arXiv:2602.02391v1 [math.PR] 2 Feb 2026

本原有理模型中的多元符号统计分析

Massimiliano Goldwurm[(1)]、Claudio Macci[(2)]、Marco Vignati[(1)]、Elena Villa[(1)]

(1) 意大利米兰大学数学系 (2) 意大利罗马第二大学数学系

摘要

我们研究多元随机变量序列的渐近行为。这些随机变量表示在有理随机模型下随机生成的长度为 n 的字符串中，给定符号集合的出现次数。假设模型的转移总权重矩阵是本原的，我们首先确定此类统计量的均值和协方差的渐近表达式。随后，我们建立两个渐近结果，将已知的单变量情形推广到不同的状态：具有速度 n的大偏差原理，并由此推出几乎必然收敛；以及多元高斯极限。此外，我们还提出了一项新的中等偏差结果，作为这两种状态之间的桥梁。我们的证明核心是统计量矩母函数的拟幂性质，这使我们能够将 G¨artner-Ellis 定理应用于大偏差和中等偏差。

关键词：大偏差、极限分布、模式统计、正则语言。

AMS 分类：60F10, 68Q45, 68Q87。

1 引言

本文致力于分析一系列多元随机变量，它们表示在一个长度为 n 的字符串中，来自给定有限字母表 Σ的符号出现的次数。该字符串由有理随机模型按照随机方式生成，记作 Σ^* ([1])。此模型在 [2] 中提出，用于研究正则语言中给定长度的词里模式的出现次数。它可以由一个在转移上配备正实权重的（非确定性）有限状态自动机定义（在 [23]中称为广义自动机 ）。对于任何 $n \in \mathbb{N}_+$，长度为 n 的字符串 $w \in \Sigma^*$ 的概率与该自动机在输入 w时所有接受计算的总权重成正比。这些随机模型将任何正则语言中给定长度字符串的均匀随机生成作为特例。它们还包括模式统计分析中用于随机生成给定长度文本的其他经典概率模型，例如伯努利模型或马尔可夫模型 [17, 18, 9, 16]。

在单变量情形中，当 Σ 缩减为二元字母表时，有理模型已在此前的多篇论文中得到研究，并已获得关于给定长度的随机词中单个符号

[1]传统上， Σ^* 是 Σ 上所有字符串的集合。

出现次数的若干性质。特别是，当转移权重矩阵为本原矩阵时，这一数量服从高斯极限分布，具有局部极限定理的各种性质，并满足大偏差原理 [2, 4, 11, 13]。

本工作的主要目标是将其中一些结果扩展到多元情形，即字母表 Σ 包含 $\ell+1$ 个元素，且所研究的随机变量具有$\ell>1$ 个分量。这里，我们尤其关注 [13, 14]中最近在单变量情形下研究的大偏差性质。我们回顾一下，大偏差性质是概率论中的一个经典课题 [5, 6], ，通常在模式统计 [3, 7],的背景下，以及更一般地在组合结构 [15, 8]的分析中进行研究。

为介绍我们的结果，记$\mathbf{Y}_n$ 为一个具有 ℓ个分量的随机变量，它表示在有理随机模型下按随机方式生成的长度为 n 的字符串中， Σ 的前 ℓ 个元素的出现次数。假设定义该模型的自动机的权重矩阵具有本原性，我们确定了 Y_n的均值和协方差矩阵的渐近表达式。特别地，对于某个 $\boldsymbol{\beta}\in\mathbb{R}^{\ell}$，有 $\mathbb{E}[\mathbf{Y}_n]=\boldsymbol{\beta}n+o(n)$。此外，在相同的本原性假设下，我们证明以下结果：

- $\frac{\mathbf{Y}_n-\boldsymbol{\beta}n}{\sqrt{n}}=\sqrt{n}\left(\frac{\mathbf{Y}_n}{n}-\boldsymbol{\beta}\right)$ 依分布收敛于一个中心高斯随机变量；
- $\{\mathbf{Y}_n/n\}$ 满足速度为 n 的大偏差原理，这意味着概率$P(\|\mathbf{Y}_n/n-\boldsymbol{\beta}\|\geq\delta)$以指数速度衰减至零（对每个 $\delta>0$），从而保证$\mathbf{Y}_n/n$几乎必然收敛于 $\boldsymbol{\beta}$；
- 中等偏差结果成立，从而在上述两种渐近状态之间架起桥梁。

我们注意到，尽管前两项结果与单变量情形 $\ell=1$ [2, 13] 的已知结果一致，但中等偏差结果则是一项全新的贡献。大偏差与中等偏差均通过应用 G¨artner-Ellis 定理得到，而这之所以可行，是由于 $\mathbf{Y}_n$ 的矩母函数具有拟幂性质。

本文内容安排如下：在第 2 节中，我们回顾有关大偏差的标准概念与结果；在第 3 节中，我们在本文语境下定义有理模型，并引入随机变量 $\mathbf{Y}_n$；在第 4 节中，我们给出本原情形下$\mathbf{Y}_n$ 的若干性质，特别是证明其矩母函数的拟幂定理，并证明其均值与协方差矩阵的渐近表达式。第 5 节给出 $\frac{\mathbf{Y}^n-\boldsymbol{\beta}n}{\sqrt{n}}$的高斯极限分布。上述另外两项结果随后分别在第 6 节和第 7 节中加以证明。第 8 节对结果作进一步~~评述~~并加以总结，同时讨论若干可能的开放问题；最后，在附录（第 9 节）中，我们回顾一些可能对读者有用的已知性质。

2 大偏差（和中等偏差）预备知识

本节回顾大偏差（及中等偏差）的一些预备知识，并讨论取值于 $\mathbb{R}^m$ 的随机变量序列，其中 $m\geq1$ 为任意整数。主要参考文献为 [5, 6]。

首先介绍一些预备记号。对于任意一对元素 $\mathbf{x} = (x_1, \ldots, x_m)$、$\mathbf{y} = (y_1, \ldots, y_m) \in \mathbb{R}^m$，我们用 $\mathbf{x} \cdot \mathbf{y}$ 表示标量积 $\sum_{i=1}^n x_i y_i$。必要时，将元素 $\mathbf{x} \in \mathbb{R}^m$ 视为列数组，而 $\mathbf{x}'$ 表示其转置（行）数组；因此，$\mathbf{x} \cdot \mathbf{y}$ 也可以表示为 $\mathbf{x}'\mathbf{y}$。对于每个集合 $D \subset \mathbb{R}^m$，我们分别用 D^c、D° 和 $\underline{D}$表示 D的补集、内部和闭包。此外，对于任意函数 $f\colon \mathbb{R}^m \to [-\infty, \infty]$，令 $\mathcal{D}_f := \{\mathbf{x} \in \mathbb{R}^m\colon f(\mathbf{x}) \in \mathbb{R}\}$，其中 $\nabla f(\mathbf{x})$ 表示 f在 $\mathbf{x} \in \mathbb{R}^m$处的梯度，而$Hf(\mathbf{x})$ 表示 f 在 $\mathbf{x} \in \mathbb{R}^m$处的海森矩阵。最后，对于每个 $\delta \geq 0$ 以及每个 $\mathbf{x_0} \in \mathbb{R}^m$，$B_\delta(\mathbf{x_0}) := \{\mathbf{x} \in \mathbb{R}^m\colon \|\mathbf{x} - \mathbf{x_0}\| < \delta\}$ 表示以 $\mathbf{x_0}$ 为中心、半径为 δ的开球。

一个速率函数定义为下半连续映射 $I\colon \mathbb{R}^m \to [0, \infty]$；若对任意 $c \geq 0$，水平集 $\{\mathbf{x} \in \mathbb{R}^m\colon I(\mathbf{x}) \leq c\}$ 都是紧集，则称该函数为良好。我们回顾一下：下半连续函数总能在任意非空紧集上取到最小值。

现在，考虑一个正实数序列 $\{v_n\}_{n\geq 1}$，满足 $\lim_{n\to\infty} v_n = \infty$。我们称定义在同一概率空间 $(\Omega\ \mathcal{F}, P)$ 上的一列 $\mathbb{R}^m$ 值随机变量 $\{\mathbf{X}_n\}_{n\geq 1}$，以速度 v_n 和速率函数I满足大偏差原理（简称 LDP），如果以下关系成立：

$$\liminf_{n\to\infty} \frac{1}{v_n} \log P(\mathbf{X}_n \in O) \geq -\inf_{\mathbf{x}\in O} I(\mathbf{x}), \qquad \text{for all open sets } O \subseteq \mathbb{R}^m; \tag{1}$$

$$\limsup_{n\to\infty} \frac{1}{v_n} \log P(\mathbf{X}_n \in C) \leq -\inf_{\mathbf{x}\in C} I(\mathbf{x}), \qquad \text{for all closed sets } C \subseteq \mathbb{R}^m. \tag{2}$$

