arXiv:2602.07665v1 [math.ST] 7 Feb 2026

闭单纯形上的费舍尔得分

Giovanni Pistone
Statistics Initiative Collegio
Carlo Alberto 都灵 10122，意大利
giovanni.pistone@carloalberto.org

Fabio Rapallo
经济系
热那亚大学
意大利，热那亚 16126
fabio.rapallo@unige.it

Eva Riccomagno数学系
热那亚大学 意大利，热那亚
16146
eva.riccomagno@unige.it

摘要

我们将有限状态统计模型的经典解析工具加以扩展，使其能够处理零概率。借助代数统计学和信息几何的方法，我们建立了一个框架，使光滑统计模型能够到达单纯形的边界，例如含有非结构零的列联表。我们方法的核心对象是向量丛，其纤维是与每个概率分布 p 相关联的p-对比。在这一框架中，费舍尔得分以及其他关键统计概念（例如一维统计模型的熵）也能在单纯形边界上获得代数表示。

关键词 代数统计学 · 对比空间 · 信息几何 · 单参数指数族 · 得分函数

1 引言

本文研究分类数据的统计模型，特别着重于定义在概率单纯形上的单参数模型。我们集中讨论可在指数族内表示的模型，并利用其解析与几何结构。在此设定下，每个模型，即离散概率分布的一维族，都可自然地表示为单纯形上的一条曲线。这一几何观点为系统研究所考虑模型的统计、几何与代数性质提供了统一框架。

对这些曲线的分析基于对模型曲线的切空间与对比空间之间关系的研究，后者定义为编码分类模型中相关对数线性变异的向量空间（例如，参见例2.1中对 2×2 列联表的统计模型的详细说明）。这两个空间之间的联系，是刻画模型局部几何、并对由指数参数化诱导的微分结构给出统计学解释的基本工具。由此看来，本文工作属于代数统计学与信息几何相互作用的更广阔背景，有助于更清晰地理解如何将这两个领域的概念与方法共同用于统计模型的研究Gibilisco等人 [2010], Heaton与Himmelmann [2025], Dufresne等人 [2018], Duarte等人 [2021],Kobayashi与Wynn [2014]。

现有文献大多关注开单纯形的情形，即每个类别的概率都严格为正。然而，当可能出现结构零或观察到零单元格时，就有必要考虑闭单纯形。在组成数据分析框架中，这种情况具有重要意义，例如参见 Aitchison

MSC2020：62B11, 62R01, 62H17

[1986], 其中某些“成分”可能在某些组成中缺失，但不应因此从分析中排除。事实上，组成数据的经典几何方法以对数比变换为基础，而这种变换在存在结构零时定义不明确。Martin-Fernández 等人 [2011]介绍了对含零组成数据的一些扩展。

处理列联表中结构零的一种方法是将表格向量化，并排除对应于结构零的单元格。关系模型中的 Klimova 等人 [2012], Klimova 和 Rudas [2022] 以及代数对数线性模型中的 Pistone 等人 [2001],Drton 等人 [2009]都采用了这种方法。本文考虑一个略有不同的问题，即某些参数取值会使单参数统计模型落在单纯形边界上。本文的主要贡献是将开单纯形上的单参数指数模型的几何分析扩展到闭单纯形，从而提供一种连贯的处理方式，也涵盖边界情形。为此，本文提出对传统上在微分几何方法框架下（即在开单纯形中）考察的对象进行代数化转译。将费舍尔得分推广到闭单纯形是关键一步，这通过利用切空间与对比空间之间的关系得以实现。

开概率单纯形的信息几何最初由 Amari 和 Nagaoka [2000], 定义，研究的是一个仿射空间。根据 Weyl [1952], 的定义，仿射空间通过将位移定义为从一对点到向量空间的映射来刻画。我们在此使用 Aitchison[1986] 和 Pistone 和 Sempi [1995] 引入的位移 $(p,\ q) \mapsto \log \frac{q}{p} - \mathbb{E}_p\left[\log \frac{q}{p}\right]$。仿射空间使严谨的微积分成为可能，也使我们能够将统计模型视为可微簇来讨论。这种微分结构与 Pistone 等人 [2001] 开创的代数统计学方法十分契合。事实上，单项式统计模型及其环面对应模型在仿射坐标中表示为仿射子空间。诸如 Segre 簇和 Veronese 簇之类的仿射对象具有明确的统计学解释。反过来，这种抽象构造恰好提供了吉布斯统计物理学和费舍尔统计方法所需的数学基础，直至 Efron 和 Hastie [2016] 所述的《计算机时代的统计方法》。目前，关于这一主题已有大量文献。

本文结构如下。第 2 节讨论切空间与对比空间之间的关系，指出该过程的代数化是通过对比空间实现的，并取决于支撑集上的一个条件。特别是，将对比空间表示为坐标形式后，便可推导出线性方程。在第 3 节中，我们证明，即使在单纯形边界上，费舍尔得分也是对比。我们在第 4 节利用这一结果，将统计丛扩展到闭单纯形上。第 5 节和第 6 节借助这一扩展，研究数理统计和信息几何中一些经典对象在边界处的表现，例如自然梯度和平行移动。在总结性的第 7 节中，我们概述本文的主要发现，并展望未来的研究方向。

2 切丛与对比空间

为统一记号，我们考虑一个含有 d 个元素的有限集 Ω，以及概率向量 $p = (p(x) : x \in \Omega)$。例如，我们可以设有一个包含有限个 d 类别的分类随机变量 X，并从 X 中抽取一个随机样本，该样本恰好落入类别 x 的概率为 $p(x) = \text{Probability}(X = x)$。

定义 2.1。 设 Ω 是一个含有 d 个元素的有限集，且 $\mathbb{R}$ 是实数集。定义 $L(\Omega) = \mathbb{R}^\Omega$ 为 Ω 上实值函数的空间，等价地可视为 $\mathbb{R}^d$。

1. 概率单纯形是由分量非负且和为一的 d 维实值向量构成的集合

$$\Delta(\Omega) = \left\{ p \in \mathbb{R}^\Omega \,\middle|\, p(x) \geq 0, \sum_{x \in \Omega} p(x) = 1 \right\}$$

2. 对比空间 $C(\Omega)$ 是零和向量的集合

$$C(\Omega) = \left\{ p \in \mathbb{R}^\Omega \,\middle|\, \sum_{x \in \Omega} p(x) = 0 \right\}$$

3. 对于 $p \in \Delta(\Omega)$， p-对比的空间定义为

$$\left\{ u \in \mathbb{R}^\Omega \,\middle|\, \mathbb{E}_p[u] = \sum_{x \in \Omega} p(x)u(x) = 0 \right\}$$

定义 2.1 的第 3 项用于第 3 节。这里我们集中讨论完整单纯形。嵌入 $\Delta(\Omega) \subset \mathbb{R}^d$通过位移函数 $(p, q) \mapsto q - p \in C(\Omega)$ 为单纯形赋予仿射空间的结构，如下

$$\sum_{x \in \Omega} (q(x) - p(x)) = 0$$

仿射结构自然引出了速度向量空间的概念（例如参见 [Schwartz, 1981, Ch. 3]）。

与本文相关的内容如下。给定参考点 $p \in \Delta(\Omega)$ 和一条可微曲线 $t \mapsto \gamma(t) \in \Delta(\Omega)$，该曲线相对于原点 p 的表达式为 $t \mapsto \gamma_p(t) = \gamma(t) - p$，它在向量空间 $C(\Omega)$ 中演化，因为 $\sum_{x \in \Omega} (\gamma(x; t) - p(x)) = 0$，其中 $\gamma(x; t)$ 是 $\gamma(t) \in \mathbb{R}^\Omega$ 的第 x 个分量。曲线 γ 被视为一个单参数统计模型，而 γ_p 是相对于参考点 p 的对比。γ 和γ_p的速度定义为它们关于 t的一阶导数，二者相等，即 $\dot{\gamma}_p(t) = \dot{\gamma}(t)$，且不依赖于 p。在本文讨论的更一般框架中，我们会遇到速度表达式 $t \mapsto \gamma_p(t)$ 依赖于参考点的情形。因此，我们给出以下速度空间的定义；如下文定理 2.1 所示，该空间与对比空间一致。

对于可微曲线 $t \mapsto \gamma(t) \in \Delta(\Omega)$，设 Ω 是包含零的实数轴上的一个开区间，且对该开区间内的每个 t，都有 $\sum_{x \in \Omega} \gamma(x; t) = 1$ 和 $\sum_{x \in \Omega} \dot{\gamma}(x; t) = 0$。

定义 2.2。

1. $p \in \Delta(\Omega)$ 处的切空间是 $p \in \Delta(\Omega)$ 处速度的集合，定义为$T_p\Delta(\Omega) = \left\{ v \in \mathbb{R}^\Omega \,\middle|\, \text{存在可微的 } \gamma: t \mapsto \gamma(t) \in \Delta(\Omega),\ \gamma(0) = p,\ \gamma \in C^1,\ \dot{\gamma}(0) = v \right\}$
2. 将每个点 p 处的切空间汇集起来，构成切丛，定义为

$$T\Delta(\Omega) = \{(p, v) \in \Delta(\Omega) \times T_p\Delta(\Omega)\}$$

定义 2.2 是微分几何中一种一般结构的非常特殊的情形；例如参见 Lang [1995]。给定流形 M，所有经过 $p \in M$ 的曲线的速度所构成的向量空间 T_pM 称为 p处的切空间，而切丛 TM 是所有切空间构成的丛。在每个 p 处，速度向量空间称为 p 处的纤维，记作 T_pM。

定理 2.1。令 支撑集 $\mathrm{p}\, v$ 为向量 $v \in \Omega$的支撑集，即 Ω 中 v 具有非零分量的子集 *t. Then*

$$T\Delta(\Omega) = \left\{ (p, v) \,\middle|\, p \in \Delta(\Omega), v \in \mathbb{R}^\Omega, \sum_{x \in \Omega} v(x) = 0, \operatorname{Supp} v \subseteq \operatorname{Supp} p \right\}, \tag{1}$$

