望上确界的已有界。此外，我们推导出高斯过程下界（5）的信息论版本，其与我们的上界相差一个乘法常数。在应用方面，我们预计这些结果有助于从数据压缩和最小描述长度原理 [8], [9]的角度分析机器学习算法的统计性能。

## II. 预备知识

### A. 度量空间上的变长编码

设有一个紧致度量空间 $(T,d)$，其直径为 $= 1 T$。定义在 $T$ 上的变长码（VLC）是一个二元组 $(\pi, f)$，其中 $\pi : T \to T$ 是一个映射，其像 $\pi(T)$ 是 $T$ 的离散子集；而 $f : \pi(T) \to \{0,1\}^*$ 是对 $\pi(T)$ 中元素的一种唯一可译表示，其编码为

有限长度二进制字符串。特别地，它满足 Kraft–McMillan 不等式

$$\sum_{s\in\pi(T)} 2^{-\ell(f(s))} \le 1, \tag{6}$$

其中，$\ell(\cdot)$ 表示二进制字符串的长度。我们说，VLC $(\pi, f)$ 以分辨率 $\rho \ge 0$ 运行，当且仅当对于所有 $t \in T$ 都有 $d(t, \pi(t)) \le \rho$（如果 $\rho = 0$，那么 $T$ 必然是离散的，并且 $(\pi, f)$ 是无损的）。构造此类 VLC 的一种自然方法是固定一个 $\rho$-划分 $\mathcal{A}$ of $T$，即由 $A \subseteq T$ 组成的互不相交子集族，其直径至多为 $\rho$，且并集为整个 $T$；如果 $\rho > 0$，则由 $T$的紧致性，总能找到一个有限划分。然后，在每个 $A \in \mathcal{A}$中选取一个点 $s_A$，取 $\pi(t) := s_{A(t)}$，其中 $A(t)$ 是包含 $t$的 $\mathcal{A}$ 中的唯一元素。以这种方式构造的映射 $\pi$ 显然是幂等的，即 $\pi \circ \pi = \pi$。最后，令 $\pi(T) = \{s_A : A \in \mathcal{A}\}$，我们可以选择任意满足 (6) 的映射 $f : \pi(T) \to \{0,1\}^*$。例如，我们可以选取任意 $\mu \in \mathcal{P}(T)$，使得对于所有 $A \in \mathcal{A}$ 都有 $\mu(A) > 0$，并令 $f$ 为具有码长 $\ell(f(s_A)) = \lceil \log \mu^{-1}(A) \rceil$，$A \in \mathcal{A}$的香农码。

---

我们将使用在多种尺度上运行的变长码序列。若满足以下条件，我们称 $\mathcal{C} = \{(\pi_k, f_k)\}_{k\in\mathbb{N}}$ 为可容许：

(i) 对每个 $k \in \mathbb{N}$，$\pi_k$ 都是幂等的，并且存在映射 $\varphi_{k|k+1} : T \to T$，使得 $\pi_k = \varphi_{k|k+1} \circ \pi_{k+1}$；(ii) $(\pi_k, f_k)$ 的分辨率为 $\rho_k$，并且 $\rho_k \to 0$ 如同 $k \to \infty$；(iii) 对每个 $t$ 以及每个 $k$，$\ell(f_k \circ \pi_k(t)) \le \ell(f_{k+1} \circ \pi_{k+1}(t))$。

我们还定义 $(\pi_0, f_0)$，其中 $\pi_0(T) = \{t_0\}$，且对于一个固定但任意的点 $t_0 \in T$，$\ell(f_0(t_0)) = 0$。条件 (i) 表示，对于每个 $t$，编码 $\pi_k(t)$ 可由任意 $\pi_m(t)$ $m > k$ 确定，而无需知道 $t_0$。不失一般性，我们可以取 $\varphi_{k|k+1} = \pi_k$，因为 $\pi_k = \varphi_{k|k+1} \circ \pi_{k+1} = \varphi_{k|k+1} \circ \pi_{k+1} \circ \pi_{k+1} = \pi_k \circ \pi_{k+1}$。这意味着由 $\pi_{k+1}$ 诱导的划分 $\mathcal{A}_{k+1}$ 是 $\mathcal{A}_k$ 的细化，并且对于每个 $t \in T$，集合 $A_k(t)$（直径为 $\rho_k$）是嵌套的，即 $T \equiv A_0(t) \supseteq A_1(t) \supseteq A_2(t) \supseteq \dots$。条件 (i) 和 (ii) 意味着 $\rho_k \searrow 0$ 随着 $k \to \infty$ 而变化。最后，条件 (iii) 限制每个 $t$ 的二进制描述，使其随着 $k$ 增大而变得更加“富有信息量”。