(1) 和 (2) 的一个等价表述如下：

$$-\inf_{\mathbf{x}\in D^\circ} I(\mathbf{x}) \leq \liminf_{n\to\infty} \frac{1}{v_n} \log P(\mathbf{X}_n \in D) \leq \limsup_{n\to\infty} \frac{1}{v_n} \log P(\mathbf{X}_n \in D) \leq -\inf_{\mathbf{x}\in \overline{D}} I(\mathbf{x}) \tag{3}$$

对所有 $D \subset \mathbb{R}^m$。

根据上述大偏差原理的定义，可以推导出一些在本文语境下有用的结果和简单观察：

1. 如果对于某个 $D \subset \mathbb{R}^m$，有 $\inf_{\mathbf{x}\in D^\circ} I(\mathbf{x}) = \inf_{\mathbf{x}\in D} I(\mathbf{x})$，那么

$$\lim_{n\to\infty} \frac{1}{v_n} \log P(\mathbf{X}_n \in D) = -\inf_{\mathbf{x}\in D} I(\mathbf{x});$$

因此，粗略地说，当 $_n \in D$ 较大时，$P(\mathbf{X}_n \in D) \approx e^{-v_n \inf_{x\in D} I(\mathbf{x})}$（并且，如果 $\inf_{x\in D} I(\mathbf{x}) > 0$，则其以指数衰减趋于零）。那么，$\inf_{x\in D} I(\mathbf{x})$ 的值越大，当 n 较大时，事件 $\{\mathbf{X}_n \in D\}$ 发生的可能性就越小。特别是，当对每个 n 都有 $P(\mathbf{X}_n \in D) = 0$ 时，$\inf_{x\in D} I(\mathbf{x}) = +\infty$。

2. 如果随机变量 $\mathbf{X}_n$ 取值于闭集 $C_\bullet$，也就是说，对每个 n 都有 $P(\mathbf{X}_n \in C_\bullet) = 1$，那么速率函数 $I(\mathbf{x})$ 在补集 $C_\bullet^c$ 上等于无穷。

3. 可以证明（参见第 9.2 节），如果 I 是一个良好速率函数，并且存在$\mathbf{x}_0$，使得 $I(\mathbf{x}) = 0$ 当且仅当 $\mathbf{x} = \mathbf{x}_0,$，那么序列 $\{\mathbf{X}_n\}_{n\geq 1}$ 依概率收敛到$\mathbf{x}_0$。此外，如果 $v_n = n$，则该收敛几乎必然发生。

在我们的工作中，序列 $\{\mathbf{X}_n\}_{n\geq 1}$的形式为 $\{\mathbf{Y}_n/n\}_{n\geq 1}$，其中 $\mathbf{Y}_n=(Y_{n,1},\ldots,Y_{n,m})$，每个分量 $Y_{n,i}$ 取值于 $\{0,1,\ldots,n\}$，并且有 $Y_{n,1}+\cdots+Y_{n,m}\leq n$。因此，上文提到的集合 $C_\bullet$ 就是传统单纯形，即

$$C_\bullet=\{(x_1,\ldots,x_m)\in\mathbb{R}^m : x_1,\ldots,x_m\geq 0, x_1+\cdots+x_m\leq 1\}.$$

我们将证明序列 $\{\mathbf{X}_n\}_{n\geq 1}$ 满足速度为 $v_n=n$ 的大偏差原理。由此还可得出，相应的速率函数仅在单点$\mathbf{x}_0$, 处取零；该点将在下文确定并记为 $\boldsymbol{\beta}$。

我们还将给出一个关于中等偏差的结果。文献中使用这一术语来指代一类适当的大偏差原理，它们连接了以下两种渐近状态:

- $\{\mathbf{X}_n-\mathbf{x}_0\}_{n\geq 1}$ 收敛到零;
- $\{\sqrt{n}(\mathbf{X}_n-\mathbf{x}_0)\}_{n\geq 1}$ 弱收敛到具有适当协方差矩阵 V 的中心正态分布。

更确切地说，对于满足 $a_n\to 0$ 和 $na_n\to\infty$ 的任意正数列 $\{a_n\}_{n\geq 1}$，我们将证明序列 $\{\sqrt{na_n}(\mathbf{X}_n-\mathbf{x}_0)\}_{n\geq 1}$ 满足速度为 $1/a_n$ 的大偏差原理，并具有适当的良好速率函数 J。因此，在某种意义上，上述两种渐近状态分别对应于 $a_n=1/n$ （此时 $na_n\to\infty$ 不成立）和 $a_n=1$ （此时 $a_n\to 0$ 不成立）。

关于 $\{\mathbf{X}_n\}_{n\geq 1}$ 以 n为速度的大偏差结论，以及中等偏差结论，将通过适当应用*Gärtner–Ellis* 定理 （例如参见[5, Th. 2.3.6]）来获得，该定理的表述回顾如下。为此，我们首先回顾：若凸函数 $f\colon \mathbb{R}^m\to(-\infty,\infty]$ 满足以下条件，则称其为本质光滑： $\mathcal{D}_f^\circ$的内部非空， f在 $\mathcal{D}_f^\circ$上处处可微，并且 f是陡峭的，即当 $\{\mathbf{x}_n\}_{n\geq 1}$是 $\mathcal{D}_f^\circ$ 中收敛到 $\mathcal{D}_f$边界上某一点的序列时，$\lim_{n\to\infty}\|\nabla f(\mathbf{x}_n)\|=\infty$。

定理1（Gärtner-Ellis 定理）设 $\{\mathbf{Z}_n\}_{n\geq 1}$ 是一个取值于 $\mathbb{R}^m$的随机变量序列，且存在函数 $\Lambda:\mathbb{R}^m\to(-\infty,\infty]$，其定义为

$$\Lambda(\mathbf{t}):=\lim_{n\to\infty}\frac{1}{v_n}\log\mathbb{E}\left[e^{v_n\,\mathbf{t}\cdot\mathbf{Z}_n}\right],\qquad\forall\ \mathbf{t}\in\mathbb{R}^m.$$

假设原点 $\mathbf{0}=(0,..,0)\in\mathbb{R}^m$属于内部 $\mathcal{D}_\Lambda^\circ$，并且 Λ本质光滑且下半连续。那么 $\{\mathbf{Z}_n\}_{n\geq 1}$ 满足大偏差原理，其速度为 v_n，良好速率函数 Λ^* 定义为

$$\Lambda^*(\mathbf{x}):=\sup_{\mathbf{t}\in\mathbb{R}^m}\{\mathbf{t}\cdot\mathbf{x}-\Lambda(\mathbf{t})\}.$$

注1 定理*1*中的函数 Λ^* 被称为函数 Λ 的勒让德-芬切尔变换，它是一个凸的下半连续函数。如果我们可以应用 *Gärtner-Ellis* 定理，那么速率函数仅在 $\mathbf{x}=\nabla\Lambda(\mathbf{0})$ 处取零。通过凸分析中的一些标准论证可以证明这一性质，我们将在第 *9.1* 节回顾这些论证。因此，如上文第 *3* 点所述， $\{\mathbf{Z}_n\}_{n\geq 1}$ 依概率收敛到 $\nabla\Lambda(\mathbf{0})$。

注2 在 *G¨artner-Ellis* 定理的假设下（尤其是由 Λ的可微性），我们有

$$\mathbf{x} \in \mathcal{D}_{\Lambda^*}^o \iff \mathbf{x} \in \{\nabla\Lambda(\mathbf{t}) : \mathbf{t} \in \mathcal{D}_\Lambda^o\} \ .$$

注3 上述协方差矩阵 V 是函数 Λ （针对速度为 n 的初始大偏差原理）的海森矩阵，在 $\mathbf{t} = \mathbf{0}$ 处取值。

之后将应用盖特纳–埃利斯定理的两种情形（即速度为 n的初始大偏差原理，以及中等偏差区域）中，函数 Λ 在整个 $\mathbb{R}^m$上均有限且可微；因此，两种情形下都显然满足陡峭性条件。

3 有理模型中的多元符号统计

本节定义本文所用的有理随机模型。

给定正整数 ℓ，考虑一个有限字母表 $\Sigma = \{a_1, \ldots, a_\ell, b\}$。按惯例， Σ^* 表示由 Σ中的所有字符串构成的集合， ϵ 是 Σ^*中的空词；而对于每个 $w \in \Sigma^*$ 和每个 $x \in \Sigma$， $|w|$ 表示 w 的长度，$|w|_x$ 表示 w中 x 的出现次数。显然， Σ^*与 ϵ 以及字符串间的连接运算共同构成一个幺半群（称为 Σ上的自由幺半群）。对于每个 $n \in \mathbb{N}$，我们还用 Σ^n 表示集合 $\{w \in \Sigma^* : \ |w| = n\}$。此外，对于整数 $m \geq 1$，固定两个（列）向量 ξ， $\eta \in \mathbb{R}_+^m$、一个矩阵$B \in \mathbb{R}_+^{m\times m}$以及一个矩阵集合$A_i \in \mathbb{R}_+^{m\times m}$， $i = 1$，...， ℓ，其中 $\mathbb{R}_+ = [0, \infty)$ 。