证明。我们区分 p 是否位于单纯形边界上。如果 p 不在单纯形边界上（也就是说，对所有 $x \in \Omega$，$p(x) > 0$），那么对任意对比 v，线性曲线 $\gamma(t) = p + tv$ 在零的某个邻域内对任意 t都属于单纯形，且 $\gamma(0) = p$、$\dot{\gamma}(0) = v$；因此，数对 (p, v) 属于 $T_p\Delta(\Omega)$。（这定义了信息几何中所用的单纯形内部的切丛（例如参见 Pistone [2020]）。）类似的论证也适用于整个单纯形。如果 p 位于单纯形边界上，并且 v 是一个对比，使得只要 $p(x) = 0$，就有 $v(x) = 0$，那么 $\gamma(t) = p + tv$ 在零的某个邻域内对任意 t 都属于单纯形，且 $\gamma(0) = p$、$\dot{\gamma}(0) = v$；因此，数对 (p, v) 也属于 $T_p\Delta(\Omega)$。

□

定理 2.1 给出了切丛的一种等价定义。事实上，方程 (1) 无须借助速度这一概念，就给出了切丛的定义。它表明，仅凭对比空间不足以定义 $T\Delta(\Omega)$ 中的速度；速度的定义还取决于其作用位置 p 是否位于边界上。关键之处在于纳入与所有开面相关的对比空间之并，也就是关于支撑集的条件。

如果 p 是一个顶点，那么方程 (1) 中的条件仅在向量 v 的所有分量都等于零时才成立，这对应于一个将概率 1 赋给顶点 p 的退化统计模型。事实上，不存在经过顶点的可微曲线。

方程 (1) 给出的切丛定义很有意思，因为对比空间在本质上具有代数性质，具体体现在两个方面。首先，分量和为零这一条件是代数条件，尽管它不是单项式条件。其次，每个纤维都由特定坐标的消失来定义，这是单项式簇最简单的例子。这些性质将在第 2.1 节中得到精确说明。

作为本节的结论，请注意：当单参数统计模型是一条可微曲线时，定理 2.1 使我们能够将切向量与速度这一几何概念转化为对比空间这一代数概念。这一点将在本文后续部分加以利用。

2.1 坐标子空间

方程 (1) 中定义的切丛并非微分几何中传统的切丛，因为它是一组维数各不相同的空间的集合，单纯形的每个面对应其中一个空间。我们的定义与代数几何的描述相一致：在对比空间（ $\mathbb{R}^\Omega$ 的一个子空间）的方程之外，我们还加入了描述缺陷支撑（坐标的单项式簇）的方程。本文特别关注定义在单纯形边界上的模型，并说明如何参照列联表来实现这一点。

我们进一步分析方程 (1) 的结构。事实上，完整单纯形的切丛定义要求为每个面分别引入向量空间。定理 2.1 的证明表明，沿单纯形的开面，切丛可根据 $\Delta(\Omega)$ 中元素的零点集自然分解。

令 $C(\Omega) = \left\{ v \in \mathbb{R}^\Omega \,\middle|\, \sum_{x\in\Omega} v(x) = 0 \right\}$ 为对比空间。对于顶点 $I \subseteq \Omega$ 的每个非空子集，定义其支撑位于 I 的对比空间

$$C_I(\Omega) = \{ v \in C(\Omega) \,|\, v(x) = 0 \text{ if } x \notin I \} = \operatorname{Span}\left(e_x - e_y, x, y \in I \right) ,$$

其中 e_x 是一个 d 维向量，其在位置 $x \in I$ 的坐标为 1，其余坐标为 0。任意 C_I 定义了一个对比空间，与定理 2.1 中针对 p 内 I 的特定结构零集合对切丛的刻画相符。每个 $C_I(\Omega)$ 都是由以下方程定义的线性子簇

$$\sum_{x\in\Omega} v(x) = 0 \quad \text{and} \quad v(x) = 0 \text{ if } x \notin I \, ;$$

也就是说，它是 $C(\Omega)$ 与由零条件给出的坐标子簇 的交集。我们得到一个线性簇与一个单项式簇的交集。例如，在应用统计学中，尤其是列联表理论中，避免使用隐式方程会很有帮助。通过引入完整对比空间的一组向量基，再进行代数消元，可以寻求一种完全避开对比条件的表示（见例2.1）。

例2.1。 考虑四点情形，$\Omega = \{1,2\}^2$。那么

$$C(\Omega) = \left\{ u \in \mathbb{R}^\Omega \,\middle|\, \sum_{x\in\Omega} u(x) = 0 \right\} = \mathbf{1}^\perp$$

其中$\mathbf{1}$ 是一个每个元素均为 1 的 2×2 矩阵， $\mathbf{1}^\perp$ 是相对于矩阵的弗罗贝尼乌斯内积而言与它正交的空间。 $C(\Omega)$的一组向量基为

$$c_{12} = e_{11} - e_{12} = \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}$$
$$c_{21} = e_{11} - e_{21} = \begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix}$$
$$c_{22} = e_{11} - e_{22} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

因此，一般对比可以写成

$$c = x \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} + y \begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix} + z \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} x+y+z & -x \\ -y & -z \end{bmatrix}$$

四个面的切平面分别由以下方程定义

$$x + y + z = 0 \quad x = 0 \quad y = 0 \quad z = 0$$

分别对应于多项式环 $\mathbb{R}[x,y,z]$ 中由 $xyz(x+y+z)$ 生成的多项式理想。也就是说，二元列联表饱和模型中的对比空间是一个实数值三维向量空间，其基为 c_{12}， c_{21}， c_{22}。当系数满足条件 $xyz(x+y+z) = 0$ 时，该对比属于一个与三维单纯形的某个面相切的空间。

可以用列联表中的结构零来直接解释如下。举例来说，取指标集 $I = \{(1,2)\} \subset \Omega$，由 $x = 0$ 给出，并对应于对比集 $C_{\{(1,2)\}} =$

$\begin{bmatrix} y+z & 0 \\ -y & -z \end{bmatrix}$。这些速度与所有在位置 (1,2) 具有结构零的列联表相容，而统计模型

$$\gamma(t) = t \begin{bmatrix} y+z & 0 \\ -y & -z \end{bmatrix} + \begin{bmatrix} p_{11} & 0 \\ p_{21} & p_{22} \end{bmatrix} \in \Delta(\Omega)\,,$$

当 p_{ij} 非负且 $p_{11}+p_{21}+p_{22}=1$ 时，对于零点邻域内的 t，该定义良好 o.

Ex例2.2（例2.1续）。对比空间的另一组可能的基为

$$c'_{12} = e_{11}+e_{21}-e_{12}-e_{22} = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}$$
$$c'_{21} = e_{11}+e_{12}-e_{21}-e_{22} = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$$
$$c'_{22} = e_{11}+e_{22}-e_{12}-e_{21} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

其中最后一个向量是独立性模型的基本马尔可夫移动，参见 Diaconis 和 Sturmfels[1998]。使用基 $\{c'_{12}, c'_{21}, c'_{22}\}$，一般对比可表示为

$$c = x\begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} + y\begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} + z\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} x+y+z & -x+y-z \\ x-y-z & -x-y+z \end{bmatrix}$$

以及与由以下多项式生成的理想相对应的切空间

$$(x+y+z)(-x+y-z)(x-y-z)(-x-y+z)\,.$$

例2.3。在与例2.1相同的设定中，考虑位于与 (1,1) 相对的面的一个元素 p，也就是说，

$$\{p \in \Delta(\Omega) \,|\, p(1,1) = 0\}$$

根据定理 2.1，切 可微曲线 γ： $t \mapsto \gamma(t)$（其中 $\gamma(0) = p$）的空间与
对比集合一一对应

$$\{v \in C(\Omega) \,|\, v(1,1) = 0\}$$

类似地，对于另一组顶点，这为写出列联 g=表中在 (1，1) 处有一个结构零的单参数统计模型的切空间提供了一种简便方法。

3 费舍尔得分

得分是费希尔统计学Fisher [1922] 的关键组成部分。本节首先论证，得分可解释为广义仿射空间中的速度（见 Pistone [2020]），并将以下定义应用于一个
。开单纯形中的一维可微统计模型，记作 Δ(Ω)。接下来，我们将在定理 3.1 中将其推广到闭单纯形。仿射空间在数学建模中得到系统应用，这一做法始于 Weil[1946]（最初发表于 1918 年）。

定义 3.1。

。1. p 到 q 在 Δ(Ω)中的位移定义为

$$s_p(q) = \log\frac{q}{p} - \mathbb{E}_p\left[\log\frac{q}{p}\right]$$

。其中， Δ(Ω) 是单纯形的内部集合，且这些比值按分量定义。

。 。2. 随机变量 $u \in \mathbb{R}^\Omega$ 从 $p \in \Delta(\Omega)$ 到 $q \in \Delta(\Omega)$ 的指数传输（其中 $\mathbb{E}_p[u] = 0$）定义为

$$^{e}\mathbb{U}_p^q u = u - \mathbb{E}_q[u]\,.$$

第 6 节将更详细地讨论传输的概念。随机变量u 在 q 处的指数传输相对于 q以零为中心。特别地，从 p 到 q 的传输将 p-对比映射为 q-对比，其中 p-对比是满足 $\mathbb{E}_p[u]=0$ 的任意随机变量 u （见定义 4.1）。

接下来，对于 p，q，$r\in$ $\overset{\circ}{\Delta}(\Omega)$ 广义平行四边形法则成立

$$s_p(q)+{}^{\mathrm{e}}\mathbb{U}_q^p s_q(r)=s_p(r)\,.$$

设 $t\mapsto\gamma(t)$ 是开单纯形中的一条可微曲线。以 p 为原点的曲线 γ 的表达式（即 $\gamma(0)=p$）为

$$t\mapsto\gamma_p(t)=s_p(\gamma(t))=\log\frac{\gamma(t)}{p}-\mathbb{E}_p\left[\log\frac{\gamma(t)}{p}\right]$$

其一阶导数的表达式为

$$\frac{d}{dt}\gamma_p(t)=\frac{d}{dt}\log\gamma(t)-\mathbb{E}_p\left[\frac{d}{dt}\log\gamma(t)\right]\,.$$

注意，这与第 2 节有类似之处，这也体现在记号上。在通常的仿射框架中，从p到 q 的位移为 $q-p$，因此导数 $\frac{d}{dt}(\gamma(t)-p)=\dot\gamma(t)$ 不依赖于原点 p，正如第 2 节中所提到的。我们采用不依赖于原点的移动标架导数，

$$\frac{d}{dt}\gamma_p(t)\Big|_{p=\gamma(t)}=\frac{d}{dt}\log\gamma(t)\,,$$

因为 $\mathbb{E}_p\left[\dot\gamma_{(t}\right]\Big|_{p=\gamma(t)}=\mathbb{E}_{\gamma(t)}\left[\frac{d}{dt})\log\gamma(t)\right]=0$。