构造一列可容许的 VLC 的一种自然方法是，首先取一列递增的划分 $(\mathcal{A}_k)_{k\ge1}$，使得 $T$ 的每个 $A \in \mathcal{A}_k$ 的直径至多为 $\rho_k$，其中 $\rho_k \searrow 0$ 为 $k \to \infty$。映射 $\pi_k$ 的构造方式与单个划分的情形完全相同，因此条件 (i)–(ii) 将自动满足。为构造满足条件 (iii) 的二进制编码 $f_k$，选取某个 $\mu_1 \in \mathcal{P}(T)$，使得对所有 $A \in \mathcal{A}_1$，$\mu_1(A) > 0$；然后，对每个 $k \ge 1$ 和每个 $B \in \mathcal{A}_k$，选取一个概率测度 $\nu_{k+1}(\cdot|B) \in \mathcal{P}(T)$，使得对每个 $A \in$，$\nu_{k+1}(A|B) > 0$ $\mathcal{A}_{k+1}(B) := \{A \in \mathcal{A}_{k+1} : A \subseteq B\}$，其他情况则为 0。由此定义一列概率测度 $(\mu_k)_{k\ge1}$：

$$\mu_{k+1}(\cdot) = \sum_{B\in\mathcal{A}_k} \nu_{k+1}(\cdot|B)\mu_k(B), \tag{7}$$

并且我们可以选择每个 $f_k$，使得（忽略取整问题）对于每个 $t \in T$ 和每个 $k$，都有

$$\begin{aligned}\ell(f_{k+1} \circ \pi_{k+1}(t)) &\approx \log \frac{1}{\mu_{k+1}(A_{k+1}(t))} \\ &= \log \frac{1}{\mu_k(A_k(t))} + \log \frac{1}{\nu_{k+1}(A_{k+1}(t)|A_k(t))} \\ &\ge \log \frac{1}{\mu_k(A_k(t))} \\ &\approx \ell(f_k \circ \pi_k(t)),\end{aligned}$$

其中在第二行中，我们用到了这样一个事实：$A_{k+1}(t) \in \mathcal{A}_{k+1}(A_k(t))$，因此 $\mu_{k+1}(A_{k+1}(t)) = \nu_{k+1}(A_{k+1}(t)|A_k(t))\mu_k(A_k(t))$。

## B. 关于 Fernique–Talagrand 泛函的一些事实

Fernique–Talagrand 泛函 (3) 可以用递增划分序列来表示。为此，按照 Bednorz [10], ，我们定义概率测度对 $\mu, \nu \in \mathcal{P}(T)$ 上的如下泛函：

$$M(\mu,\nu) := \int_T \int_0^1 \sqrt{\log \frac{1}{\mu(B(t,\varepsilon))}}\, \mathrm{d}\varepsilon\, \nu(\mathrm{d}t).$$

那么 $I_\mu(t) = M(\mu, \delta_t)$，其中 $\delta_t$ 是以 $t$ 为中心的狄拉克测度，并且界 (2) 和 (5) 可以表示为

$$\mathbf{E}\Big[\sup_{t\in T} X_t\Big] \lesssim \inf_\mu \sup_t M(\mu,\delta_t) = \inf_\mu \sup_\nu M(\mu,\nu)$$

以及（对于配备诱导度量的 $T$ 上的高斯过程）

$$\mathbf{E}\Big[\sup_{t\in T} X_t\Big] \gtrsim \inf_\mu \sup_t M(\mu,\delta_t) = \inf_\mu \sup_\nu M(\mu,\nu).$$

于是，我们得到 $I_\mu(t)$ 和$M(\mu,\nu)$ [10]的如下上界：设 $\mathcal{A} = (\mathcal{A}_k)_{k\ge1}$ 是 $T$ 的一个递增划分序列，使得对于所有 $A \in \mathcal{A}_k$ 以及某个 $r > 1$，所有 $k$ 均有 $\mathrm{diam}(A) \le r^{-k}$。那么，令 $\mathcal{A}_0 = \{T\}$，我们有

$$I_\mu(t) \le \sum_{k>0} r^{-k+1} \sqrt{\log \frac{\mu(A_{k-1}(t))}{\mu(A_k(t))}}$$

and

$$M(\mu,\nu) \le \sum_{k>0} r^{-k+1} \sum_{B\in\mathcal{A}_{k-1}} \sum_{A\in\mathcal{A}_k(B)} \nu(A) \sqrt{\log \frac{\mu(B)}{\mu(A)}}.$$

我们可以用信息论术语来表达这些界。对于任意 $k \ge 1$ 以及任意满足 $j < k$ 的 $B \in \mathcal{A}_j$，我们定义熵和交叉熵

$$\begin{aligned}H_{\mathcal{A}_k|B}(\mu) &:= -\sum_{A\in\mathcal{A}_k} \mu(A|B)\log\mu(A|B), \\ H_{\mathcal{A}_k|B}(\mu,\nu) &:= -\sum_{A\in\mathcal{A}_k} \mu(A|B)\log\nu(A|B).\end{aligned}$$