现在，令 μ: $\Sigma^* \to \mathbb{R}_+^{m\times m}$为由以下映射生成的幺半群同态：对每个 $i = 1, \ldots, \ell$，令 $\mu(b) = B$ 和$\mu(a_i) = A_i$。显然， $\mu(\epsilon) = I$，并且对于每个满足 $w_i \in\Sigma$ 的 $w = w_1 \cdots w_n \in \Sigma^+$，对每个 i，都有 $\mu(w) = \mu(w_1) \cdots \mu(w_n)$。我们将函数 r: $\Sigma^* \to \mathbb{R}_+$ 定义为将每个 $w \in \Sigma^*$ 映射到值 $\xi'\mu(w)\eta$ 的映射。根据文献 [1, 20]， r 是非交换变量 $a_1, \ldots, a_\ell$ 上的有理形式级数，可表示为形式$r = \sum_{w\in\Sigma^*}(r, w)w$，其中（$r, w$）表示 r 在字符串 w 处的取值，即（r, w） $= \xi'\mu(w)\eta$。我们称三元组（ξ, μ, η）为 r 的线性表示。显然，这样的三元组（ξ, μ, η）可视为字母表 Σ上的有限状态自动机，其转移以正实数值加权。因此，每个矩阵 A_i （分别为 B）表示标记为 a_i（分别为 b）的转移权重，而 ξ （分别为 η）表示初始状态（分别为终止状态）的权重。

在本文中，我们始终假设集合 $\{w \in \Sigma^n : \ (r, w) \ > 0\}$ 对每一个足够大的$n \in \mathbb{N}_+$ 均为非空（从而 $\xi \neq \mathbf{0} \neq \eta$），并且所有 A_i 与 B 都是非零矩阵（即，它们每一个都至少有一个正元素）。注意，对每一个 $n \in \mathbb{N}$，都可以容易地计算与 Σ^n中的词相关联的 r 的所有值的和：事实上，定义 $M = A_1 + \cdots + A_\ell + B$后，我们有

$$\sum_{w\in\Sigma^n}(r, w) = \xi' \sum_{w\in\Sigma^n} \mu(w)\eta = \xi' \left(\prod_{i=1}^{n} \sum_{w_i\in\Sigma} \mu(w_i) \right) \eta = \xi' M^n \eta \ . \tag{4}$$

因此，我们可以考虑定义在集合 Σ^n 上的概率测度 $\mathbf{Pr}$:

$$\Pr(w) := \frac{(r, w)}{\sum_{x\in\Sigma^n}(r, x)} = \frac{\xi'\mu(w)\eta}{\xi' M^n \eta} \ , \qquad \forall \ w \in \Sigma^n \ . \tag{5}$$

请注意，若 r 是语言 $L \subseteq \Sigma^*$ 的特征级数，则 $\mathbf{Pr}$ 是集合 $L \cap \Sigma^n$ 上的均匀概率函数。还要注意，当 M 是随机矩阵、ξ 是随机数组且 $\eta' = (1,1\ldots,1)$ 时，就得到传统马尔可夫模型（用于在 Σ^* 中随机生成一个字符串）。

那么，在此前的假设下，对于每个 $n \in \mathbb{N}_+$，我们都可以定义多元随机变量

$$\mathbf{Y}_n = (|w|_{a_1}, |w|_{a_2}, \ldots, |w|_{a_\ell}) \ ,$$

其中，w 是在 Σ^n 中随机选取的，其概率为 $\mathbf{Pr}(w)$。我们有 $\mathbf{Y}_n = (Y_{n,\,1}, \ldots, Y_{n,\,\ell}) \in \{0, 1, \ldots, n\}^\ell$ 且 $\sum_{i=1}^{\ell} Y_{n,\,i} \leq n$。因此，$\mathbf{Y}_n$ 的取值属于集合 $\mathbf{Sim}_n$，该集合定义如下：

$$\mathrm{Sim}_n := \left\{ \mathbf{k} = (k_1, \ldots, k_\ell) \in \mathbb{N}^\ell : \sum_{i=1}^{n} k_i \leq n \right\} \ .$$

此外，由关系式 (4) 和 (5)，$\mathbf{Y}_n$ 的概率函数为 y

$$\begin{aligned} p_n(\mathbf{k}) := P(\mathbf{Y}_n = \mathbf{k}) &= \frac{\displaystyle\sum_{|w|=n, |w|_{a_1}=k_1, \ldots, |w|_{a_\ell}=k_\ell} (r, w)}{\displaystyle\sum_{w \in \Sigma^n} (r, w)} = \\ &= \frac{[\mathbf{x}^{\mathbf{k}}]\ \xi'(A_1 x_1 + \cdots + A_\ell x_\ell + B)^n \eta}{\xi' M^n \eta} \ , \qquad \forall\ \mathbf{k} \in \mathrm{Sim}_n \ , \end{aligned} \tag{6}$$

其中，对于任意多项式 $g(\mathbf{x}) \in \mathbb{R}[\mathbf{x}],$，我们用 $[\mathbf{x}^{\mathbf{k}}]g(\mathbf{x})$ 表示 $g(\mathbf{x})$ 中次数为 $\mathbf{k}$ 的单项式 $x_1^{k_1} \cdots x_\ell^{k_\ell}$ 的系数。

现在，考虑 $\mathbf{Y}_n$ 的矩母函数 $\Psi_n(\mathbf{t})$，其中 $\mathbf{t} \in \mathbb{R}^\ell$。定义

$$h_n(\mathbf{t}) := \xi'(A_1 e^{t_1} + \cdots + A_\ell e^{t_\ell} + B)^n \eta \ , \tag{7}$$

并利用关系式 (6)，可得

$$\Psi_n(\mathbf{t}) := \mathbb{E}(e^{\mathbf{t}\cdot\mathbf{Y}_n}) = \sum_{\mathbf{k} \in \mathrm{Sim}_n} p_n(\mathbf{k}) e^{\mathbf{t}\cdot\mathbf{k}} = \frac{h_n(\mathbf{t})}{h_n(\mathbf{0})} \ , \qquad \forall\ \mathbf{t} \in \mathbb{R}^\ell \ . \tag{8}$$

我们还定义

$$\Phi_n(\mathbf{t}) := \log \Psi_n(\mathbf{t}) = \log \mathbb{E}(e^{\mathbf{t}\cdot\mathbf{Y}_n}) \ .$$

众所周知，[10, 8] $\Phi_n(\mathbf{t})$ 可用于计算 $\mathbf{Y}_n$ 的均值和协方差。在这里，我们有

$$\mathbb{E}[\mathbf{Y}_n] := (\mathbb{E}[Y_{n,1}], \ldots \mathbb{E}[Y_{n,\ell}]) = \nabla \log \Psi_n(\mathbf{t})|_{\mathbf{t}=\mathbf{0}} = \nabla \Phi_n(\mathbf{0}) = \frac{\nabla h_n(\mathbf{0})}{h_n(\mathbf{0})} \ , \tag{9}$$

$$\mathrm{Cov}(\mathbf{Y}_n) := [\mathrm{Cov}(Y_{n,i}, Y_{n,j})]_{i,j \in \{1,\ldots,\ell\}} = H \log \Psi_n(\mathbf{t})|_{\mathbf{t}=\mathbf{0}} = H\Phi_n(\mathbf{0}) \ , \tag{10}$$

其中，回顾一下，$H\Phi(\mathbf{0})$ 是 $\Phi(\mathbf{t})$ 在点 $\mathbf{t} = \mathbf{0}$ 处的海森矩阵：

$$H\Phi_n(\mathbf{0}) = \Big[\frac{\partial^2 \Phi_n(\mathbf{t})}{\partial t_i \partial t_j} \Big|_{\mathbf{t}=\mathbf{0}} \Big]_{i,j \in \{1,\ldots,\ell\}} .$$

特别是，这意味着

$$\mathrm{Cov}(Y_{n,i}, Y_{n,j}) = \frac{1}{h_n(\mathbf{0})} \frac{\partial^2 h_n(\mathbf{t})}{\partial t_i \partial t_j}\Big|_{\mathbf{t}=\mathbf{0}} - \frac{1}{h_n(\mathbf{0})^2} \frac{\partial h_n(\mathbf{t})}{\partial t_i}\Big|_{\mathbf{t}=\mathbf{0}} \frac{\partial h_n(\mathbf{t})}{\partial t_j}\Big|_{\mathbf{t}=\mathbf{0}}, \qquad \forall i, j \in \{1, \ldots, \ell\} .$$