定义 3.2。对于由开单纯形中的可微曲线 $t\mapsto\gamma(t)$ 给出的单参数统计模型，得分函数是 $\log\gamma(t)$ 的导数，即 $\frac{d}{dt}\log\ \gamma(t)$。

因此，从上述精确意义上说，得分是在仿射空间中的速度，其中速度是相对变化率。定理 3.1 展示了如何将这种方法推广到闭单纯形。

定理 3.1。如果 $t\mapsto\gamma(t)\in\mathbb{R}^\Omega$ 是 $\Delta(\Omega)$ 中的一条可微曲线，那么它的（通常意义下的）速度是一条曲线 $t\mapsto\dot\gamma(t)$，位于对比空间 $C(\Omega)$ 中。此外，对于每个 t，都存在 $s(t)\in\mathbb{R}^\Omega$，使得

$$\dot\gamma(t)=s(t)\,\gamma(t)\quad\textit{and}\quad\operatorname{Supp}\dot\gamma(t)=\operatorname{Supp}s(t)\cap\operatorname{Supp}\gamma(t)\,,\tag{2}$$

其中，方程(2) 中的第一个等式是逐分量成立的，即对所有 $x\in\Omega$，$\dot\gamma(x;t)=s(x;t)\ \gamma(x;t)$。

证明。若对某个 $x\in\Omega$ 和某个 $\bar t$，有 $\gamma(x;\bar t)=0$，即曲线在时刻 $\bar t$触及单纯形的一个面，则实值函数 $t\mapsto\gamma(x;t)$ 在 $t=\bar t$处取得最小值，因此 $\dot\gamma(x;\bar t)=0$。这证明了对每个 $x\in\Omega$，$\dot\gamma$ 相对于 γ绝对连续，简记为 $\dot\gamma\ll\gamma$。结论可由拉东-尼科迪姆定理直接得到。 □

上述统计论证是一般凸性论证的一个特例。具体而言，若 $t\mapsto x(t)$ 是闭凸集 $K\subset\mathbb{R}^\Omega$ 中的一条可微曲线，且 $x(\bar t)\in\partial K$，则 $\dot x(\bar t)$ 属于 $x(\bar t)$ 处的某个支撑超平面。因此，得分的概念纳入了次微分理论；该理论由 Rockafellar [1970]之后系统发展。不过，我们旨在找到一种能利用模型光滑性的代数框架。方程 (2) 中的常微分方程类似于对 x 中的指数函数$f(x)=\exp(s\,x)$ 求导后得到的常微分方程。对于 a，b和 γ，只要相关计算有意义，就有 $\frac{d}{dt}\gamma(t)=s(t)\ \gamma(t)$，其解为$\gamma(b)=\gamma(a)\exp\left(\int_a^b s(t)\ dt\right)$。在这种情况下，$\gamma$在 a中与 b具有相同的支撑；也就是说，常微分方程方法假设支撑恒定，而本文关注的是支撑变化。

如果曲线γ保持在 概率单纯形的内部，则原始的Fisher定义的得分 $s(t)$成立 s,

$$s(t)=\frac{\dot\gamma(t)}{\gamma(t)}=\frac{d}{dt}\log\gamma(t)\,.\tag{3}$$

否则， $s(t)$ 在支撑集$\gamma(t)$上唯一确定，而在 $\gamma(t)$的零集上则是任意的。

让我们从 Watanabe[2009] 中统计模型的奇点理论的角度来讨论得分的概念。如果模型 γ 可微，则在每一点 $x \in \Omega$处成立

$$\gamma(x;t) - \gamma(x;s) = (t-s)R_1(x;s,t) \quad \text{where } R_1(x;s,t) = \int_0^1 \dot{\gamma}(x, s+u(t-s))\, du$$

以及 $\lim_{t\to s} R_1(x;\ s,\ t) = \dot{\gamma}(x;s)$。假设在 $s=0, \gamma(x;0) = \dot{\gamma}(x;0) = 0$ 处存在一个临界点，因此

$$\gamma(x;0) = 0 \quad \Rightarrow \quad \gamma(x;t) = tR_1(x;0,t) \quad \text{with } \lim_{t\to 0} R_1(x;0,t) = 0\,, \tag{4}$$

and

$$\dot{\gamma}(x;t) = \dot{\gamma}^{\star}(x;t) t R_1(x;0,t)$$

其中 $\dot{\gamma}^{\star}(t) = s(t)$（参见定义 3.3）。

如果模型 γ 二阶可微，则

$$\dot{\gamma}^{\star}(x;t)\gamma(x;t) = \dot{\gamma}(x;t) = R_1(x;0,t) + t\frac{\partial}{\partial t}R_1(x;0,t)\,, \quad \lim_{t\to s}\frac{\partial}{\partial t}R_1(x;s,t) = \frac{1}{2}\ddot{\gamma}(x;s)\,.$$

或者，我们可以使用展开式

$$\gamma(x;t) = \gamma(x;s) + (t-s)\dot{\gamma}(x;s) + (t-s)^2 R_2(x;s,t) \quad \text{with } \lim_{t\to s} R_2(x;s,t) = \frac{1}{2}\ddot{\gamma}(x;s)\,,$$

从而得到，在 $s=0$处的奇点处，

$$\gamma(x;t) = t^2 R_2(x;0,t)\,,$$

因此

$$\dot{\gamma}(x;t) = \dot{\gamma}^{\star}(x;t) t^2 R_2(x;t)\,.$$

例 3.1（熵产生）。熵产生率是 $l(p) = -p\log p$ 关于时间的一阶导数。在沿统计模型计算熵产生时，需要检查$l(p)$的临界点。设概率模型 $\gamma(t)$ 在 $(x,0)$ 处有一个孤立零点。在非临界点处，我们有 $-\frac{d}{dt}l(\gamma(x;\ t)) = (\log(\gamma(x;\ t)) + 1)\,\dot{\gamma}(t)$。否则，假设 $\gamma \in C^2$，

$$t^{-1}l(\gamma(x;t)) = t^{-1}(t^2 R_2(t)\log(t^2 R_2(t))) = t^{-1}(l(t^2)R_2(t) + t^2 l(R_2)) \to 0 \ \text{ for } t \mapsto 0\,.$$

因此，熵函数在 $t=0$ 处可微。

示例 3.2（平方根变换）。变换 $p \mapsto 2\sqrt{p} = \rho$ 将概率单纯形映射到半径为 2 的球面，而球面是一个黎曼流形。这是对概率单纯形诱导黎曼结构的一种方法。例如，Amari [2009] 讨论了此类变换在统计学中的应用。球面上的导数与单纯形上的得分之间的关系为

$$\frac{d}{dt}\rho(t) = \frac{d}{dt}2\sqrt{\gamma(t)} = \frac{\dot{\gamma}^{\star}(t)\gamma(t)}{\sqrt{\gamma(t)}} = \frac{1}{2}\dot{\gamma}^{\star}(t)\rho(t)\,.$$

3.1 速度方程

接下来，我们借助定理3.1，以代数形式表述得分的性质。给定一条可微曲线$\gamma \in \Delta(\Omega)$，对于每个 t，其一阶导数都是对比空间中的一个元素与曲线本身的乘积。

在环 $\mathbb{R}^{\Omega}$中，所有形如 $s\gamma$ 的向量构成的集合是由 γ 生成的理想。因此，导数满足

$$\dot{\gamma} \in C(\Omega) \cap \mathrm{Ideal}(\gamma)$$

其中，$\mathrm{Ideal}(\gamma(t))$ 宽泛地定义为从 Ω 到 $\mathbb{R}$ 的所有函数的集合；对于任意给定的 t，这些函数都是 $\gamma(t)$ 的倍数。这一性质意味着，导数的支撑集等于或包含于模型曲线 γ 的支撑集。反过来，所有表示相同速度的得分 s 的集合由以下方程定义

$$\sum_{x\in\Omega}\gamma(x) = 1,\ \sum_{x\in\Omega}\dot{\gamma}(x) = 0,\ \dot{\gamma}(x) = s(x)\gamma(x) \text{ for all } x \in \Omega\,.$$

为了解决奇异性 性方程（3），给定一个 $\gamma \in \Delta(\Omega)$，令 $S(t)$ 表示 γ 在时刻的支撑集的示性函数 e t。对于每个 t，$S(t)\colon \Omega \mapsto \mathbb{R}$满足

$$S^2 = S\,, \quad S\gamma = \gamma\,, \quad S \in \mathrm{Ideal}(\gamma)\,, \tag{5}$$

对于每个 $S^2 - S$，所有满足方程（5）的 $\gamma \in \Delta(\Omega)$ 构成的集合，与 S 上的开单纯形重合。方程（5）表明，对于 $\Delta(\Omega)$ 中的一条光滑曲线，有 $\dot{\gamma}(t) \in \text{Ideal}(\gamma(t))$. 事实上，这可通过考虑 Ω 上所有实值函数构成的集合与多项式环之间的实向量空间同构来得出；该多项式环由与 Ω 的基数相同数量的不定元构成，并模去由和为一条件生成的理想。Pistone 等人 [2001]。

特别地，我们可以将 $\dot{\gamma} = \dot{\gamma}/\gamma$ 速度的定义推广到单纯形边界，如下所示。

定义 3.3（速度）。若 $S(t) =$ 支撑集 $\gamma(t)$ 且 $S(t) = \check{\gamma}(t)\gamma(t)$，则速度 $\dot{\gamma}(t)$ 由任意曲线定义。$t \mapsto (\gamma(t)\ \dot{\gamma}(t))$ 使得

$$(\dot{\check{\gamma}}(t) - \dot{\gamma}(t)\check{\gamma}(t))\ \gamma(t) = 0\,. \tag{6}$$

与例如 Chirco 和 Pistone [2022], 所讨论的标准情形相比，我们允许支撑集可变，因此切空间的维数也可变。对于开单纯形，得分为 $\dot{\gamma}(t) = \frac{d}{dt}\log\gamma(t)$。$*$记法是在非参数信息几何的相关文献中引入的。例如参见教程 Pistone [2020]。