利用詹森不等式和界 $\inf_\mu \sup_t M(\mu,\delta_t) \le \sup_\mu M(\mu,\mu)$ [10], [11],，我们可以得到如下结果：

命题1。对于任何 $r > 1$,

$$\mathbf{E}\Big[\sup_{t\in T} X_t\Big] \lesssim \inf_{\mathcal{A}} \sup_{\mu} \sum_{k>0} r^{-k+1} \sum_{B\in\mathcal{A}_{k-1}} \mu(B)\sqrt{H_{\mathcal{A}_k|B}(\mu)},$$

其中，下确界取遍满足对所有 $A \in \mathcal{A}_k$，$k \geq 1$，$\operatorname{diam}(A) \leq r^{-k}$ 的 $T$ 的所有递增序列 $\mathcal{A} = (\mathcal{A}_k)_{k\geq 1}$ of 划分。

III. 上界及其若干推论

为简化起见，在本文余下部分中，我们假设 $T$ 有限（在所有相关随机过程均满足可分性这一温和假设下，$T$ 为无限的情形并不会困难多少）。设 $\mathcal{C} = \{(\pi_k, f_k)\}_{k\in\mathbb{N}}$ 是 $T$ 上的 VLC 的可容许序列，其中每个 $(\pi_k, f_k)$ 均以分辨率 $\rho_k$ 运行。再设 $\boldsymbol{p} = (p_k)_{k\geq 1}$ 是 $\mathbb{N}$ 上的正概率质量函数，并定义泛函

$$\bar{\sigma}_{(\mathcal{C},\boldsymbol{p})}(t) := \sum_{k>0} \rho_{k-1}\sqrt{\ln \frac{2^{\ell(f_k\circ\pi_k(t))+1}}{p_k}}, \tag{8}$$

其中我们设定了 $\rho_0 = \operatorname{diam}(T) \equiv 1$。以下是对 [5, Thm. 6] 的推广：

定理 1. 设 $(X_t)_{t\in T}$ 是满足增量条件 (1) 的中心化随机过程。设 $(\mathcal{C}, \boldsymbol{p})$ 使得 $\bar{\sigma}_{(\mathcal{C},\boldsymbol{p})}(t)$ 对所有 $t \in T$ 都是有限的。固定任意点 $t_0 \in T$。那么，对于任意 $u > 0$,

$$|X_t - X_{t_0}| \leq \bar{\sigma}_{(\mathcal{C},\boldsymbol{p})}(t)(u+1), \qquad \forall t \in T$$

其概率至少为 $1 - e^{-u^2}$。

证明。由于 $T$ 是有限的，对于每个 $t \in T$，存在某个 $k = k(t) \in \mathbb{N}$，使得 $t = \pi_k(t) = \pi_{k+1}(t) = \dots$。因此，我们可以构造链式分解 [1, p。 19]

$$X_t - X_{t_0} = \sum_{k>0} (X_{\pi_k(t)} - X_{\pi_{k-1}(t)}).$$

对于每个 $k$，令 $S_k := \pi_k(T)$，并通过以下方式定义 $\xi_k : S_k \to \mathbb{R}$

$$\xi_k(s) := d(s, \pi_{k-1}(s))\sqrt{\ln \frac{2^{\ell(f_k(s))+1}}{p_k} + u^2}$$

以及事件 $E_k := \big\{\exists s \in S_k : |X_s - X_{\pi_{k-1}(s)}| > \xi_k(s)\big\}$。利用增量条件 (1) 和 Kraft–McMillan不等式，我们有

$$\begin{aligned}
\mathbf{P}[E_k] &\leq \sum_{s\in S_k} \mathbf{P}[|X_s - X_{\pi_{k-1}(s)}| > \xi_k(s)] \\
&\leq 2\sum_{s\in S_k} \exp\left(-\frac{\xi_k^2(s)}{d^2(s,\pi_{k-1}(s))}\right) \\
&= 2\sum_{s\in S_k} \exp\left(-\ln\frac{2^{\ell(f_k(s))+1}}{p_k} - u^2\right) \\
&= p_k e^{-u^2} \sum_{s\in S_k} 2^{-\ell(f_k(s))} \\
&\leq p_k e^{-u^2}.
\end{aligned}$$

Then

$$\begin{aligned}
&\mathbf{P}\left[\exists t \in T, |X_t - X_{t_0}| > \sum_{k>0} \xi_k(\pi_k(t))\right] \\
&\leq \mathbf{P}\left[\exists t \in T, \sum_{k>0}\left(|X_{\pi_k(t)} - X_{\pi_{k-1}(t)}| - \xi_k(\pi_k(t))\right) > 0\right] \\
&\leq \mathbf{P}\left[\bigcup_k E_k\right] \\
&\leq e^{-u^2},
\end{aligned}$$