4 本原情形的分析

本节研究矩阵 M 为本原矩阵时 $\{\mathbf{Y}_n\}$的渐近行为。

回顾一下，若存在正整数 n 使得$T^n > 0$（即 T^n的所有元素都严格为正），则称矩阵 $T \in \mathbb{R}_+^{m\times m}$为本原矩阵。在这种情况下，著名的佩龙-弗罗贝尼乌斯定理成立。这里我们回顾这一结果及其一些在本文语境中有用的推论（例如参见 [21, 第 1.1]节）。该定理指出，任一本原矩阵 T 都有一个模最大的唯一特征值 λ，通常称为 T的佩龙-弗罗贝尼乌斯特征值；此外， λ 为实正数，是 T 的特征多项式的单根，并且存在严格为正的左、右特征向量。此外，已知当 $n \to \infty$

$$T^n = \lambda^n \; uv'(1 + O(\varepsilon^n)) , \tag{11}$$

对某个 $\varepsilon \in$ （0，1）成立，其中 v'和 u 分别是 T关于λ的左、右特征向量，并经过归一化，使得 $v'u = 1$。显然， v 和 u 是严格为正的数组。还证明了矩阵 uv' 满足恒等式

$$uv' = \frac{\mathrm{Adj}(\lambda I - T)}{\phi'(\lambda)} , \tag{12}$$

其中 $\phi(x) = \det(xI - T)$，且进一步有 $\phi'(\lambda) > 0$。

现在，设 $\mathbf{Y}_n$ 由线性表示 (ξ, μ, η) 定义，并假设矩阵$M = A_1 + \cdots + A_\ell + B$是本原的。记 λ 为 M的佩龙-弗罗贝尼乌斯特征值，设v' 和 u 是 M相对于 λ的严格正左、右特征向量，并将其归一化，使得$v'u = 1$。那么，根据关系式 (11)，有 $M^n = \lambda^n \; uv'(1 + O(\varepsilon^n))$，其中某个 $\varepsilon \in (0, 1)$，并且矩阵 uv' 也可写成式 (12) 的形式。

$\{\mathbf{Y}_n\}$ 的渐近性质取决于如下定义的多元函数 $y(\mathbf{t})$。对于每个 $\mathbf{t} \in \mathbb{R}^\ell$，考虑由下式给出的矩阵 $M(\mathbf{t})$

$$M(\mathbf{t}) = A_1 e^{t_1} + A_2 e^{t_2} + \cdots + A_\ell e^{t_\ell} + B , \tag{13}$$

并令 $y = y(\mathbf{t})$ 为由下列方程定义的隐函数

$$\det(Iy - M(\mathbf{t})) = 0 ,$$

初始条件为 $y(\mathbf{0}) = \lambda$。由于对于每个 $\mathbf{t} \in \mathbb{R}^\ell$， $M(\mathbf{t})$ 都是本原的， $y(\mathbf{t})$ 是其 Perron-Frobenius 特征值，因此对所有 $\mathbf{t} \in \mathbb{R}^\ell$， $y(\mathbf{t}) > 0$ 。因此， $y(\mathbf{t})$ 是一个正函数，在整个 $\mathbb{R}^\ell$上都有定义且解析。

此外，关系式 (11) 和 (12) 可应用于 $M(\mathbf{t})$，这使我们能够证明矩母函数 $\Psi_n(\mathbf{z})$满足以下“准幂”条件；在我们的语境中，这是一项关键性质。

定理 2 给定第 *3* 节中定义的线性表示 (ξ, μ, η)，假设矩阵 $M = A_1 + \cdots + A_\ell + B$ 是本原的，并令 λ 为其佩龙-弗罗贝尼乌斯特征值。那么，存在一个解析严格正函数 $r : \mathbb{R}^\ell \to \mathbb{R}_+$，满足 $r(\mathbf{0}) = 1$，使得对任意 $\mathbf{t} \in \mathbb{R}^\ell$，都有

$$\Psi_n(\mathbf{t}) = r(\mathbf{t}) \left(\frac{y(\mathbf{t})}{\lambda} \right)^n (1 + O(\varepsilon_{\mathbf{t}}^n)) \ , \tag{14}$$

对于某个 $\varepsilon_{\mathbf{t}} \in (0,1)$。

证明。 对矩阵 $M(\mathbf{t})$ 应用关系式 **(11)**，对于任意 $\mathbf{t} \in \mathbb{R}^\ell$，可得

$$M(\mathbf{t})^n = y(\mathbf{t})^n \, u_{\mathbf{t}} v'_{\mathbf{t}} \, (1 + O(\varepsilon_{\mathbf{t}}^n)) \ ,$$

对于某个 $\varepsilon_{\mathbf{t}} \in (0,1)$，其中 $v'_{\mathbf{t}}$ 和 $u_{\mathbf{t}}$ 是 $M(\mathbf{t})$ 相对于 $y(\mathbf{t})$ 的严格正左、右特征向量，使得 $v'_{\mathbf{t}} u_{\mathbf{t}} = 1$。这意味着

$$h_n(\mathbf{t}) = \xi' M(\mathbf{t})^n \eta = y(\mathbf{t})^n \, \xi' u_{\mathbf{t}} v'_{\mathbf{t}} \eta \, (1 + O(\varepsilon_{\mathbf{t}}^n)) \ , \tag{15}$$

因此

$$\Psi_n(\mathbf{t}) = \frac{h_n(\mathbf{t})}{h_n(\mathbf{0})} = \left(\frac{y(\mathbf{t})}{\lambda} \right)^n r(\mathbf{t}) \, (1 + O(\varepsilon_{\mathbf{t}}^n)) \ ,$$

其中 $r(\mathbf{t}) = \frac{\xi' u_{\mathbf{t}} v'_{\mathbf{t}} \eta}{\xi' u v' \eta}$。显然，$r(\mathbf{0}) = 1$。此外，对 $M(\mathbf{t})$ 应用方程 **(12)**，可知 $u_{\mathbf{t}} v'_{\mathbf{t}}$ 的所有元素都是解析函数，这证明 $r(\mathbf{t})$ 也是解析的，并且在整个 $\mathbb{R}^\ell$ 上严格为正。□

现在，当矩阵 M 为本原矩阵时，我们来求 $Y_{n,i}$ 在 $i \in \{1, \ldots, \ell\}$ 下的均值、方差和协方差。下文中，对于任意一对指标 i, j，我们假设 $y'_i(\mathbf{t}) := \partial_i y(\mathbf{t}) = \frac{\partial y}{\partial t_i}(\mathbf{t})$ 且 $y''_{ij}(\mathbf{t}) := \partial_{ij} y(\mathbf{t}) = \frac{\partial^2 y}{\partial t_i \partial t_j}(\mathbf{t})$，并对 $r(\mathbf{t})$ 使用相同的记号。在单变量情形下，即对于 $\ell = 1$，当 M 为本原矩阵时，$\mathbb{E}(\mathbf{Y}_n)$ 和 $\mathrm{Var}(\mathbf{Y}_n)$ 的渐近性质已为人所知，因而可应用于我们的情形。不过，我们在此更倾向于给出一个基于定理 2 的直接证明；该证明将分析扩展到协方差，并使我们能够在后续章节中使用论证中的一些步骤。

对于任意 $\mathbf{t} \in \mathbb{R}^\ell$，我们考虑 $\{h_n(\mathbf{t})\}$ 的生成函数，即

$$\begin{aligned} \sum_{n \geq 0} h_n(\mathbf{t}) w^n &= \xi' \left(\sum_{n \geq 0} M(\mathbf{t})^n w^n \right) \eta = \xi' \left(I - w M(\mathbf{t}) \right)^{-1} \eta \\ &= \frac{\xi' \mathrm{Adj}(I - w M(\mathbf{t}) \eta}{\det(I - w M(\mathbf{t}))} \end{aligned}$$

注意，分母的根是 $M(\mathbf{t})$ 的特征值的倒数。因此，我们有

$$\sum_{n \geq 0} h_n(\mathbf{t}) w^n = \frac{p(w, \mathbf{t})}{\prod_{i=1}^m (1 - w \alpha_i(\mathbf{t}))}$$