此外，关于 $\gamma(t)$ 的 $s(t) = \dot{\gamma}(t)$ 的期望值对于所有 t 都为零，事实上

$$\mathbb{E}_{\gamma(t)}\left[s(t)\right] = \sum_{x\in\Omega} s(x;t)\gamma(x;t) = \sum_{x\in\Omega} \dot{\gamma}(x;t) = 0\,.$$

3.2 例子

在以下例子中，我们考虑由多项式方程描述的模型。根据求导法则（方程 (2)），速度由曲线与得分的乘积给出，直接可得该模型在切丛中的相应方程。

例 3.3（二维单纯形中的一条直线）。令 $\Omega = \{1,2,3\}$。对每个 $t \in [0,1/2]$，下面的曲线在 $\Delta(\Omega)$ 中给出一个概率密度函数

$$p_1 = t \qquad p_2 = t \qquad p_3 = 1 - 2t$$

它与一个一维 面相交于 $\Delta(\Omega)$ 的 $t = 1/2$，并与 $\Delta(\Omega)$ 的一个顶点相交于 $t = 0$。它在 t i 中的一阶导数 s（1，1，-2）。其隐式 这个一维统计模型的 t 多项式形式为

$$p_1 - p_2 = 0 = p_3 + 2p_1 - 1 \tag{7}$$

在模型方程（7）中加入切向条件 $\dot{p}_i = s_i\, p_i$，其中 $i = 1$，2，3，并加入和为一约束以及速度约束

$$p_1 - p_2 = 0 = p_3 + 2p_1 - 1 \qquad \dot{p}_i = s_i\, p_i \text{ for } i = 1,2,3 \qquad p_1 + p_2 + p_3 - 1 = 0 = \dot{p}_1 + \dot{p}_2 + \dot{p}_3$$

我们考虑由以下元素生成的 $\mathbb{R}[p, \dot{p}, s]$ 理想

$$p_1 - p_2, \quad p_3 + 2p_1 - 1, \qquad \dot{p}_i - s_i\, p_i \text{ for } i = 1,2,3 \qquad p_1 + p_2 + p_3 - 1, \quad \dot{p}_1 + \dot{p}_2 + \dot{p}_3$$

在导数规则为 $\dot{p}_i = s_i p_i$ 的微分环中，我们得到 $\dot{p}_1 - \dot{p}_2$，$\dot{p}_3 + 2\dot{p}_1$，这正如预期，因为在模型上，等式 $\dot{p}_1 - \dot{p}_2 = 0 = \dot{p}_3 + 2\dot{p}_1$ 在 $t = 0$ 或 $1/2$ 时也成立。用概率变量和得分变量表示，该模型由两个多项式方程 $s_1p_1 - s_2p_2 = 0 = s_3p_3 + 2s_1p_1 = 0$ 给出，它们提供了切丛的隐式表达式。

例 3.4 (2×2 列联表)。上述计算可应用于双参数统计模型，例如 2×2 独立性模型；这里用它对例 3.5 进行初步计算。令 p_{ij},i，$j = 1$，2表示单元格概率，并考虑环 $\mathbb{R}[p_{ij}, \dot{p}_{ij}, s_{i,j}$：$i$，$j = 1$，$2]$。我们强调，$p$是概率，而 $\dot{p}$是对比，因此

$$p_{11} + p_{12} + p_{21} + p_{22} = 1 \tag{8}$$

$$\dot{p}_{11} + \dot{p}_{12} + \dot{p}_{21} + \dot{p}_{22} = 0 \tag{9}$$

在边际独立模型下，我们有两类方程：

1. 一个用于独立性条件的二项式方程

$$p_{11}p_{22} - p_{12}p_{21} = 0 \tag{10}$$

2. 以及定义这些得分的条件方程

$$\dot{p}_{ij} = s_{ij}p_{ij} \qquad i,j = 1,2 \tag{11}$$

对于 p 个参数的一元变量可微函数，对方程（10）关于该参数求导，并利用方程（11），可得

$$\dot{p}_{11}p_{22} + p_{11}\dot{p}_{22} - \dot{p}_{12}p_{21} - p_{12}\dot{p}_{21} = 0$$
$$s_{11}p_{11}p_{22} + p_{11}s_{22}p_{22} - s_{12}p_{12}p_{21} - p_{12}s_{21}p_{21} = (s_{11}+s_{22})p_{11}p_{22} - (s_{12}+s_{21})p_{12}p_{21} = 0 \tag{12}$$

我们将方程 (9) 和 (11) 相加，得到 Fisher得分条件，并可得

$$s_{11}p_{11} + s_{12}p_{12} + s_{21}p_{21} + s_{22}p_{22} = 0 \tag{13}$$

方程 (12) 和 (13) 用概率变量和得分变量表示该模型。在开单纯形中，由于 $p_{11}p_{22} = p_{12}p_{21} \neq 0$，我们得到一个仅涉及得分变量的简单方程

$$s_{11} + s_{22} - s_{12} - s_{21} = 0\,.$$

这与经典结果一致：在独立性条件下，得分呈线性关系。

例 3.5. （2×2 具有独立性和边际齐性的表）。我们通过添加条件 $p_{12} - p_{21} = 0$ 来修改例 3.4 中的独立性模型，其显式形式为

$$p_{11} = (1-\theta)^2$$
$$p_{12} = p_{21} = \theta(1-\theta)$$
$$p_{22} = \theta^2$$

因此，我们使用以下方程处理三个概率

1. 模型

$$p_{11}p_{22} - p_{12}^2 = 0 \tag{14}$$

2. 得分条件

$$\dot{p}_{ij} = s_{ij}p_{ij} \qquad\qquad (i,j) \in \{(1,1),(1,2),(2,2)\} \tag{15}$$

对于 p 变量的单参数可微函数，推导方程 (14) 可得

$$\dot{p}_{11}p_{22} + p_{11}\dot{p}_{22} - 2p_{12}\dot{p}_{12} = 0$$

并利用方程 (15)，我们得到关于 p 和 s 变量的模型方程

$$s_{11}p_{11}p_{22} + p_{11}s_{22}p_{22} - 2s_{12}p_{12}^2 = 0$$

$$(s_{11}+s_{22})p_{11}p_{22} - 2s_{12}p_{12}^2 = 0$$

在开单纯形中，由于 $p_{11}p_{22} = p_{12}^2$，我们得到关于得分的一个线性条件

$$s_{11} + s_{22} - 2s_{12} = 0$$

对于可表示为二项式方程的模型，以下结果将这一事实推广了。

引理 3.2. 设 p_i 为不定元，且对所有 i，$\dot{p}_i = s_i\, p_i$。给定单项式 p^α，则其导数为 $\langle \alpha, s\rangle p$.

证明。这可由导数的莱布尼茨乘积法则推出。 □

定理 3.3。给定二项式方程 $p^\alpha - p^\beta = 0$，则在开单纯形上：

$$\langle (\alpha - \beta), s\rangle = 0 \tag{16}$$

证明。 $p^\alpha - p^\beta = 0$ 的导数必须为零，而前一个引理意味着关于 s 的线性方程在开单纯形上必须成立。 □

作为推论，对于二项式模型这一特殊情形，定理 3.3 给出了一组替代方程，这些方程只涉及不定元 p 和 s（而 $\dot{p}$ 已被消去）。

4 统计丛

在本节中，我们根据第 3 节所讨论的将 Fisher 得分视为速度的概念，重新定义速度丛。在定理 4.1 中，我们证明了这一新概念的一致性。指数族提供了从速度丛的纤维到概率单纯形的自然映射。我们采用如下形式的一维指数族

$$\gamma(t) = \mathrm{e}^{ts-\psi(t)} \cdot p\,, \quad \mathbb{E}_p[s] = 0\,, \quad \psi(t) = \log \mathbb{E}_p\left[\mathrm{e}^{ts}\right]\,.$$

参见 Brown [1986] 关于指数族。Pistone 和 Sempi [1995] 使用 K_p 记号来表示累积量函数。即,

$$q = \mathrm{e}^{u-K_p(u)} \cdot p\,, \quad \mathbb{E}_p[u] = 0\,, \quad K_p(u) = \log \mathbb{E}_p\left[\mathrm{e}^{u}\right]\,.$$

定义 4.1. 设 $L(\Omega) = \mathbb{R}^\Omega$ 为 Ω 上实值函数的向量空间。

1. 统计丛是对的集合

$$S\Delta(\Omega) = \{(p,s)\,|\,p \in \Delta(\Omega), s \in L(\Omega) \text{ such that } \mathbb{E}_p[s] = 0\}\,. \tag{17}$$

p 处的纤维为 $S_p\Delta(\Omega) = L_0(p) = \{s \in L(\Omega)\,|\,\mathbb{E}_p[s] = 0\}$。每个纤维均被赋予双线性型$\langle s, v\rangle_p = \mathbb{E}_p[sv] = \mathrm{Cov}_p(s,v)$，对于 $s, v \in L_0(p)$。

2. 得分丛是 $p \in \Delta(\Omega)$ 上所有 p-得分纤维的并，定义如下:

$$\{s \in L(\Omega)\,|\,\text{there exists } \dot{p} \in C(\Omega) \text{ such that } \dot{p} = sp\}\,.$$

注意，定理 2.1 和定义 4.1 类似，一个以代数几何语言表述，另一个则以概率统计语言表述。下面的定理 4.1 表明，统计丛确实是一个切丛。

定理 4.1. 统计丛等于得分丛。

证明. 若 $\dot{p} = sp$，则 $\mathbb{E}_p[s] = \sum_x \dot{p}(x) = 0$。反过来，设 $p \in \Delta(\Omega)$，并设 $s \in L(\Omega)$ 满足 $\mathbb{E}_p[s] = 0$。定义指数族

$$\gamma(t) = \mathrm{e}^{ts-\psi(t)} \cdot p\,.$$

于是,

$$\left.\frac{d}{dt}\gamma(t)\right|_{t=0} = \left.\left(s - \frac{d}{dt}\psi(t)\right)\gamma(t)\right|_{t=0} = \left.\left(s - \mathbb{E}_{\gamma(t)}[s]\right)\gamma(t)\right|_{t=0} = sp\,.$$