其中最后一步是由关于权重$p_k$的假设得出的。因此，由于 $d(\pi_k(t), \pi_{k-1}(t)) \leq \rho_{k-1}$,

$$|X_t - X_{t_0}| \leq \bar{\sigma}_{(\mathcal{C},\boldsymbol{p})}(t) + u\sum_{k>0}\rho_{k-1}, \quad \forall t \in T$$

其概率至少为 $1 - e^{-u^2}$。由于根据 $f_k$ 的唯一可解码性，$\ell(f_k \circ \pi_k(t)) \geq 1$ 对于$k \geq 1$ 成立：$S_k \to \{0,1\}^*$，因此可得 $\frac{2^{\ell(f_k\circ\pi_k(t))+1}}{p_k} \geq 4$，从而

$$\sum_{k>0}\rho_{k-1} \leq \bar{\sigma}_{(\mathcal{C},\boldsymbol{p})}(t)$$

对所有 $t \in T$ 均成立。证明完毕。 □

推论 1。在与定理 1 相同的假设下,

$$\mathbf{E}\Big[\sup_{t\in T} X_t\Big] \leq 2\inf_{(\mathcal{C},\boldsymbol{p})} \sup_{t\in T} \bar{\sigma}_{(\mathcal{C},\boldsymbol{p})}(t).$$

证明。令 $\bar{a} :=$ 上确界$_{t\in T}\,\bar{\sigma}_{(\mathcal{C},\boldsymbol{p})}(t)$，并令 $Z :=$ 上确界$_{t\in T}\,|X_t - X_{t_0}|$。根据定理 1，对所有 $u > 0$，$\mathrm{P}[\bar{Z} > \bar{a} + u] \leq e^{-u^2/\bar{a}^2}$，因此

$$\begin{aligned}
\mathbf{E}\Big[\sup_{t\in T} X_t\Big] &= \mathbf{E}\Big[\sup_{t\in T}(X_t - X_{t_0})\Big] \\
&\leq \mathbf{E}[\bar{Z}] \\
&= \int_0^\infty \mathbf{P}[\bar{Z} > r]\,\mathrm{d}r \\
&= \int^{\bar{a}} \mathbf{P}[\bar{Z} > r]\,\mathrm{d}r + \int_{\bar{a}}^\infty \mathbf{P}[\bar{Z} > r]\,\mathrm{d}r \\
&\leq \bar{a} + \int_0^\infty e^{-u^2/\bar{a}^2}\,\mathrm{d}u \\
&\leq 2\bar{a}.
\end{aligned}$$

现在对所有 $(\mathcal{C}, \boldsymbol{p})$ 取下确界。 □

A. 一些推论

仔细检查定理 1 的证明可知，我们并未充分利用变长码序列的所有性质。为了得到更多结果，将泛函 $\bar{\sigma}_{(\mathcal{C},\boldsymbol{p})}$ 分解如下会很方便:

$$\begin{aligned}
\bar{\sigma}_{(\mathcal{C},\boldsymbol{p})}(t) &\leq \sigma_{\mathcal{C}}(t) + \sigma'_{(\mathcal{C},\boldsymbol{p})}(t) \\
&:= \sum_{k>0}\rho_{k-1}\sqrt{\ell(f_k\circ\pi_k(t))} + \sum_{k>0}\rho_{k-1}\sqrt{\ln\frac{2}{p_k}}
\end{aligned}$$

我们总是可以选择 $\boldsymbol{p}$，使得 $\sigma'_{(\mathcal{C},\boldsymbol{p})}(t)$ 有一个绝对常数上界——例如，取 $p_k = 2^{-k}$ 和 $\rho_k = r^{-k} \le 2^{-k}$ 时，我们将有

$$\sigma'_{(\mathcal{C},\boldsymbol{p})}(t) \le \sum_{k>0} 2^{-k+1}\sqrt{(k+1)\ln 2} \le 3.$$

因此，在下文中我们可以专注于 $\sigma_{\mathcal{C}}(t)$。

定理 2. 设 $\mathcal{C} = \{(\pi_k, f_k)\}_{k\ge 1}$ 是一个可容许的 VLC 序列，且对于某个 $r \ge 2$ 有 $\rho_k = r^{-k}$。则

$$\sigma_{\mathcal{C}}(t) \lesssim \sum_{k>0} r^{-k+1}\sqrt{\ell(f_k\circ\pi_k(t)) - \ell(f_{k-1}\circ\pi_{k-1}(t))}. \quad (9)$$