其中， $p(w,\mathbf{t})$ 是关于 w 的一个适当多项式，而 $e^{t_i}(i=1,\ldots,m)$ 中的值 $\alpha_i(\mathbf{t})$ 是 $M(\mathbf{t})$ 的特征值。根据 $M(\mathbf{t})$ 的本原性，其中只有一个与 $y(\mathbf{t})>0$ 重合，而其余所有值的模都小于 $y(\mathbf{t})$。由上面的最后一个恒等式可知， $h_n(\mathbf{t})$ 是以下形式的有限项之和

$$cn^k\alpha_i(\mathbf{t})^n$$

其中 $k\in\mathbb{N}$ 和 c 是常数（可能取决于 i 和 k）。因此， $h_n(\mathbf{t})$ 的偏导数也都是相同形式的项的和。这意味着，出现在 (15) 和 (14) 右侧的量 $O(\varepsilon_{\mathbf{t}}^n)$ 满足以下关系

$$\nabla O(\varepsilon_{\mathbf{t}}^n)=O(\varepsilon^n) \tag{16}$$

对于某个 $\varepsilon\in(0,1)$，它可能取决于 $\mathbf{t}$。

4.1 均值

由方程 (9) 可知

$$\mathbb{E}[\mathbf{Y}_n]=\nabla\log\Psi_n(\mathbf{t})_{|\mathbf{t}=\mathbf{0}}=\nabla\log\mathbb{E}[e^{\mathbf{t}\cdot\mathbf{Y}_n}]_{|\mathbf{t}=\mathbf{0}}. \tag{17}$$

应用拟幂方程 (14)，并考虑关系 (16)，可得

$$\begin{aligned}\mathbb{E}[\mathbf{Y}_n]=\nabla\log\Psi_n(\mathbf{t})_{|\mathbf{t}=\mathbf{0}}&=\nabla\log\left(r(\mathbf{t})\left(\frac{y(\mathbf{t})}{\lambda}\right)^n(1+O(\varepsilon_{\mathbf{t}}^n))\right)_{|\mathbf{t}=\mathbf{0}}\\&=\frac{\nabla r(\mathbf{0})}{r(\mathbf{0})}+n\frac{\nabla y(\mathbf{0})}{y(\mathbf{0})}+\nabla\log\left(1+O(\varepsilon_{\mathbf{t}}^n)\right)_{|\mathbf{t}=\mathbf{0}}=\boldsymbol{\beta}n+\mathbf{c}+O(\varepsilon^n),\end{aligned} \tag{18}$$

这意味着

$$\mathbb{E}[\mathbf{Y}_n]=\boldsymbol{\beta}n+\mathbf{c}+O(\varepsilon^n)\ ,\ \text{where}\ \ \boldsymbol{\beta}=\frac{\nabla y(\mathbf{0})}{\lambda}\ \text{and}\ \ \mathbf{c}=\nabla r(\mathbf{0}). \tag{19}$$

此外，根据 [2, 4], 中出现的结果可知，对于每个 $i\in\{1,\ldots,\ell\}$， $y_i'(\mathbf{0})=vA_iu$。这意味着

$$\beta_i=\frac{y_i'(\mathbf{0})}{\lambda}=\frac{v'A_iu}{\lambda}\ ,\quad \text{and hence } 0<\beta_i<1 \text{ and } 0<\sum_{i=1}^{\ell}\beta_i<1. \tag{20}$$

对 $M(\mathbf{t})$ 应用相同的推理，我们还可得出偏导数 $y_i'(\mathbf{t})$ 也是严格正的，即

$$y_i'(\mathbf{t})=v_{\mathbf{t}}'A_ie^{t_i}u_{\mathbf{t}}>0 \tag{21}$$

对于每个 $i\in\{1,\ldots,\ell\}$ 以及每个 $\mathbf{t}\in\mathbb{R}^\ell$。

4.2 协方差

由方程 (10) 可知，对于每个 $i,\ j \in \{1,\dots,\ell\}$，有

$$\begin{aligned}\mathrm{Cov}(Y_{n,i},Y_{n,j}) &= \mathbb{E}[Y_{n,i}Y_{n,j}] - \mathbb{E}[Y_{n,i}]\mathbb{E}[Y_{n,j}] = \\ &= \partial_{ij}\log\mathbb{E}[e^{\mathbf{t}\cdot\mathbf{Y}_n}]|_{\mathbf{t}=\mathbf{0}} = \partial_{ij}\log\Psi_n(\mathbf{t})|_{\mathbf{t}=\mathbf{0}}\ .\end{aligned}$$

Then，应用拟幂关系 (14)，结合 (18) 和 (16)，我们得到 n

$$\partial_{ij}\log\Psi_n(\mathbf{t}) = \partial_j\Big(\frac{\partial_i r(\mathbf{t})}{r(\mathbf{t})} + n\frac{\partial_i y(\mathbf{t})}{y(\mathbf{t})} + O(\varepsilon^n)\Big)\ ,$$

因此

$$H\log\Psi_n(\mathbf{t}) = H\log r(\mathbf{t}) + nH\log y(\mathbf{t}) + O(\varepsilon^n)$$

由此可得,

$$\mathrm{Cov}(\mathbf{Y}_n) = nH\log y(\mathbf{0}) + H\log r(\mathbf{0}) + O(\varepsilon^n) = n\Gamma + C + O(\varepsilon^n)\ ,$$

其中，矩阵 $\Gamma = [\gamma_{ij}]$ 和 $C = [c_{ij}]$ 定义为

$$\begin{aligned}\gamma_{ij} &= \partial_{ij}\log y(\mathbf{t})_{|\mathbf{t}=\mathbf{0}} = \frac{\partial_{ij}y(\mathbf{t})\ y(\mathbf{t}) - \partial_i y(\mathbf{t})\partial_j y(\mathbf{t})}{y(\mathbf{t})^2}_{|\mathbf{t}=\mathbf{0}} = \frac{y''_{ij}(\mathbf{0})}{\lambda} - \frac{y'_i(\mathbf{0})y'_j(\mathbf{0})}{\lambda^2}\ , \qquad (22)\\ c_{ij} &= r''_{ij}(\mathbf{0}) - r'_i(\mathbf{0})r'_j(\mathbf{0})\ .\end{aligned}$$

特别地，每个$Y_{n,i}$ 的方差为

$$\begin{aligned}\mathrm{Var}(Y_{n,i}) &= \gamma_{ii}n + c_{ii} + O(\varepsilon^n)\ ,\ \text{ where}\\ \gamma_{ii} &= \frac{y''_{ii}(\mathbf{0})}{\lambda} - \left(\frac{y'_i(\mathbf{0})}{\lambda}\right)^2 \ \text{ and }\ c_{ii} = r''_{ii}(\mathbf{0}) - (r'_i(\mathbf{0}))^2\ .\end{aligned}$$

此外，在我们的情形下应用 [2, 定理 3]，不难看出，对每个$i \in \{1,\dots\ell\}$都有 $\gamma_{ii} > 0$ 。这意味着，在我们的假设下，对每个 i，$\mathrm{Var}(Y_{n,\ i}) = \Theta(n)$。

显然，由前面的等式可得以下极限:

$$\mathrm{Cov}\left(\frac{Y_{n,i}}{\sqrt{n}},\frac{Y_{n,j}}{\sqrt{n}}\right) \to \gamma_{ij}\ ,\quad \text{for } n\to\infty\ , \qquad (23)$$

特别地,

$$\mathrm{Var}\left(\frac{Y_{n,i}}{\sqrt{n}}\right) \to \gamma_{ii}\ ,\quad \text{for } n\to\infty\ . \qquad (24)$$

5 多元高斯分布极限

在本原假设下，我们证明随机变量 $\frac{\mathbf{Y}^n - n\boldsymbol{\beta}}{\sqrt{n}}$仍具有（多元）高斯极限分布。证明基于对其矩母函数的研究，以及定理 2 中提出的拟幂性质。

首先，我们固定一些记号。对于任意数组 $\mathbf{m} \in \mathbb{R}^\ell$ 和任意对称半正定矩阵 $C \in \mathbb{R}^{\ell\times\ell}$，令 $N(\mathbf{m}\ C)$ 为一个具有 ℓ 个分量、均值为$\mathbf{m}$ 、协方差矩阵为 C的多元高斯随机变量。我们回顾一下 [10]， $N(\mathbf{m}\ C)$ 的矩母函数为

$$\Phi_{\mathbf{m}C}(\mathbf{t}) := \exp\left\{\mathbf{m}'\mathbf{t} + \frac{1}{2}\mathbf{t}'C\mathbf{t}\right\}\ , \qquad \mathbf{t} \in \mathbb{R}^\ell\ .$$

定理 3 设(ξ, μ, η) 是一个线性表示，且矩阵 $M = A_1 + \cdots + A_\ell + B$是本原的，并设 λ 为其佩龙-弗罗贝尼乌斯特征值。另外，设 $\boldsymbol{\beta}$ 和 Γ 如上一节所定义。则$\mathbf{Y}^n_{\sqrt{n}} - n\boldsymbol{\beta}$依分布收敛到随机变量 $N(\mathbf{0}, \Gamma)$。