最后一个等号成立，是因为 $\gamma(0) = p$ 且 $\mathbb{E}_p[s] = 0$。 □

定理 4.1 表明，对于任意 $p \in \Delta(\Omega)$，向量空间 $L_0(p) = \{u \in L(\Omega)\,|\,\mathbb{E}_p[u] = 0\}$ 等于 p 处所有可能费舍尔得分的纤维。空间 $L_0(p)$ 并不能唯一地 表示所有得分，因为如果 $u,\ v \in L_0(p)$ 和 $(u-v)p = 0$，也就是说，u 和 v 在支撑集 $\mathrm{Supp}\,p$ 上相等但在其他地方不同，那么它们表示相同的得分。为确保唯一性，考虑等价关系

$$(p,u),(p,v) \in S\Delta(\Omega)\,, \quad (p,u) \sim (p,v) \Leftrightarrow (u-v)p = 0\,, \tag{18}$$

即，在 $\mathrm{Supp}\,p$ 上的 $u = v$。

定义 4.2。 约化统计丛 是关于关系 $\sim$ 的统计丛商集,

$$S_{\mathrm{r}}\Delta(\Omega) = \left\{(p,u)\,\middle|\,p \in \Delta(\Omega),\ u \in \frac{L_0(p)}{\sim} = L_0(\mathrm{Supp}\,p, p)\right\}\,, \tag{19}$$

其中 $L_0(\mathrm{Supp}\ p, p) = \{u \in L(\mathrm{Supp}\ p)\,|\,\mathbb{E}_p[u] = 0\}$ 是一个希尔伯特空间，对于 $(u,v) \mapsto \langle u, v\rangle_p$。

统计丛继承自 $L_0(p)$ 的实向量空间结构。此外，它由以下半代数条件描述:

$$\sum_x p(x) = 1,\ \sum_x u(x)p(x) = 0,\ p(x) \geq 0\,.$$

以下命题将得分解释为统计丛中的速度。

命题 4.2. 设 $t \mapsto \gamma(t)$ 是概率单纯形中的一条 C^1 曲线。设 $F\colon\ \mathbb{R}^\Omega \to \mathbb{R}$, $F \in C^1$ 是模型上的一个隐式关系，即 $F(\gamma(t)) = 0$ 对所有 t均成立。那么对于 F 的梯度 ∇F，有

$$\mathrm{Cov}_{\gamma(t)}\left(\nabla F(\gamma(t), \dot{\gamma}(t)\right) = 0\,.$$

证明。 我们有 $\mathbb{E}_{\gamma(t)}\left[\dot{\gamma}(t)\right]=0$，因此

$$0=\frac{d}{dt}F(\gamma(t))=\sum_x\frac{\partial}{\partial\gamma(x)}F(\gamma(x;t))\dot{\gamma}(x;t)\gamma(x;t)=\mathbb{E}_{\gamma(t)}\left[\nabla F(\gamma(t))\dot{\gamma}(t)\right]=\mathrm{Cov}_{\gamma(t)}\left(\nabla F(\gamma(t)),\dot{\gamma}(t)\right).$$

□

根据力学的标准形式体系（Landau 和 Lifshits [1976] 以及 Chirco 等人 [2022]），我们在下文的定义 4.3 中为统计丛引入一种可微结构。

定义 4.3（统计拉格朗日量）。

1. 拉格朗日量是一个映射
$$\Lambda\colon S\Delta\left(\Omega\right)\ni(p,u)\mapsto\Lambda(p,u)\in\mathbb{R}.$$

2. 拉　　格朗日量若与方程（18）中的等价关系相容，则称其为统计的，即
$$(u-v)p=0\quad\text{implies}\quad\Lambda(p,u)=\Lambda(p,v).$$

若 $t\mapsto(\gamma(t)\ u(t))\in S\Delta(\Omega)$ 是一条可微曲线，且 Λ 是定义在包含统计丛的开集上的可微函数，则

$$\frac{d}{dt}\Lambda(\gamma(t),u(t))=\mathbb{E}_{\gamma(t)}\left[\nabla_1\Lambda(\gamma(t),u(t))\dot{\gamma}(t)\right]+\nabla_2\Lambda(\gamma(t),u(t))\cdot\dot{u}(t))$$

其中，∇_1 表示对 Λ 的第一个自变量求偏导，∇_2 表示对第二个自变量求偏导。

例 4.1. 考虑 $\Lambda(p,u)=K_p(u)=\log\mathbb{E}_p\left[\mathrm{e}^u\right]$，其中 $(p,u)\in S\Delta(\Omega)$。若 $(u-v)\ p=0$，则 $\Lambda(p,u)=\Lambda(p,v)$，因此累积量泛函是一个统计拉格朗日量。若 $t\mapsto(\gamma(t)\ u(t))$ 可微，则

$$\begin{aligned}\frac{d}{dt}K_{\gamma(t)}(u(t))&=\mathbb{E}_{\gamma(t)}\left[\mathrm{e}^{u(t)}\right]^{-1}\sum_x\frac{d}{dt}\gamma(x;t)\mathrm{e}^{u(t)}\\&=\mathrm{e}^{-K_{\gamma(t)}(u(t))}\sum_x(\dot{\gamma}(x;t)\mathrm{e}^{u(x;t)}+\gamma(x;t)\mathrm{e}^{u(x;t)}\dot{u}(x;t))\\&=\mathbb{E}_{\gamma(t)}\left[\dot{\gamma}(t)\mathrm{e}^{u(t)-K_{\gamma(t)}(u(t))}+\mathrm{e}^{u(t)-K_{\gamma(t)}(u(t))}\dot{u}(t)\right].\end{aligned}$$

在上面的方程中，因子 $\gamma^u(t)=\mathrm{e}^{u(t)-K_{\gamma(t)}(u(t))}\cdot\gamma(t)$ 是一个新的模型。它满足

$$\frac{d}{dt}K_{\gamma(t)}(u(t))=\mathbb{E}_{\gamma^u(t)}\left[\dot{\gamma}(t)+\dot{u}(t)\right].$$

累积量函数的导数的代数性质已在 Pistone and Wynn [1999] 以及 Pistone and Wynn [2006] 中讨论过。在 $u=\dot{\gamma}$ 的情况下，我们有加速度这一概念。此处不作讨论，详见 Chirco 和 Pistone [2022] 的教程。

给定 p，考虑由所有满足 $k\,p=0$ 的 $k\in\mathbb{R}^\Omega$ 构成的向量空间 J_p。那么，u 和 v 表示相同的得分，当且仅当 $u-v\in J_p$。从代数角度看，公式（19）中的约化统计丛是一个半代数簇。事实上，令 $\mathbb{R}[p,u]$ 为以 p（x） u（x） $x\in\Omega$ 为不定元的多项式环。统计丛由以下多项式的零点集给出

$$\sum_x p(x)-1$$
$$\sum_x u(x)p(x)$$

并由不等式 $p(x)\geq0$。注意，J_p 是一个理想，因此

$$S\Delta(\Omega)=V(\sum_x p(x)-1,\sum_x u(x)p(x),p(x)\geq0)/J_p$$

5 自然梯度

将费舍尔得分扩展到闭单纯形，使数理统计中广为人知的微分计算得以推广。一个显著的例子是克拉美-罗不等式，Radhakrishna Rao [1945], 提出了这一不等式，它为给定单参数模型 $\gamma(t)$ 中任意估计量的方差提供了一个统计界。定理 3.1 中的位置$\dot{\gamma} = \overset{\star}{\gamma}\gamma$ 是推导该不等式的关键环节；更一般地说，也是计算 Amari [1998]以及 Amari 和 Nagaoka [2000]的自然梯度的关键环节。

命题 5.1. 令 $t \mapsto \gamma(t)$ 为概率单纯形中的一条可微曲线， g 为定义在Ω上的随机变量。则

$$\left(\frac{d}{dt}\,\mathbb{E}_{\gamma(t)}\left[g\right]\right)^2 \leq \mathrm{Var}_{\gamma(t)}\left(\overset{\star}{\gamma}(t)\right)\ \mathrm{Var}_{\gamma(t)}\left(g\right)\ .$$

证明。 令 $G(t) = \mathbb{E}_{\gamma(t)}\left[g\right]$ 为 g 的期望值。则

$$\dot{G}(t) = \sum_x \dot{\gamma}(x;t)g(x) = \sum_x \overset{\star}{\gamma}(x;t)g(x)\gamma(x;t) = \sum_x \overset{\star}{\gamma}(x;t)(g(x) - \mathbb{E}_{\gamma(t)}\left[g\right])\gamma(x;t) = \mathrm{Cov}_{\gamma(t)}\left(\overset{\star}{\gamma}(t), g\right)\ .$$

这是由 $\dot{\gamma}(x;t)$ 属于切丛这一事实得出的:

$$0 = \sum_x \overset{\star}{\gamma}(x;t)\gamma(x;t) = \sum_x \dot{\gamma}(x;t)$$

结论由此可通过应用施瓦茨不等式得出。 □

更一般地，设 $G\colon \Delta(\Omega) \to \mathbb{R}$ 是一个统计估计量，使得 $G(\gamma)$ 可微。由命题 5.1，我们可以写成:

$$\begin{aligned}\frac{d}{dt}G(\gamma(t)) &= \sum_x \frac{\partial\, G(\gamma(t))}{\partial\gamma(x)}\,\dot{\gamma}(x,t) = \sum_x \frac{\partial\, G(\gamma(t))}{\partial\gamma(x)}\,\overset{\star}{\gamma}(x,t)\,\gamma(x,t)\\ &= \mathbb{E}_{\gamma(t)}\left[\left(\nabla G(\gamma(t)) - \mathbb{E}_{\gamma(t)}\left[\nabla G(\gamma(t))\right]\right)\overset{\star}{\gamma}(t)\right]\\ &= \left\langle \nabla G(\gamma(t)) - \mathbb{E}_{\gamma(t)}\left[\nabla G(\gamma(t))\right],\ \overset{\star}{\gamma}(t)\right\rangle_{\gamma(t)}\ .\end{aligned}$$

其中，最后一个等式源于以下事实：对于任意满足 $\mathbb{E}_\gamma\left[u\right] = \mathbb{E}_\gamma\left[v\right] = 0$ 的 u， v，都有 $\left\langle u, v\right\rangle_\gamma = \mathbb{E}_\gamma\left[uv\right]$。这引出了下面的定义。