证明。设 $\ell_k(t) := \ell(f_k\circ\pi_k(t))$，其中 $\ell_0(t) \equiv 0$。由 $\mathcal{C}$ 的可容许性可知，$(\ell_k(t))_{k\ge 1}$ 是一个正整数非递减序列。因此，

$$\begin{aligned}\sigma_{\mathcal{C}}(t) &= \sum_{k>0} r^{-k+1}\sqrt{\ell_k(t)}\\ &\le \sum_{k>0} r^{-k+1}\sqrt{\ell_k(t)-\ell_{k-1}(t)} + \sum_{k>0} r^{-k+1}\sqrt{\ell_{k-1}(t)}\\ &= \sum_{k>0} r^{-k+1}\sqrt{\ell_k(t)-\ell_{k-1}(t)} + r^{-1}\sigma_{\mathcal{C}}(t),\end{aligned}$$

这与 $(1-r^{-1})^{-1} \le 2$ 一起给出 (9)。 □

对于每个 $k > 0$，上界 (9) 中平方根下的量表示对描述 $\pi_{k-1}(t)$ 进行细化所需的额外码长：将其在分辨率 $r^{-k+1}$ 下的描述细化为在分辨率 $r^{-k}$ 下更精确的描述 $\pi_k(t)$。现有文献中的若干结果可以作为特殊情形得到。

首先，我们可以用标记网 [4], [12] 恢复上界。在 $(T,d)$ 上的标记网是一个二元组 $(\mathcal{A}, L)$，其中 $\mathcal{A} = (\mathcal{A}_k)_{k\ge 1}$ 是 $T$ 的递增划分序列，使得对所有 $A \in \mathcal{A}_k$ 和所有 $k \ge 1$，$\mathrm{diam}(A) \le r^{-k}$，并且 $L: \mathcal{A} \to \mathbb{N}$ 是一个标记函数，满足条件 $\{L(A) : A \in \mathcal{A}_{k+1}(B)\} = \{1, \dots, |\mathcal{A}_{k+1}(B)|\}$，对所有 $B \in \mathcal{A}_k$ 和所有 $k$ 均成立。

定理 3. 对于每个标记网 $(\mathcal{A}, L)$，都存在一个可容许的 VLC 序列，使得

$$\sigma_{\mathcal{C}}(t) \le c_1 \sum_{k>0} r^{-k+1}\sqrt{\log L(A_k(t))} + c_2, \quad (10)$$

其中 $c_1, c_2 > 0$ 是通用常数。

证明。我们将采用第 II-A 节中所述的构造。为此，定义一个次概率测度

$$\mu'_1(\cdot) = \frac{6}{\pi^2}\sum_{A\in\mathcal{A}_1}\frac{1}{L^2(A)|A|}\mathbf{1}_A(\cdot)$$

使得对每个 $A \in \mathcal{A}_1$，$\mu'_1(A) = \frac{6}{\pi^2 L^2(A)}$，并且对每个 $k$ 和 $B \in \mathcal{A}_k$，定义一个次概率测度

$$\nu'_{k+1}(\cdot|B) := \frac{6}{\pi^2}\sum_{A\in\mathcal{A}_{k+1}(B)}\frac{1}{L^2(A)|A|}\mathbf{1}_A(\cdot),$$

使得对于每个 $A \in \mathcal{A}_{k+1}(B)$，都有 $\nu'_{k+1}(A|B) = \frac{6}{\pi^2 L^2(A)}$。因此，存在概率测度 $\mu_1(\cdot) \ge \mu'_1(\cdot)$ 以及 $\nu_{k+1}(\cdot|B) \ge \nu'_{k+1}(\cdot|B)$。那么，令 $\mu_k$ 通过 (7) 递归定义，我们有

$$\begin{aligned}&\ell(f_{k+1}\circ\pi_{k+1}(t)) - \ell(f_k\circ\pi_k(t))\\ &\le \log\frac{1}{\mu_{k+1}(A_{k+1}(t))} - \log\frac{1}{\mu_k(A_k(t))} + 1\\ &= \log\frac{1}{\nu_{k+1}(A_{k+1}(t)|A_k(t))} + 1\\ &\le \log\frac{1}{\nu'_{k+1}(A_{k+1}(t)|A_k(t))} + 1\\ &\le 2\log L(A_{k+1}(t)) + \log\frac{\pi^2}{6} + 1\end{aligned}$$

于是由定理 2 得到不等式 (10)。 □

类似地，我们可以得到 Bednorz 的一个界 [10]，其中涉及一个递增划分，如上所示：

定理 4. 对于定理 3 中的每个划分序列，以及每个 $\mu \in \mathcal{P}(T)$，都存在一个由 VLC 构成的可容许序列 $\mathcal{C}$，使得

$$\sigma_{\mathcal{C}}(t) \le \sum_{k>0} r^{-k+1}\sqrt{\log\frac{\mu(A_{k-1}(t))}{\mu(A_k(t))}} + 2 \quad (11)$$