证明。 首先注意到 Γ 是对称且半正定的，因为它是协方差矩阵的极限（由关系式 (23) 和 (24) 得出）。

接着，记 $M_n(\mathbf{t})$ 为 $\mathbf{Y}^n_{\sqrt{n}} - n\boldsymbol{\beta}$的矩母函数，即

$$M_n(\mathbf{t}) := \mathbb{E}\Big[\exp\Big\{\mathbf{t}'\frac{\mathbf{Y}_n - n\boldsymbol{\beta}}{\sqrt{n}}\Big\}\Big]\ .$$

只需证明对于任意$\mathbf{t} \in \mathbb{R}^\ell$， $n \to \infty$, $M_n(\mathbf{t})$ 收敛到 $\exp\left\{\frac{1}{2}\mathbf{t}'\Gamma\mathbf{t}\right\}$，即可得到该结果（例如参见 [10, 8]）。为此，直接应用定理 2，可得

$$\begin{aligned} M_n(\mathbf{t}) &= \Psi\left(\frac{\mathbf{t}'}{\sqrt{n}}\mathbf{Y}_n\right)\cdot\exp\{-\mathbf{t}'\boldsymbol{\beta}\sqrt{n}\} \\ &= r(\mathbf{t}/\sqrt{n})\left(\frac{y(\mathbf{t}/\sqrt{n})}{\lambda}\right)^n\left(1 + O(\varepsilon^n_{\mathbf{t}/\sqrt{n}})\right)\exp\{-\mathbf{t}'\boldsymbol{\beta}\sqrt{n}\} \\ &= \exp\left\{\log r(\mathbf{t}/\sqrt{n}) + n\log\frac{y(\mathbf{t}/\sqrt{n})}{\lambda} + O(\varepsilon^n_{\mathbf{t}/\sqrt{n}}) - \mathbf{t}'\boldsymbol{\beta}\sqrt{n}\right\}\ . \end{aligned}$$

回想起 $r(\mathbf{t})$ 在 $\mathbf{0}$ 处是解析的， $r(\mathbf{0}) = 1$ 且 $\boldsymbol{\beta} = \nabla y(\mathbf{0})/\lambda$，通过对前述各项进行适当的一阶和二阶展开，我们得到

$$\begin{aligned} M_n(\mathbf{t}) &= \exp\left\{\frac{\mathbf{t}'}{\sqrt{n}}\nabla r(\mathbf{0}) + O(1/n) + n\left[\frac{\mathbf{t}'}{\sqrt{n}}\frac{\nabla y(\mathbf{0})}{\lambda} + \frac{1}{2}\frac{\mathbf{t}'}{\sqrt{n}}\Gamma\frac{\mathbf{t}'}{\sqrt{n}} + O(n^{-3/2})\right] - \mathbf{t}'\boldsymbol{\beta}\sqrt{n}\right\} \\ &= \exp\left\{\frac{1}{2}\mathbf{t}'\Gamma\mathbf{t} + O(1/\sqrt{n})\right\}, \end{aligned}$$

由此证明得证。 □

6 大偏差结果

本节证明，若矩阵 M 是本原的，则序列 $\{\mathbf{Y}_n/n\}$ 满足大偏差原理，其速度为 n。这意味着同一序列几乎必然收敛到 $\boldsymbol{\beta}$ 。

定理 4 设(ξ,μ,η)是一个线性表示，且矩阵 $M=A_1+\cdots+A_\ell+B$是本原的；设 λ是其佩龙-弗罗贝尼乌斯特征值。则序列 $\{\mathbf{Y}_n/n\}$ 满足大偏差原理，其速度为 n，良好速率函数为

$$G^*(\mathbf{x}):=\sup_{\mathbf{t}\in\mathbb{R}^\ell}\left\{\mathbf{t}\cdot\mathbf{x}-\log\frac{y(\mathbf{t})}{\lambda}\right\}. \tag{25}$$

此外，$\{\mathbf{Y}_n/n\}$ 几乎必然收敛到 $\boldsymbol{\beta}$。

证明。为了应用定理1，我们来考察函数

$$G(\mathbf{t}):=\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}\left[e^{n\frac{\mathbf{t}\cdot\mathbf{Y}_n}{n}}\right]\ ,\qquad\forall\ \mathbf{t}\in\mathbb{R}^\ell\ ,$$

由定理 2 可得:

$$\begin{aligned}G(\mathbf{t}) &= \lim_{n\to\infty}\frac{1}{n}\log\Psi_n(\mathbf{t})\\ &= \lim_{n\to\infty}\frac{1}{n}\log\Big(r(\mathbf{t})\Big(\frac{y(\mathbf{t})}{\lambda}\Big)^n(1+O(\varepsilon_{\mathbf{t}}^n))\Big)\\ &= \lim_{n\to\infty}\Big(\frac{\log r(\mathbf{t})}{n}+\log\frac{y(\mathbf{t})}{\lambda}+\frac{\log(1+O(\varepsilon_{\mathbf{t}}^n))}{n}\Big)\\ &= \log\frac{y(\mathbf{t})}{\lambda}=\log\frac{y(\mathbf{t})}{y(\mathbf{0})}.\end{aligned}$$

显然，G在整个 $\mathbb{R}^\ell$上都是有限值且可微的。因此，正如注3所述，G 是本质光滑的。由定理1可知，序列$\{\mathbf{Y}_n/n\}$ 满足大偏差原理，其速度为 n，良好速率函数为 $G^*(\mathbf{x})$，由式（25）给出。

现在注意到，关系式（19）意味着 $\nabla G(\mathbf{0})=\frac{\nabla\mathbf{y}(\mathbf{0})}{\lambda}=\boldsymbol{\beta}$。因此，由注1和第 2 节的第 3 点可知，$\{\mathbf{Y}_n/n\}$几乎必然收敛到 $\boldsymbol{\beta}$，因为其大偏差原理的速度为 n。□

如第 2 节第 2 点所述，根据我们关于线性表示（ξ,μ,η）的假设（在第 3 节中确立），上面给出的函数 G^* 可能仅在集合中取有限值

$$C_\bullet=\{\mathbf{x}=(x_1,\ldots,x_\ell)\in\mathbb{R}^\ell:x_1,\ldots,x_\ell\geq 0,x_1+\cdots+x_\ell\leq 1\},$$

而 $G^*(\mathbf{x})=+\infty$ 对每个 $\mathbf{x}\in C_\bullet^c$ 都成立。这也与注2一致；事实上，对于每个 $\mathbf{t}\in\mathbb{R}^\ell$ 以及每个 $i\in\{0,1,\ldots\ell\}$，应用 (21) 可得

$$\frac{\partial G}{\partial t_i}(\mathbf{t})=\frac{y_i'(\mathbf{t})}{y(\mathbf{t})}=\frac{v_{\mathbf{t}}'A_ie^{t_i}u_{\mathbf{t}}}{v_{\mathbf{t}}'M(\mathbf{t})u_{\mathbf{t}}}\ \in\ (0,1)\ .$$

特别地，由关系式 (13) 可知，由于 $B\neq[0]$ 以及 $v_{\mathbf{t}}$ 和 $u_{\mathbf{t}}$ 的所有分量都严格为正，这意味着

$$0<\sum_{i=1}^{l}\frac{\partial G}{\partial t_i}(\mathbf{t})<1\ ,\quad\forall\ \mathbf{t}\in\mathbb{R}^\ell\ .$$

7 中度偏差结果

本节基于前文使用的相同本原假设，证明一个中等偏差结果。这项工作完全是原创的，并且将单变量情形$\ell = 1$ 作为一个特例纳入适用范围。正如第 2 节所述，该结果在两种收敛行为之间架起了一座数学桥梁：$\mathbf{Y}_n/n$ 几乎必然收敛到一个常数，以及$\frac{\mathbf{Y}_n - \boldsymbol{\beta} n}{\sqrt{n}}$弱收敛到一个中心高斯随机变量，或者以其等价形式 $\sqrt{n}(\mathbf{Y}_n/n - \boldsymbol{\beta})$ 弱收敛到一个中心高斯随机变量。

定理 5 假设定理 *4* 的相同条件成立，并令 $\boldsymbol{\beta}$ 和 Γ 分别由 (20)和 (22)定义（即 $\boldsymbol{\beta} = \nabla y(\mathbf{0})$ 以及 $\Gamma = H \log y(\mathbf{0})$）。那么，对于满足以下条件的任意正实数序列 $\{a_n\}$

$$a_n \to 0 \qquad and \qquad n a_n \to \infty \ ,$$

我们有：对$\mathbf{Y}_n - n\boldsymbol{\beta}$ $\sqrt{n/a_n}$，以 $1/a_n$ 为速度、速率函数为

$$J^*(\mathbf{x}) = \sup_{\mathbf{t}} \Big\{ \mathbf{t} \cdot \mathbf{x} - \frac{1}{2} \mathbf{t}' \Gamma \mathbf{t} \Big\} \ .$$