定义 5.1 （自然梯度）。 设 $G\colon \Delta(\Omega) \to \mathbb{R}$ 满足：对于所有可微曲线 γ， $t \mapsto G(\gamma(t))$ 是无穷可微的。 G 的自然梯度 grad G 是统计丛的一个截面，即 $(\gamma,\mathrm{grad}\ G(\gamma)) \in S\Delta(\Omega)$，使得

$$\frac{d}{dt}G(\gamma(t)) = \left\langle \mathrm{grad}\, G(\gamma(t))\ ,\ \overset{\star}{\gamma}(t)\right\rangle_{\gamma(t)} \tag{20}$$

自然梯度是相对于由 Fisher 得分描述的相对速度，以及由协方差给出的标量积而言的梯度。在下面的章节中，我们将通过例子说明，对于单项式模型，这些运算可通过代数延拓到闭单纯形。

5.1 例子

例 5.1 （由例 3.1 得出）。 假设 γ 是无穷可微的，并令 $l(\gamma) = -\gamma\log\gamma$。那么，如例 3.1 所示，熵 l 是可微的，且熵产生为

$$\frac{d}{dt}l(\gamma(x;t)) = -(\log(\gamma(x;t)) + 1)\ \overset{\star}{\gamma}(x;t)\ \gamma(x;t)\,.$$

这里我们计算熵沿曲线 γ 的自然梯度，其中 $\gamma(t) \in \Delta\Omega$，且熵定义为$H(\gamma(t)) = -\sum_{x\in\Omega}\gamma(x;t)\log\gamma(x;t)$，对所有 t 均成立。

对 x 求和，可得

$$\begin{aligned}\frac{d}{dt}H(\gamma(t)) &= -\sum_x \left(\log(\gamma(x;t)) + 1\right)\ \overset{\star}{\gamma}(x;t)\ \gamma(x;t) = \mathbb{E}_{\gamma(t)}\left[-(\log\gamma(t) + 1)\overset{\star}{\gamma}(t)\right]\\ &= \mathbb{E}_{\gamma(t)}\left[\left(-(\log\gamma(t) + 1) - \mathbb{E}_{\gamma(t)}\left[-(\log\gamma(t) + 1)\right]\right)\overset{\star}{\gamma}(t)\right]\\ &= \left\langle -\log\gamma(t) - H(\gamma(t))\ ,\ \overset{\star}{\gamma}(t)\right\rangle_{\gamma(t)}\end{aligned}$$

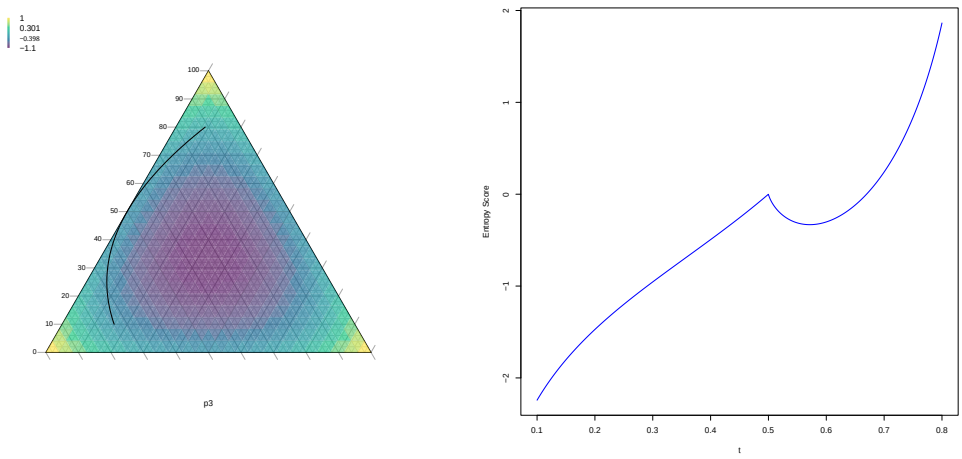


图1：例 5.1。具有缺陷支撑的熵。左图：熵的热图。右图：熵产生。

注意，$(-\log \gamma(t) - H(\gamma(t)) \quad \dot{\gamma}) \in S\Delta(\Omega)$，并且我们利用了以下事实：$\mathbb{E}_{\gamma(t)}\left[\mathbb{E}_{\gamma(t)}\left[-(\log \gamma(t) + 1)\right]\dot{\gamma}(t)\right] = \mathbb{E}_{\gamma(t)}\left[-(\log \gamma(t) + 1)\right]\mathbb{E}_{\gamma(t)}\left[\dot{\gamma}(t)\right] = 0$，因为 $\mathbb{E}_{\gamma(t)}\left[\dot{\gamma}(t)\right] = 0$。此外，请注意，等式链表明 $\frac{d}{dt}H(\gamma(t))$ 等于 $-\log \gamma$ 与 $\dot{\gamma}$ 之间的协方差。

熵的自然梯度为

$$(\gamma\,,\ \operatorname{grad} H(\gamma)) = (\gamma\,,\ -\log\gamma - H(\gamma)) \in S\Delta\,(\Omega)\ .$$

注意 $\operatorname{grad} H(\gamma) \quad = 0$ if $-\log \gamma = H(\gamma)$，也就是说，γ 在其支撑上为常数。这是最大熵的情形。

例 5.2（由例 5.1 得出）。图1提供了一个数值示例。在 $\Omega = \{1,2,3\}$ 上，考虑模型

$$]\,.1,.8[\ni t \mapsto \left(t, \left(t-\frac{1}{2}\right)^2, \frac{3}{4} - t^2\right)\ .$$

具有孤立临界点 $\gamma(1/2) = (1/2,\ 0,\ 1/2)$。得分（见方程（3））为

$$\dot{\gamma}(t) = \left(t^{-1}, 2\left(t-\frac{1}{2}\right)^{-1}, 2t\left(t^2-\frac{3}{4}\right)^{-1}\right)\ , \quad t \neq \frac{1}{2}\,.$$

熵产生为

$$-\frac{d}{dt}H(\gamma(t)) = \log t + 2\left(t-\frac{1}{2}\right)\log\left(t-\frac{1}{2}\right)^2 - 2t\log\left(\frac{3}{4}-t^2\right)\ .$$

注意到奇点在 $t = 1/2$处得到了补偿。

例 5.3（玻尔兹曼-吉布斯模型）。图 2 是概率单纯形中一条光滑曲线的示例，展示了触及边界的两种方式。在顶点 1 和 3 处的渐近行为是指数族的典型特征，而切向触及则是统计物理学中玻尔兹曼-吉布斯模型的特征；参见 [Landau and Lifshits，1980 年，第 3] 章。令 U 为一个非负随机变量，其最小值为 $U = 0$，并令 V 为任意随机变量。在 $\Omega = \{1,2,3\}$ 上考虑模型

$$\gamma(x;\beta) \propto \exp\left(-\beta^{-2}U(x) + \beta V(x)\right) = (U(x)=0)\exp\left(\beta V(x)\right) + (U(x>0)\exp\left(-\beta^{-2}U(x) + \beta V(x)\right)\ ,$$

其中，β 是一个非零实值 参数。归一化常数（在物理学中称为配分函数）为

$$Z(\beta) = \sum_{x\in\Omega}\exp\left(-\beta^{-2}U(x) + \beta V(x)\right) = \sum_{U(x)=0}\exp\left(\beta V(x)\right) + \sum_{U(x)>0}\exp\left(-\beta^{-2}U(x) + \beta V(x)\right)\ .$$

令 $\#A$ 表示集合 A 的元素个数，并令 $(U(x) = 0)$ 表示事件 $U(x) = 0$ 的示性函数。当极限

$$\lim_{\beta\to 0}\exp\left(-\beta^{-2}U + \beta V\right) = (U=0) \quad \text{and} \quad \lim_{\beta\to 0} Z(\beta) = \#\{U=0\}\ ,$$

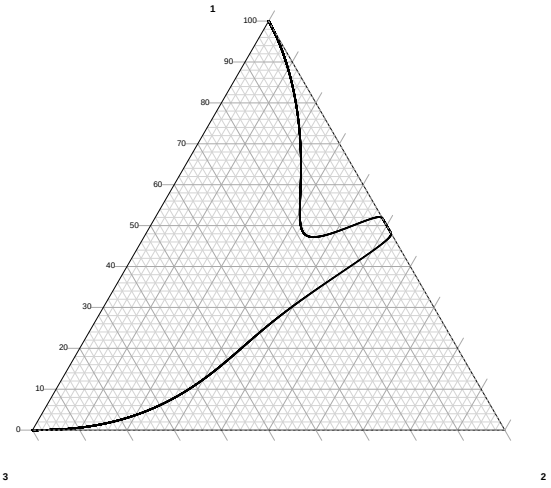


图 2：例 5.3，其中 $U = (0,0,1)$ 和 $V = (0,1,1.8)$。在 $\beta = 0$ 处，曲线以指数阶的接触方式到达边界点 $(1/2, 1/2, 0)$。对于 $\beta \to \pm\infty$，曲线分别趋向于与 V的最大值和最小值相关联的顶点。

存在时，我们可以通过以下方式将模型扩展到 $\beta = 0$

$$\gamma(x;\beta) = \begin{cases} \exp\left(-\beta^{-2}U(x) + \beta V - \log Z(\beta)\right) & \text{if } \beta \neq 0, \text{ and} \\ \lim_{\beta\to 0}\gamma(x;\beta) = \dfrac{(U(x)=0)}{\#\{U=0\}} & \text{if } \beta = 0, \end{cases}$$

使得曲线

$$\mathbb{R} \ni \beta \mapsto \gamma(\beta) \in \Delta(\Omega)$$

是连续的。累积量函数 $\psi(\beta) = \log Z(\beta)$ 在 β 处的导数为

$$\dot{\psi}(\beta) = \mathbb{E}_\beta\left[2\beta^{-3}U + V\right]$$

and

$$\lim_{\beta\to 0}\dot{\psi}(\beta) = \lim_{\beta\to 0} Z(\beta)^{-1}\sum_x \mathrm{e}^{-\beta^{-2}U(x)+\beta V(x)}(2\beta^{-3}U(x) + V(x)) = 0\,.$$