证明. 不失一般性，我们可以假设对所有 $A \in \mathcal{A}_k$ 和所有 $k$，$\mu(A) > 0$。我们沿用定理 2 证明中的相同步骤，但对所有 $k$ 和所有 $B \in \mathcal{A}_k$，取 $\mu_1 = \mu$ 和 $\nu_{k+1}(\cdot|B) = \mu(\cdot|B)$。于是，对每个 $B \in \mathcal{A}_k$ 和 $A \in \mathcal{A}_{k+1}(B)$，有 $\nu_{k+1}(A|B) = \mu(A|B) = \frac{\mu(A)}{\mu(B)}$，并且

$$\begin{aligned}&\ell(f_{k+1}\circ\pi_{k+1}(t)) - \ell(f_k\circ\pi_k(t))\\ &\le \log\frac{\mu(A_k(t))}{\mu(A_{k+1}(t))} + 1\\ &= \log\frac{\mu(A_k(t))}{\mu(A_{k+1}(t))} + 1.\end{aligned}$$

对 $k$ 求和，得到 (11)。 □

事实上，使用由单个 $\mu \in \mathcal{P}(T)$ 构造的可容许 VLC 序列，而非由递归定义的序列 $(\mu_k)_{k\ge 1}$ 构造的可容许 VLC 序列，并不会失去一般性：

定理 5. 对任意 $\boldsymbol{p} = (p_k)_{k\ge 1}$，以及由递增划分序列 $\mathcal{A} = (\mathcal{A}_k)_{k\ge 0}$ 和概率测度序列 $(\mu_k)_{k\ge 1}$ 通过 (7) 构造的任意可容许变长码序列 $\mathcal{C} = \{(\pi_k, f_k)\}_{k\ge 1}$，存在一个概率测度 $\mu \in \mathcal{P}(T)$ 和一个可容许变长码序列 $\mathcal{C}' = \{(\pi'_k, f'_k)\}_{k\ge 1}$，使得

$$\bar{\sigma}_{(\mathcal{C}',\boldsymbol{p})}(t) \le \bar{\sigma}_{(\mathcal{C},\boldsymbol{p})}(t) + \sum_{k>0} r^{-k+1}\sqrt{\log\frac{2}{p_k}}, \quad t \in T. \quad (12)$$

证明。令 $\mathcal{C} = \{(\pi_k, f_k)\}_{k\ge 1}$ 按照所示由 $\mathcal{A}$ 和 $(\mu_k)_{k\ge 1}$ 构造而成。给定 $\boldsymbol{p}$，构造混合体 $\mu = \sum_{k>0} p_k\mu_k$。然后，保持相同的划分序列 $\mathcal{A} = (\mathcal{A}_k)_{k\ge 1}$，构造 $\mathcal{C}' = \{(\pi'_k, f'_k)\}_{k\ge 1}$，使得 $\pi'_k = \pi_k$

并且 $f'_k$ 满足 $\ell(f'_k \circ \pi'_k(t)) = \lceil \log \mu \frac{1}{A_k(t)} \rceil$，对每个 $t \in T$ 和每个 $k \in T$ 都成立。然后

$$\begin{aligned}\ell(f'_k \circ \pi'_k(t)) &\le \log \frac{1}{\mu(A_k(t))} + 1 \\ &\le \log \frac{1}{\mu_k(A_k(t))} + \log \frac{2}{p_k} \\ &\le \ell(f_k \circ \pi_k(t)) + \log \frac{2}{p_k}.\end{aligned}$$

将其与 $\rho_k = r^{-k}$ 一起代入 (8)，即可得到 (12)。 □

反过来，任取一组由 VLC $\}_{k\ge1}$ 和 $_{k\ge1}$ 构成的可容许序列 $\mathcal{C}$ $=\{(\pi_k, f_k)\}_{k\ge1}$，即可得到一个概率测度

$$\mu \propto \sum_{k>0} p_k \sum_{s\in\pi_k(T)} 2^{-\ell(f_k(s))}\delta_s$$

（例如，参见 [5, 第 4] 节），且不难验证，对每个$t \in T$和每个 $k$,

$$\log \frac{1}{\mu(A_k(t))} \le \ell(f_k \circ \pi_k(t)) + \log \frac{1}{p_k}.$$

因此，本节发展的工具可用于给出命题 1 的一个自足证明。事实上，利用 Maurer [5], 提出的一种构造，我们可以完全省去对递增分划序列的显式最小化，如命题 1 所示——取而代之的是，给定 $\mu \in \mathcal{P}(T)$（或等价地，给定一个唯一可译码 $f : T \to \{0,1\}^*$），在分辨率级别 $r^{-k}$ 上选择映射 $\pi_k$ 和 $f_k$，使得对每个 $t \in T$，二进制字符串 $f_k \circ \pi_k(t)$ 由无损表示 $f(t)$ 截断至最少比特数得到，并且基于这一截断表示，所有可能与 $t$ “混淆”的 $t' \in T$ 构成的集合，其直径至多为 $r^{-k}$。然后将该集合取作 $A_k(t)$。详细内容留待本文完整版介绍。