证明。 对于任意 $n \in \mathbb{N}$ 和任意$\mathbf{t} \in \mathbb{R}^\ell$，令

$$J_n(\mathbf{t}) := \frac{1}{1/a_n} \log \mathbb{E}\Big[\exp\Big\{ \frac{\mathbf{t}'}{a_n} \frac{\mathbf{Y}_n - n\boldsymbol{\beta}}{\sqrt{n/a_n}} \Big\} \Big] \ .$$

为计算 $\lim_{n\to\infty} J_n(\mathbf{t})$，我们应用定理 2。对于每个$\mathbf{t} \in \mathbb{R}^\ell$，可得

$$\begin{aligned}
J_n(\mathbf{t}) &= a_n \log \mathbb{E}\Big[\exp\Big\{ \frac{\mathbf{t}'(\mathbf{Y}_n - n\boldsymbol{\beta})}{\sqrt{n a_n}} \Big\} \Big] \\
&= a_n \Big(\log \mathbb{E}\Big[\exp\Big\{ \frac{\mathbf{t}' \mathbf{Y}_n}{\sqrt{n a_n}} \Big\} \Big] - \frac{\mathbf{t}' n \boldsymbol{\beta}}{\sqrt{n a_n}} \Big) \\
&= a_n \Big(\log \Big[r(\mathbf{t}/\sqrt{n a_n}) \Big(\frac{y(\mathbf{t}/\sqrt{n a_n})}{\lambda} \Big)^n \Big(1 + O(\varepsilon^n_{\mathbf{t}/\sqrt{n a_n}}) \Big) \Big] - \frac{\mathbf{t}' n \boldsymbol{\beta}}{\sqrt{n a_n}} \Big) \\
&= \underbrace{a_n \log r(\mathbf{t}/\sqrt{n a_n})}_{(*)} + \underbrace{a_n \log \Big(1 + O(\varepsilon^n_{\mathbf{t}/\sqrt{n a_n}}) \Big)}_{(**)} + \underbrace{n a_n \Big(\log \frac{y(\mathbf{t}/\sqrt{n a_n})}{\lambda} - \frac{\mathbf{t}' \boldsymbol{\beta}}{\sqrt{n a_n}} \Big)}_{(***)} \ .
\end{aligned}$$

由于 $r(\mathbf{t})$在$\mathbf{0}$ 附近解析，且 $r(\mathbf{0}) = 1$，根据关于 a_n 的假设，显然最后一个表达式中的项 $(*)$ 当 $n \to \infty$时收敛于 0。此外，$(**)$也趋于 0，因为$\mathbf{t}/\sqrt{n a_n}$ 趋近于$\mathbf{0}$，因此 $(**) = O(a_n \epsilon^n)$，其中某个 $\epsilon \in (0,1)$接近 $\epsilon_{\mathbf{0}}$。

现在我们利用 $\log y(\mathbf{s}) \lambda$ 的麦克劳林展开直到二阶，证明 $(***)$ 趋于 $\frac{1}{2}\mathbf{t}' \Gamma \mathbf{t}$，

二阶:

$$\begin{aligned}(***) &= na_n\Big(\log \underbrace{\frac{y(\mathbf{0})}{\lambda}}_{=\lambda/\lambda=1} + \frac{\mathbf{t}'}{\sqrt{na_n}}\underbrace{\frac{\nabla y(\mathbf{0})}{y(\mathbf{0})}}_{\boldsymbol{\beta}} + \frac{1}{2}\frac{\mathbf{t}'}{\sqrt{na_n}}\Gamma\frac{\mathbf{t}}{\sqrt{na_n}} + o\Big(\frac{1}{na_n}\Big) - \frac{\mathbf{t}'\boldsymbol{\beta}}{\sqrt{na_n}}\Big)\\ &= na_n\Big(\frac{\mathbf{t}'\boldsymbol{\beta}}{\sqrt{na_n}} + \frac{1}{2na_n}\mathbf{t}'\Gamma\mathbf{t} + o\Big(\frac{1}{na_n}\Big) - \frac{\mathbf{t}'\boldsymbol{\beta}}{\sqrt{na_n}}\Big)\\ &= \frac{1}{2}\mathbf{t}'\Gamma\mathbf{t} + o(1)\ .\end{aligned}$$

由于函数 $J(\mathbf{t}) = \lim_{n\to\infty} J_n(\mathbf{t}) = \frac{1}{2}\mathbf{t}'\Gamma\mathbf{t}$ 对每个 $\mathbf{t} \in \mathbb{R}^\ell$ 都是有限且可微的，我们可以应用定理1并得到结果。□

备注4 如果 Γ 可逆（即 $det(\Gamma) \neq 0$），那么 $J^*(\mathbf{x})$ 中的 sup 在满足$\mathbf{x} = \Gamma\mathbf{t}$的 $\mathbf{t}$ 处取得，也就是在 $\mathbf{t} = \Gamma^{-1}\mathbf{x}$处。因此，回想起 $(\Gamma^{-1})' = (\Gamma^{-1})$，我们有

$$\begin{aligned}J^*(\mathbf{x}) &= (\Gamma^{-1}\mathbf{x})'\mathbf{x} - \frac{1}{2}(\Gamma^{-1}\mathbf{x})'\Gamma\Gamma^{-1}\mathbf{x}\\ &= \mathbf{x}'(\Gamma^{-1})'\mathbf{x} - \frac{1}{2}\mathbf{x}'(\Gamma^{-1})'\mathbf{x}\\ &= \frac{1}{2}\mathbf{x}\Gamma^{-1}\mathbf{x}.\end{aligned}$$

8 结论

在本文中，我们研究了一个多元随机变量序列 $\{\mathbf{Y}_n\}$ 的渐近行为；该序列表示随机生成的长度为 n的字符串中，一组符号的出现次数，这些字符串按照有理随机模型生成。我们考虑了标准情形，即模型中转移总权重的矩阵是本原的。在这一假设下，对于某个 $\boldsymbol{\beta} \in \mathbb{R}^\ell$，有 $\mathbb{E}[\mathbf{Y}_n] = \boldsymbol{\beta}n + o(n)$，并证明了以下结果：

- $\frac{\mathbf{Y}_n - \boldsymbol{\beta}n}{\sqrt{n}} = \sqrt{n}\big(\frac{\mathbf{Y}_n}{n} - \boldsymbol{\beta}\big)$ 依分布收敛于一个协方差矩阵为 Γ的中心高斯随机变量；
- $\{\mathbf{Y}_n/n\}_{n\geq 1}$ 满足速度为 n的大偏差原理，这意味着$\mathbf{Y}_n/n$几乎必然收敛于 $\boldsymbol{\beta}$；
- 对于每个满足 $a_n \to 0$ 和 $na_n \to \infty$的正数列 $\{a_n\}_{n\geq 1}$，$\{\sqrt{na_n}(\mathbf{Y}_n/n - \boldsymbol{\beta})\}_{n\geq 1}$ 满足速度为 $1/a_n$ 且具有适当良好速率函数 J的大偏差原理。

本分析使用的主要工具包括：根据本原矩阵的性质推导出的$\mathbf{Y}_n$的矩母函数拟幂定理，以及经典的 G¨artner-Ellis 定理；该定理非常适用于我们的情形。

为完整起见，有必要确定上述协方差矩阵 Γ何时是（严格）正定的。事实上，只有在这种情况下，极限高斯变量

才会在 $\mathbb{R}^\ell$ 上具有全支撑和正密度。尽管我们通常知道 Γ 的对角元素严格为正，但这并不能保证矩阵是正定的。

另一个值得进一步研究的自然问题是，之前的结果能否推广到转移权重矩阵不是本原（也不是不可约），但具有两个或更多（严格连通）分量的情形。在这些情形下，已经得到了不同的极限分布 [4, 12, 11]，人们可能会想知道，在什么条件下大偏差原理仍然成立。

9 附录：一些已知结果

本节回顾一些与大偏差有关的已知结果，这些结果在前文中用到，但在文献中往往没有明确提及，并且可能对读者有所帮助。这些性质尤其涉及速率函数零点的唯一性，以及依概率或几乎必然收敛到常数。

9.1 速率函数的唯一零点

命题 1 如果 *G¨artner-Ellis* 定理适用（即定理*1*），那么速率函数Λ^* 仅在 $\mathbf{x} = \nabla\Lambda(\mathbf{0})$ 处为零。

证明。容易验证，速率函数 Λ^* 在 $\mathbf{x} = \nabla\Lambda(\mathbf{0})$ 处为零。现在我们用反证法，假设 Λ^* 在不止一个点处等于零。那么，由于其凸性，函数 Λ^* 在一个区间上恒等于零；因此函数 Λ^* 不是严格凸的。接着，根据凸分析中的一个标准论证（参见例如 [19, 定理 26.3]），函数 Λ 将不可微。这与 G¨artner-Ellis 定理的假设矛盾，故证明完成。□