注意，我们利用了指数增长，即 $\lim_{\alpha\to\infty}\mathrm{e}^{-\alpha U(x)}\alpha^{3/2}U(x) = 0$，其中 $\alpha = 1/\beta^2$，来证明曲线可微，其导数为

$$\dot{\gamma}(x,\beta) = \begin{cases} \gamma(x;\beta)\left((2\beta^{-3}U(x) + V(x)) - \mathbb{E}_\beta\left[2\beta^{-3}U + V\right]\right) & \text{if } \beta \neq 0, \text{ and} \\ 0 & \text{if } \beta = 0, \end{cases}$$

观察可知，费舍尔得分为

$$\dot{\gamma}^\star(x;\beta) = \begin{cases} 2\beta^{-3}(U(x) - \mathbb{E}_\beta[U]) + (V(x) - \mathbb{E}_\beta[V]) & \text{if } \beta \neq 0, \\ 0 & \text{if } \beta = 0 \text{ and } U(x) = 0, \\ \text{anything} & \text{if } \beta = 0 \text{ and } U(x) > 0. \end{cases}$$

费舍尔得分的渐近方差，即费 Fisher 信息，为

$$\mathbb{E}_{\gamma(\beta)}\left[\dot{\gamma}^\star(\beta)^2\right] = \sum_{\dot{\gamma}^\star(x;\beta)>0}\frac{\dot{\gamma}(x;\beta)^2}{\gamma(x;\beta)}$$

对于 $\beta \to 0$，它会趋于零，因此该模型是奇异的；例如，参见 Watanabe [2009]。

图 2 中的三元图对应于 $U = (0,\ 0,\ 1)$ 和 $V = (0,\ 1,\ 1。8)$ 的情形。该模型关于参数是无穷可微的，且不能化为代数形式。如果取 $U = 0$，则 V-部分是指数形式，因此关于参数 $\theta_1 = \mathrm{e}^{\beta}$ 是单项式形式，如 [Pistone 等人，2001 年，第 5] 章所述。如果取 $V = 0$ 并采用参数化 $\theta_2 = \mathrm{e}^{-1/\beta^2}$，则该模型成为单项式统计模型。若将这两个单项式模型记为 $\gamma_i(\theta_i)$、$i = 1,\ 2$，则我们的例子可以看作一个约束单项式模型

$$\gamma(\theta_1, \theta_2) \propto \gamma_1(\theta_1)\gamma_2(\theta_2)\ , \quad \text{subject to } \log^2 \theta_1 \log \theta_2 + 1 = 0.$$

例 5.4。（密度混合）给定 $p,\ q \in \Delta(\Omega)$，考虑混合

$$\gamma(t) = (1-t)p + tq \qquad t \in [0,1]$$

在这种情况下，方程（3）变为

$$q - p = s(t)\left(p + t(q-p)\right)\ .$$

在 $t = 0$ 处，我们得到 $q - p = s\,p$；在 $t = 1$ 处，我们得到 $q - p = s\,q$。

对于 $t \in]0,\ 1[$，支撑集 $\gamma(t)$ = 支撑集 $p \cup$ 支撑集 q，它不依赖于 t，得分为

$$s(x;t) = \begin{cases} \frac{d}{dt}\log(p(x) + t(q(x)-p(x))) = \dfrac{q(x)-p(x)}{p(x)+t(q(x)-p(x))} & \text{if } x \in \operatorname{Supp} p \cup \operatorname{Supp} q \\ \text{arbitrary} & \text{otherwise} \end{cases}$$

对于 $t = 1$，支撑集 $\gamma(1)$ = 支撑集 q，得分为 $\frac{q}{(x)-p(x)\,q(x)}$ 若 $x \in$ 支撑集 q，否则任意取值。类似地，支撑集 $\gamma(0)$ = 支撑集 p，得分为 $\frac{q}{(x)-p(x)\,p(x)}$ 若 $x \in$ 支撑集 p，否则任意取值。

6 测地线

在开概率单纯形的通常几何中，方程 $\ddot{\gamma}(t) = 0$、$\dot{\gamma}(t) = \dot{\gamma}(0)$ 和 $\gamma(t) = \gamma(0) + t\dot{\gamma}(0)$ 等价，并定义了仿射参数化的特殊曲线，通常称为测地线。在一般仿射空间中，尤其是在第 4 节定义的统计丛中，也具有相同的性质。主要区别在于，不同点处的切空间彼此不同，因此我们需要较弱意义上的恒定速度概念，以及一种特殊的坐标选择。为此，我们引入纤维之间的映射，通过“传输”得分，将点 p 处的得分与点 q处的得分对应起来。我们考虑统计学中更自然的特殊情形：指数传输和仿射传输。参见 Amari [1987] 以及 Kass 和 Vos [1997], ，其中纤维未经过中心化，所用名称为希尔伯特丛。另请参阅教程 Chirco 和 Pistone [2022] 以及 Pistone [2020]。

如定义 3.1 所示，对于所有 $p, q \in \Delta(\Omega)$，指数平行传输 为

$${}^{\mathrm{e}}\mathbb{U}_p^q v = v - \mathbb{E}_q\left[v\right]\ .$$

指数平行传输显然满足

$$\begin{aligned} &{}^{\mathrm{e}}\mathbb{U}_p^q \colon S_p\Delta\left(\Omega\right) \to S_q\Delta\left(\Omega\right)\ , \\ &{}^{\mathrm{e}}\mathbb{U}_q^r\,{}^{\mathrm{e}}\mathbb{U}_p^q = {}^{\mathrm{e}}\mathbb{U}_p^r\ , \quad {}^{\mathrm{e}}\mathbb{U}_p^p = I\ . \end{aligned}$$

一个初始速度为 $\dot{q}(0)$、初始位置为 $q(0)$ 的指数测地线 在通过传输识别速度空间时具有恒定速度，

$$\dot{q}(t) = {}^{\mathrm{e}}\mathbb{U}_{q(0)}^{q(t)}\dot{q}(0)\ ,$$

也就是说，

$$\dot{q}(t) = \left({}^{\mathrm{e}}\mathbb{U}_{q(0)}^{q(t)}\dot{q}(0)\right) q(t) = \left(\dot{q}(0) - \mathbb{E}_{q(t)}\left[\dot{q}(0)\right]\right) q(t)\,. \tag{21}$$

方程（21）在给定初始条件下的解为指数族

$$q(t) = \mathrm{e}^{t\dot{q}(0) - \psi(t)} \cdot q(0)\,. \tag{22}$$

注意到，$\operatorname{Supp} q(t) = \operatorname{Supp} q(0)$，因此

$$\psi(t) = \sum_{q(0)>0} q(0) \log \frac{q(0)}{q(t)} = \mathrm{D}\left(q(0) \,\|q(t)\right)\ .$$

若取 $t=1$，则从 p 到 q 的指数位移 $s_p(q)$定义为

$$q = \mathrm{e}^{s_p(q) - \mathrm{D}(p\,\|q)} \cdot p \quad \text{and} \quad \mathbb{E}_p\left[s_p(q)\right] = 0\,.$$

也就是说,

$$s_p(q) = \begin{cases} \log\frac{q}{p} - \mathbb{E}_p\left[\log\frac{q}{p}\right] & \text{on } \operatorname{Supp} p \\ \text{arbitrary} & \text{otherwise} \end{cases}$$

在 $\operatorname{Supp} p = \operatorname{Supp} q$ 的条件下。

反之 在坐标图中， $s_{q_0}(q(t)) = t\,\dot{q}_0$ 显示仿射形式。对于二阶导数，我们计算 ute

$$\lim_{h\to 0} h^{-1}\left({}^{\mathrm{e}}\mathbb{U}^{q(t)}_{q(t+h)}\dot{q}(t+h) - \dot{q}(t)\right)\,,$$

对于测地线，该值为零。

Amari 和 Nagaoka [2000] 引入了指数传输的对偶。协方差是 $S\Delta(\Omega)$ 上的一个度量，定义如下:

$$(q, v, w) \mapsto \langle v, w\rangle_q = \operatorname{Cov}_q(v, w)\,, \quad q \in \Delta(\Omega), v, w \in S_q\Delta(\Omega)\,. \tag{23}$$

对于所有的 $(p,u) \in S_p\Delta(\Delta(\Omega))$，以及 $(q,w) \in S_q\Delta(\Delta(\Omega))$，假设 $q\,{}^{\mathrm{m}}\mathbb{U}^q_p u = pu$，并计算:

$$\begin{aligned}\left\langle {}^{\mathrm{m}}\mathbb{U}^q_p u, w\right\rangle_q &= \sum_x q(x)\,{}^{\mathrm{m}}\mathbb{U}^q_p u(x) w(x) = \sum_x p(x)u(x)w(x) \\ &= \sum_x p(x)u(x)(w(x) - \mathbb{E}_p[w]) = \left\langle u, {}^{\mathrm{e}}\mathbb{U}^p_q w\right\rangle_p\end{aligned}$$

其中 ${}^{\mathrm{e}}\mathbb{U}^p_q w = (\text{支撑集 } p)(w - \mathbb{E}_p[w])$。

接下来，固定一个一般的 $(p,u) \in S\Delta(\Omega)$，并考虑 $(q,v) \in\in S\Delta(\Omega)$ 的方程

$$pu = qv$$

我们有

1. 若 $x \in$ 支撑集 $p^c \cap$ 支撑集 q^c，则方程变为 $0 = 0$2. 若 $x \in$ 支撑集 $p \cap$ 支撑集 q^c，则方程变为 $p(x)u(x) = 0$，也就是说， $u(x) = 0$3. 若 $x \in$ 支撑集 $p^c \cap$ 支撑集 q，则方程变为 $q(x)v(x) = 0$，也就是说， $v(x) = 0$4. 若 $x \in$ 支撑集 $p \cap$ 支撑集 q，则 $v = \frac{p}{q}u$ –

如果我们删除不一致的条目 2，即假设 $\operatorname{Supp} p \subset \operatorname{Supp} q$，那么条目 3 和条目 4 变为

$$\begin{aligned} v(x) &= 0 && x \in \operatorname{Supp} q \setminus \operatorname{Supp} p \\ v(x) &= \frac{p}{q}u && x \in \operatorname{Supp} p \end{aligned}$$

总之，对于 $\operatorname{Supp} p \subset \operatorname{Supp} q$，定义 ${}^{\mathrm{m}}\mathbb{U}^q_p S_p\Delta(\Omega) \to S_q\Delta(\Omega)$，其中 $pu = q\,{}^{\mathrm{m}}\mathbb{U}^q_p u$，以包含 Δ 的边界。