## IV. 高斯过程的下界

现在考虑过程 $(X_t)_{t\in T}$ 为高斯过程且 $T$ 赋予典范度量 (4) 的情形。我们将使用 $L$、$L_1$、.....表示各种绝对常数。注意，在不同出现位置，这些常数的取值不一定相同。下界的主要技术工具是下面对 Sudakov 下界 [1,命题 2.4.9]的改进：

命题 2. 设 $t_1$, ....., $t_m \in T$ 满足：当 $\ell \ne \ell'$ 时，$d(t_\ell, t_{\ell'}) \ge a$。考虑 $b > 0$，并且对每个 $\ell \le m$，取一个有限集 $H_\ell \subset B(t_\ell, b)$。则对于 $H = \bigcup_{\ell\le m} H_\ell$，有

$$\mathbf{E}\Big[\sup_{t\in H} X_t\Big] \ge \frac{a}{L_1}\sqrt{\log m} - L_2 b\sqrt{\log m} + \min_{\ell\le m} \mathbf{E}\Big[\sup_{t\in H_\ell} X_t\Big].$$

当 $b \le a/(2L_1L_2)$ 时，我们有

$$\mathbf{E}\Big[\sup_{t\in H} X_t\Big] \ge \frac{a}{2L_1}\sqrt{\log m} + \min_{\ell\le m} \mathbf{E}\Big[\sup_{t\in H_\ell} X_t\Big].$$

另一个要素是在 $T$ 上构造一个递增的划分序列。我们遵循 Bednorz [10] 和 Talagrand [13] 并使用贪心过程为固定的 $r > 1$ 构造划分 $(\mathcal{A}_k)_{k\ge0}$。对于任意 $A \subseteq T$，令$G(A) := \mathrm{E}[\sup_{t\in A} X_t]$。令 $\mathcal{A}_0 = \{T\}$。为了从 $\mathcal{A}_{k-1}$为$k \ge 1$ 构造 $\mathcal{A}_k$，我们将每个 $B \in \mathcal{A}_{k-1}$ 划分为集合$A_1$、.....、$A_m$，具体如下：令 $B_0 := B$，并选择 $t_1 \in B_0$，使得

$$G(B(t_1, r^{-k-1}/2) \cap B_0) = \max_{t\in B_0} G(B(t, r^{-k-1}/2) \cap B_0).$$

令 $A_1 := B(t_1, r^{-k}/2) \cap B_0$，且 $B_1 := B_0 \setminus A_1$。现在进行迭代：如果对于 $i \ge 1$、$B_{i-1} \ne \varnothing$，选择 $t_i \in B_{i-1}$ 使得

$$\begin{aligned}&G(B(t_i, r^{-k-1}/2) \cap B_{i-1}) \\ &\quad = \max_{t\in B_{i-1}} G(B(t, r^{-k-1}/2) \cap B_{i-1}).\end{aligned}$$

令 $A_i := B(t_i, r^{-k}/2) \cap B_{i-1}$ 并令 $B_i := B_{i-1} \setminus A_i$，然后继续。对于某个 $m \le N(T, d, r^{-k}/2)$，此过程必定会终止，其中 $N(T, d, \cdot)$ 是 $(T, d)$ 的覆盖数。

现在我们可以通过编码来陈述主导测度定理了。

定理 6. 设 $(X_t)_{t\in T}$ 是一个中心高斯过程。对于任意 $r \ge \max\{2, L\}$，存在一个由 VLC 构成的可容许序列 $\mathcal{C}$，满足 $\rho_k \le r^{-k}$，使得

$$\mathbf{E}\Big[\sup_{t\in T} X_t\Big] \gtrsim \frac{1}{r}\sigma_{\mathcal{C}}(t), \qquad \forall\, t \in T. \tag{13}$$

证明. 令 $L = L_1L_2$，其中 $L_1, L_2$ 是命题 2 中的常数。给定 $r \ge L$，令 $(\mathcal{A})_{k\ge0}$ 为如上所述的递增划分序列。对于每个 $k$ 和每个 $A \in \mathcal{A}_k$，选取唯一的点 $s_A \in A$，然后令 $\pi_k(t) := s_{A_k(t)}$，其中 $A_k(t)$ 是 $\mathcal{A}_k$ 中包含 $t$的唯一元素。