9.2 收敛到速率函数的唯一零点

假设 I 是定义在 $\mathbb{R}^m$ 上的一个良好速率函数，并定义 $I(S) := \inf\{I(x) : x \in S\}$，对任意 $S \subseteq \mathbb{R}^m$。

引理1 设 $S \subseteq \mathbb{R}^m$ 是一个闭集，满足 $I(S) < \infty$。则 $I(S)$ 在某个值 $x^* \in S$ 处取到，即存在 $x^* \in S$ 使得 $I(x^*) = I(S)$。

证明。当存在（有限）下确界时，总可以构造一个由点组成的极小化序列；因此，设 $\{x_n\}$ 是 S 中的一个点序列，使得

$$\lim_{n\to\infty} I(x_n) = I(S).$$

Then，对于任意 $\varepsilon > 0$，该序列中的点最终都属于水平集

$$K_{S,\varepsilon} = \{x : I(x) \le I(S) + \varepsilon\}$$

根据假设，该水平集是紧集（因为 I 是一个良好 速率函数）。因此，该序列中的点最终属于 $S \cap K_{S,\ \varepsilon}$，这是一个紧集（因为它是闭集与紧集的交集）。因此，我们可以抽取一个子序列 $\{x_{n_h}\}$，该子序列收敛到一点 $x^* \in S \cap K_{S,\ \varepsilon}$（根据紧性性质）。通过验证 $I(x^*) = I(S)$，即可完成证明。事实上，$x^* \in S$ 是由构造得出的；此外，根据 I 的下半连续性，我们有

$$\liminf_{n_h \to \infty} I(x_{n_h}) \geq I(x^*),$$

由此可得 $I(S) \geq I(x^*)$（子序列的极限与原序列的极限相同）；最后，由于 $x^* \in S$，我们得到 $I(x^*) = I(S)$。□

推论1 假设满足大偏差原理，且其良好速率函数为 I。那么 I 至少在一点处为零。

证明。考虑大偏差原理定义中的上界 (2)，我们得到

$$\limsup_{n \to \infty} \frac{1}{v_n} \log P(X_n \in \mathbb{R}^m) \leq -I(\mathbb{R}^m).$$

因此 $0 \leq -I(\mathbb{R}^m)$，这意味着 $I(\mathbb{R}^m) \leq 0$；又由于速率函数 I 非负，我们得到 $I(\mathbb{R}^m) = 0$。因此，根据前面的引理，至少存在一点 x^*，使得$I(x^*) = 0$。□

现在我们给出一个结果：依概率收敛到良好速率函数的唯一零点。

命题2 假设我们有一列随机变量 $\{X_n\}$，它们定义在同一个概率空间 $(\Omega\ \mathcal{F}, P)$ 上，满足大偏差原理，且其良好速率函数为 I，该函数在唯一一点 x_0 处为零。那么 X_n 依概率收敛到 x_0。

证明。设 $B_\delta(x_0)$ 为以 x_0 为中心、半径为 $\delta > 0$ 的开球。我们针对闭集 $(B_\delta(x_0))^c$ 考虑大偏差原理定义中的上界 (2)，从而得到

$$\limsup_{n \to \infty} \frac{1}{v_n} \log P(X_n \in (B_\delta(x_0))^c) \leq -I((B_\delta(x_0))^c).$$

于是，对每一个 $\eta > 0$，我们最终有

$$P(X_n \in (B_\delta(x_0))^c) \leq \exp(-v_n(I((B_\delta(x_0))^c) - \eta)).$$

最后，我们证明 $P(X_n \in (B_\delta(x_0))^c) \to 0$。若 $I((B_\delta(x_0))^c) = \infty$，则显然成立。否则，若 $I((B_\delta(x_0))^c) < \infty$，由引理1可知，存在 $x^* \in (B_\delta(x_0))^c$，使得$I(x^*) = I((B_\delta(x_0))^c)$，且 $I(x^*) > 0$（因为 $x^* \in (B_\delta(x_0))^c$，故 $x^* \neq x_0$，其中 x_0 是 I 的唯一零点）。总之，我们有 $P(X_n \in (B_\delta(x_0))^c) \to 0$，因为我们可以选择$\eta \in (0, I((B_\delta(x_0))^c))$。□

值得注意的是，沿用上述命题证明中的记号，标准地应用博雷尔-坎特利引理可得以下结论

推论 2 *If* $\sum_{n\geq 1}\exp(-v_n(I((B_\delta(x_0))^c)-\eta))<\infty$，则 X_n 几乎必然收敛到 x_0.

注意，当 $v_n=n$ 时，上述推论显然成立。

最后，我们指出，如果 I 不是一个良好速率函数，我们就无法得出与命题相同的结论。为说明这一点，考虑以下反例：令 I 具有以下形式

$$I(x):=\begin{cases}\frac{1}{1+x^2} & \text{if } x\neq 0\\ 0 & \text{if } x=0.\end{cases}$$

它是一个速率函数（事实上，它是下半连续的），且仅在 $x_0=0$ 处取零；然而，对于任意 $\delta>0$，我们有 $I((B_\delta(x_0))^c)=0$。

经费

M.G. 和 M.V. 感谢米兰大学对 PSR 2024 项目的部分资助。C.M. 和 E.V. 感谢 INdAM-GNAMPA 的部分资助。C.M. 还感谢 MUR 卓越系项目对罗马第二大学数学系的资助（CUP E83C23000330006），以及罗马第二大学 METRO 研究项目的部分资助（CUP E83C25000630005）。

参考文献

[1] J. Berstel 和 C. Reutenauer。《有理级数及其语言》，施普林格出版社，纽约－海德堡－柏林，1988年。

[2] A. Bertoni、C. Choffrut、M. Goldwurm 和 V. Lonati。关于正则语言中的词内某个符号出现次数。理论计算机科学，302:431–456，2003年。

[3] J. Bourdon 和 M. R´egnier。模式的大偏差性质。离散算法杂志，24:2–11，2014年。

[4] D. de Falco、M.Goldwurm 和 V. Lonati。双分量随机模型中的符号出现频率。理论计算机科学，327 (3):269–300，2004。

[5] A. Dembo 和 O.Zeitouni。大偏差技巧与应用。第二版（1998）的修订重印本。施普林格出版社，2010。

[6] F. den Hollander。大偏差。美国数学会，2000。

[7] A. Denise 和 M. R´egnier。随机字符串上的稀有事件与条件事件。离散数学与理论计算机科学，6:191–214，2004。

[8] P. Flajolet 和 R. Sedgewick。《解析组合学》。剑桥大学出版社，2009。

[9] P. Flajolet、W. Szpankowski 和 B. Vall´ee。隐藏词统计。《美国计算机学会会刊》，53:147–183，2006。[10] B.V. Gnedenko。概率论。Gordon and Breach Science Publ.，1997。[11]M. Goldwurm、J. Lin 和 M. Vignati。本原有理模型中的符号统计局部极限定律。理论计算机科学，970:114051，2023，https://doi.org/10.1016/j.tcs.2023.114051 。[12] M. Goldwurm 和 V. Lonati。模式统计与范德蒙矩阵。理论计算机科学，356:153–169，2006。[13]M. Goldwurm 和 M. Vignati。本原有理模型中模式统计的大偏差性质。理论计算机科学，1030:115055，2025，https://doi.org/10.1016/j.tcs.2024.115055 。[14] M. Goldwurm 和 M. Vignati。符号统计大偏差速率函数分析。《自动机、语言与组合学杂志》30(1-3):127-156，2025。[15] H.-K. Hwang。组合分布的大偏差。I：中心极限定理。《应用概率年刊》，6(1):297–319，1996。[16] P. Jacquet 和 W. Szpankowski。《解析模式匹配：从 *DNA* 到 *Twitter*》剑桥大学出版社，2015。[17] P. Nicodeme、B. Salvy 和 P. Flajolet。基元统计。理论计算机科学，287(2): 593–617，2002。[18] M. R´egnier 和 W. Szpankowski。关于马尔可夫序列中的模式频率出现次数。《算法》，22 (4):621–649，1998。[19] R. T. Rockafellar。凸分析，普林斯顿大学出版社，1970。[20] A. Salomaa 和 M. Soittola。《形式幂级数的自动机理论方面》。施普林格出版社，1978。[21] E. Seneta。非负矩阵与马尔可夫链。施普林格出版社，纽约、海德堡、柏林，1981。[22] H. Touchette，勒让德-芬切尔变换浅析，arXiv:cond-mat/0502123，2005。[23] P. Turakainen。广义自动机与随机语言。《美国数学会会刊》，21: 303–309，1969。