6.1 得分作为运动坐标系中的速度

作为本节的总结，我们注意到，在支撑恒定的情况下，得分等于运动坐标系中的仿射速度。考虑一条支撑集恒定的曲线 $t \mapsto \gamma(t)$，$\operatorname{Supp}\gamma(t) = S$。那么，对于所有满足 $\operatorname{Supp} p = S$ 的 $p \in \Delta(\Omega)$

$$\frac{d}{dt}s_p(\gamma(t)) = \frac{\dot\gamma(t)}{\gamma(t)} - \mathbb{E}_p\left[\frac{\dot\gamma(t)}{\gamma(t)}\right] \quad \text{on } S\text{, and } 0 \text{ otherwise.}$$

在运动坐标系中， $p = \gamma(t)$，因此

$$\left.\frac{d}{dt}s_p(\gamma(t))\right|_{p=\gamma(t)} = \frac{\dot\gamma(t)}{\gamma(t)} - \mathbb{E}_{\gamma(t)}\left[\frac{\dot\gamma(t)}{\gamma(t)}\right] = \dot\gamma(t)\,.$$

也就是说，先求关于任意坐标原点的导数，然后再将原点置于当前点。在更一般的情形下，不存在指数图，得分则是隐式定义的。

7 讨论

采用由位移（p，q）$\mapsto \log \frac{q}{p} - \mathbb{E}_p\left[\log \frac{q}{p}\right]$ 定义的仿射空间，有助于建立恰当的微积分，并将统计模型作为可微簇来讨论。代数统计学并不陌生于 Segre 簇和 Veronese 簇这类具有明确统计学解释的仿射对象。事实上，单项式统计模型及其环面对应模型在仿射坐标中表示为仿射子空间。本文论证了微分方法与代数统计学的结合不仅提供了 Fisher 统计方法所需的数学工具，也使得边界点能够纳入其中。

事实上，微分理论适用于开概率单纯形，这意味着所有统计模型都必须具有恒定支撑。本文提供了将该理论扩展到闭概率单纯形的工具。这是发展解析和代数方法的一步，可用于处理列联表等光滑但包含零值的模型。关键对象是费舍尔得分。每个光滑的一维模型 $\gamma(t)$ 都有如下形式的导数：$\dot{\gamma}(t) = \dot{\gamma}(t)\gamma(t)$，其中 $\dot{\gamma}(t)$ 是得分，可用作仿射几何中的速度。本文证明了费舍尔构造适用于闭单纯形。具体而言，我们证明模型可以在保持光滑的同时到达单纯形边界，而且常见的解析算法在仿射框架下具有代数形式，因此可计算。梯度流最小化算法就是一个例子，它需要对梯度和速度作出恰当定义。

统计模型传统上以参数形式呈现，较少采用隐式形式。显式引入一种更大的结构——统计丛，它同时包含概率函数和得分——便提供了物理学中所谓的状态空间。状态空间为描述统计模型提供了另一种方式：将其视为一个（可能是隐式的）微分方程组的解。测地线就是一个例子。以微分形式呈现模型，可直接应用于建模（例如 SIR模型等流行病模型）、近似计算（例如变分推断）以及物理感知建模（例如借鉴拉格朗日力学的模型）。

参考文献

John Aitchison。成分数据的统计分析。《统计学与应用概率专著》。Chapman & Hall，伦敦，1986年。ISBN 0-412-28060-4。doi:10.1007/978-94-009-4109-0。Shun-ichi Amari。统计推断中的微分几何。载于国际统计学会第*46*届会议论文集，第*2*卷（东京，*1987*年），《国际统计学会通报》第52卷，第321–338页，1987年。Shun-Ichi Amari。自然梯度在学习中高效发挥作用。神经计算，10(2):251–276，1998年2月。ISSN 0899-7667。doi:10.1162/089976698300017746。URL http://dx.doi.org/10.1162/089976698300017746。Shun-ichi Amari。信息几何及其应用：凸函数与对偶平坦流形。载于 Frank Nielsen 编，视觉计算新趋势，《计算机科学讲义》第5416卷，第75–102页。施普林格，2009年。ISBN 978-3-540-xxx。Shun-ichi Amari 和 Hiroshi Nagaoka。信息几何方法。美国数学学会，2000年。由 Daishi Harada 从1993年日文原著译出。Lawrence D. Brown。统计指数族基础及其在统计决策理论中的应用。数理统计学会讲义专著丛书第9号。数理统计学会，1986年。Goffredo Chirco 和 Giovanni Pistone。以巴拿赫空间为模型的对偶仿射信息几何，2022年。URL https://arxiv.org/abs/2204.00917。arXiv:2204.00917。Goffredo Chirco、Luigi Malagò 和 Giovanni Pistone。统计丛上概率的拉格朗日与哈密顿动力学。现代物理学几何方法国际期刊，19(13):2250214.1–46，2022年8月。doi:10.1142/s0219887822502140。URL https://doi.org/10.1142/s0219887822502140。Persi Diaconis 和 Bernd Sturmfels。用于从条件分布中抽样的代数算法。《统计年刊》，26(1):363–397，1998年。ISSN 0090-5364。Mathias Drton、Bernd Sturmfels 和 Seth Sullivant。代数统计学讲义，《奥伯沃尔法赫研讨会》第39卷。Birkhäuser Verlag，2009年。ISBN 978-3-7643-8904-8。doi:10.1007/978-3-7643-8905-5。URL http: //dx.doi.org/10.1007/978-3-7643-8905-5。Eliana Duarte、Orlando Marigliano 和 Bernd Sturmfels。最大似然估计量为有理数的离散统计模型。伯努利，27(1):135–154，2021年。ISSN 1350-7265,1573-9759。doi:10.3150/20-BEJ1231。URL https: //doi.org/10.3150/20-BEJ1231。

Emilie Dufresne、Heather A. Harrington 和 Dhruva V. Raman。松弛性的几何。《代数统计学杂志》，9(1): 30–68，2018年。ISSN 1309-3452。doi:10.18409/jas.v9i1.64。URL https://doi.org/10.18409/jas.v9i1.64。Bradley Efron 和 Trevor Hastie。计算机时代的统计推断，《数理统计学会专著》第5卷。剑桥大学出版社，纽约，2016年。ISBN 978-1-107-14989-2。URL https://doi.org/10.1017/CBO9781316576533。算法、证据与数据科学。Ronald A. Fisher。理论统计学的数学基础。《伦敦皇家学会哲学汇刊：数学或物理特征论文集》，222(594–604):309–368，1922年。doi:10.1098/rsta.1922.0009。Paolo Gibilisco、Eva Riccomagno、Maria P. Rogatin 和 Henry P. Wynn。统计学中的代数与几何方法。剑桥大学出版社，2010年。Alexander Heaton 和 Matthias Himmelmann。约束优化中的欧几里得距离与最大似然回缩映射计算。计算几何，126:102147，2025年。Robert E. Kass 和 Paul W. Vos。渐近推断的几何基础。《威利统计丛书：概率与统计》。约翰·威利父子出版公司，纽约，1997年。ISBN 0-471-82668-5。doi:10.1002/9781118165980。URL http://dx.doi.org/10.1002/9781118165980。威利-因特科学出版物。Anna Klimova 和 Tamás Rudas。关系模型拟合优度检验。《统计学通讯：理论与方法》，51(23):8264–8282，2022年。Anna Klimova、Tamás Rudas 和 Adrian Dobra。列联表的关系模型。《多元分析杂志》，104(1):159–173，2012年。Kei Kobayashi 和 Henry P. Wynn。有效估计中的计算代数方法。载于信息几何理论，《信号与通信技术》，第119–140页。施普林格，查姆，2014年。ISBN 978-3-319-05316-5; 978-3-319-05317-2。Lev D. Landau 和 Eugenij M. Lifshits。《理论物理学教程：力学》，第I卷。Butterworth-Heinemann，第3版，1976年。Lev D. Landau 和 Eugenij M. Lifshits。《理论物理学教程：统计物理学》，第V卷。Butterworth-Heinemann，第3版，1980年。Serge Lang。微分流形与黎曼流形，《数学研究生教材》第160卷。施普林格出版社，第3版，1995年。ISBN 0-387-94338-2。Josep Antoni Martín-Fernández、Javier Palarea-Albaladejo 和 Ricardo Antonio Olea。处理零值，第4章，第43–58页。约翰·威利父子出版公司，2011年。doi:10.1002/9781119976462.ch4。Giovanni Pistone。概率单纯形的信息几何：简明课程。《复杂系统中的非线性现象》，23(2):221–242，2020年。Giovanni Pistone 和 Carlo Sempi。在与给定概率测度等价的所有概率测度空间上构造无限维几何结构。《统计年刊》，23(5):1543–1561，1995年10月。ISSN 0090-5364。Giovanni Pistone 和 Henry P. Wynn。有限生成累积量。《统计学新纪元》，9(4):1029–1052，1999年10月。ISSN 1017-0405。Giovanni Pistone 和 Henry P. Wynn。累积量簇。《符号计算杂志》，41(2):210–221，2006年。ISSN 0747-7171。Giovanni Pistone、Eva Riccomagno 和 Henry P. Wynn。代数统计学：统计学中的计算交换代数，《统计学与应用概率专著》第89卷。Chapman & Hall/CRC，博卡拉顿，佛罗里达州，2001年。ISBN 1-58488-204-2。Calyampudi Radhakrishna Rao。统计参数估计中可达到的精度与信息。《加尔各答数学学会通报》，37:81–91，1945年。ISSN 0008-0659。Ralph Tyrrell Rockafellar。凸分析。《普林斯顿数学丛书》第28号。普林斯顿大学出版社，1970年。Laurent Schwartz。《分析学教程》。Hermann，1981年。Sumio Watanabe。代数几何与统计学习理论，《剑桥应用与计算数学专著》第25卷。剑桥大学出版社，剑桥，2009年。ISBN 978-0-521-86467-1。doi:10.1017/CBO9780511800474。URL https://doi.org/10.1017/CBO9780511800474。André Weil。《代数几何基础》，《美国数学学会专题论文集》第29卷。美国数学学会，纽约，1946年。修订扩充版于1962年出版。

赫尔曼·外尔。空间—时间—物质 / 作者：赫尔曼·外尔。多佛出版社，纽约，1952年。1921年《时空物质》的译本。