我们归纳地构造编码 $f_k$。令 $f_0(\pi_0(t)) = \mathsf{e}$，即 $\{0,1\}^*$ 中的空字符串。假设已构造出 $f_0, \ldots, f_{k-1}$，我们为点 $s \in \pi_k(T)$ 指定编码长度 $\ell_k(s)$，并证明这一指定满足 Kraft–McMillan 不等式。因此，存在一个前缀编码 $f_k : \pi_k(T) \to \{0,1\}^*$，使得对所有 $s \in \pi_k(T)$，都有 $\ell(f_k(s)) = \ell_k(s)$。

对于每个 $s \in \pi_k(T)$，令 $i_k(s) \in \mathbb{N}$ 为集合 $A_k(s) \equiv \pi_k^{-1}(s)$ 作为分割元素在 $A_{k-1}(s) \equiv \pi_{k-1}^{-1}(s)$ 中的索引，该分割由上述贪心构造过程继承而来。于是，令 $\ell_k(s) = \ell(f_{k-1} \circ \pi_{k-1}(s)) + \log(i_k(s) + 1)^2$,，我们有

$$\begin{aligned}&\sum_{s\in\pi_k(T)} 2^{-\ell_k(s)} \\ &= \sum_{t\in\pi_{k-1}(T)} \sum_{s:\pi_{k-1}(s)=t} 2^{-\ell(f_{k-1}(s))-\log(i_k(s)+1))^2} \\ &\le \sum_{t\in\pi_{k-1}(T)} 2^{-\ell(f_{k-1}(t))} \sum_{i\ge1} \frac{1}{(i+1)^2} \le 1.\end{aligned}$$

因此，根据 Kraft–McMillan不等式，可以保证存在一个唯一可译码（前缀）码$f_k$。接着，根据构造过程和命题 2，对于每个 $t \in T$,

$$G(A_{k-1}(t)) \ge G(A_{k+1}(t)) + Lr^{-k}\sqrt{\log(i_k(\pi_k(t)))}.$$

由于对于 $i \geq 1$，$\sqrt{1+\log i} \geq \sqrt{1+\log(i+1)}$，

$$\begin{aligned}
&G(A_{k-1}(t)) + Lr^{-k} \\
&\geq G(A_{k+1}(t)) + Lr^{-k}\sqrt{\log(i_k(\pi_k(t))+1)^2} \\
&= G(A_{k+1}(t)) + Lr^{-k}\sqrt{\ell_k(t) - \ell_{k-1}(t)},
\end{aligned}$$

其中 $\ell_k(t) = \ell\mathrm{diam}(f_k \circ \pi_k)$。对 $G(A_0(t))$ 和 $G(A_1(t))$ 应用归纳法，可得

$$G(T) + L \gtrsim \sum_k r^{-k}\sqrt{\ell_k(t) - \ell_{k-1}(t)}, \qquad \forall t \in T \quad (14)$$

定理 2 以及 $G(T) \gtrsim \mathrm{diam}(T)$ 这一事实（因此可将普适常数吸收到 $G(T)$ 中）给出了所需结果。 □


致谢

作者谨感谢 Daniel Roy 指出了证明定理 1 中的一些错别字和不准确之处。


参考文献


[1] M. Talagrand，《随机过程的上界与下界》。Springer，2014。[2] X. Fernique，“高斯随机函数轨道的正则性”，载于《圣弗卢尔概率暑期学校 IV—1974》。Springer，1975，第 1–96 页。[3] M. Talagrand，“高斯过程的正则性”，《数学学报》，第 159 卷，第 99–149 页，1987。[4] O. Guédon 和 A. Zvavitch，“用树表示过程的上确界”，载于《泛函分析的几何方面》。Springer，2003，第 136–147 页。[5] A. Maurer，“主导码与测度”，预印本，见 http://www.andreas-maurer.eu/mmnotes.pdf，2010。[6] T. M. Cover 和 J. A. Thomas，《信息论基础》，第 2 版。Wiley，2006。[7] I. Kontoyiannis 和 J. Zhang，“率失真理论中的任意源模型与贝叶斯码本”，IEEE 信息论汇刊，第 48 卷，第 8 期，第 2276–2290 页，2002 年 8 月。[8] P. Massart，《集中不等式与模型选择》。Springer，2007。[9] J.-Y. Audibert 和 O. Bousquet，“结合 PAC-贝叶斯方法与通用链界”，《机器学习研究杂志》，第 8 卷，第 32 期，第 863–889 页，2007。[10] W. Bednorz，“样本有界性的主导测度方法”，《数学讨论会》，第 139 卷，第 2 期，第 205–227 页，2015。[11] X. Fernique，“有界或连续轨道过程的刻画”，《斯特拉斯堡概率研讨会》，第 12 卷，第 691–706 页，1978。[12] R. van Handel，“高维概率”，普林斯顿大学讲义，2016。[13] M. Talagrand，“主导测度定理的一个简单证明”，《几何与泛函分析》，第 2 卷，第 1 期，第 118–125 页，1992 年 3 月。