arXiv:2301.05655v2 [math.ST] 7 May 2023

概率集的收缩

Michele Caprio 和 Teddy Seidenfeld

摘要。给定一组概率测度 $\mathcal{P}$，它代表主体对某个 σ-代数 $\mathcal{F}$中各元素的知识，我们可以计算任意关注事件 $A \in \mathcal{F}$ 的概率上界和下界。如果某个生成新的信念评估的程序 A 使得该程序之后事件 A 的概率界限包含于程序之前的界限之中，那么就称该程序具有收缩性。已有充分文献记载，对于所有 $A \in \mathcal{F}$，（广义）贝叶斯更新都不能实现收缩。在本文中，我们证明，无论是否观察到证据，都可能发生收缩，并对这些可能性进行了刻画。

1. 引言

将 $\Delta(\Omega\ \mathcal{F})$称为所关注可测空间 $(\Omega\ \mathcal{F})$ 上可数可加概率测度的空间，并令 $\mathcal{P} \subseteq \Delta(\Omega\ \mathcal{F})$为一组概率测度。于是，$\underline{P}(A) = \inf_{P\in\mathcal{P}} P(A)$称为 A的下概率，而其共轭$\overline{P}(A) = 1 - \inf_{P\in\mathcal{P}} P(A^c) = \sup_{P\in\mathcal{P}} P(A)$称为 A的上概率。它们是非精确概率理论 [39]文献中的两个主要构成要素。与测度论中若一个集合的外测度和内测度相等，我们就说该集合有测度类似，若上概率和下概率相等，则它们构成一个普通概率测度，且 $\mathcal{P}$ 是单元素集。我们刚才所作的非精确概率理论与测度论之间的类比并非仅仅是启发性的：[39, 第3.1节5] 表明，将相干下概率从代数自然延拓到幂集，所得结果就是相应的内测度。

关于研究非精确概率的理由，[2, 39] 及其中的参考文献已进行了详细讨论；在本文中，我们尤其关注[16, 30]所阐述的动机。在那里，作者指出，指定概率集——从而指定其“边界元素”，即下概率和上概率——能够反映执行分析的主体所面临的模糊性。这意味着，由于主体不知道支配相关实验的真实数据生成过程，他们可能希望利用非精确概率理论的灵活性，指定一组概率测度来表示其无知。主体所面临的不确定性越高，该集合就会越“宽”，也就是说，对于所有 $A \in \mathcal{F}$，$\overline{P}(A)$与 $\underline{P}(A)$之间的差异都会更大。

本文旨在研究给定程序之后发生的收缩现象。

定义1。考虑一个关注的事件 $A \in \mathcal{F}$、一个通用的概率测度集合$\mathcal{P} \subseteq \Delta(\Omega\ \mathcal{F})$，并分别用 $\underline{P}$ 和 $\overline{P}$表示与 $\mathcal{P}$相关联的下概率和上概率，

2010 数学主题分类。主要分类：62A01；次要分类：60A10、60A99。关键词和短语。收缩、扩张、概率集、证据、条件化、遗忘。

分别。称 $\dot{\beth}$ 为一种通用程序，用于生成新的信念评估；并分别用 $\underline{P}_{\dot{\beth}}$ 和 $\overline{P}_{\dot{\beth}}$ 表示该程序产生的下概率和上概率。[1]那么，如果程序 $\dot{\beth}$（严格）一致收缩 A，记作 $\dot{\beth} \not\hookleftarrow A$，则有 $\underline{P}_{\dot{\beth}}(A) > \underline{P}(A)$ 且 $\overline{P}_{\dot{\beth}}(A) < \overline{P}(A)$。如果两个不等式中有一个是弱不等式，我们就称 $\dot{\beth}$ 弱一致收缩 A。

显然，如果 $\dot{\beth}$ 对 A进行严格一致收缩 $\dot{\beth}$ 对 A进行弱一致收缩[8, 19, 23]。在本文余下部分，我们将一致收缩（UC）简称为“收缩”。UC 有时也称为收缩 [24] 。我们更倾向于使用“收缩”这一术语——如 [29]所示——因为在信念修正文献中，“收缩”用于指完整信念的一种可修正性实例；当主体通过转向逻辑上较弱的证据集合而放弃某些当前证据时，就会发生这种情况 $\dot{}$。此外，当我们说 $\dot{\beth}$ 产生新的信念评估时，我们的意思是，程序 $\mathcal{F}$输出一组概率，用以表示主体对其中各元素的信念 [23]。这不应与 AGM理论相混淆；在 AGM理论中，程序会生成一组新的完整信念（句子集合）。我们沿用相同的术语，因为在此语境下不会产生混淆。

我们对收缩的兴趣源于涉及相反现象的令人惊讶的结果，这种现象称为扩张，最初是在对 $\mathcal{P}$ [36]进行（广义）贝叶斯更新的背景下观察到的。我们指出，$\mathcal{P}$ 不必是闭集或凸集。任取一个 $A \in \mathcal{F}$，并将 $\mathcal{P}(A) := \{P(A) : P \in \mathcal{P}\}$称为。令 $X : \Omega \to \mathbb{R}$ 为一个 $\mathcal{P}$-可测随机变量，也就是说，对所有 $P \in \mathcal{P}$，它都是 P-可测的。于是称 $\mathcal{I}$ 为一个一般指标集，并令 $\mathbf{X} := \{X = x_i\}_{i\in\mathcal{I}}$ 为与 X相关联的可测事件的样本空间。用$\mathcal{P}(A \mid x_i) := \{P(A \mid X = x_i) : P \in \mathcal{P}\}$ 表示事件 A在给定$X = x_i$条件下的条件概率集合。为避免在 P-零事件 $P \in \mathcal{P}$条件下定义条件概率所产生的问题，我们假设 $\mathcal{P}$ 中的 P's 在那些结果 $\{X = x_i\}$ 上达成一致，而这些结果是 P-零的。我们将 $E = X^{-1}(x) \subseteq \Omega$ 记为实验 $\mathbf{X}$ 之后收集的证据。收缩的一个较弱概念如下。

定义 2. 若（广义）贝叶斯更新程序记作 $\dot{\beth} = (B, E)$,[2] ，则当对一组 $\mathcal{P}$-概率为 1 的每个 x_i（即，对所有 $P \in \mathcal{P}$，一组 P-概率为 1 的 x_i）而言，$\underline{P}(A) < \underline{P}(A \mid X = x_i)$ 且 $\overline{P}(A) > \overline{P}(A \mid X = x_i)$ 时，该程序会对 A 产生严格的逐点收缩。如果两个不等式之一为弱不等式，则该程序会对 A 产生弱逐点收缩。

广义贝叶斯更新仅对 A 产生逐点收缩，是指它对 A 产生逐点收缩，但 $\inf_{i\in\mathcal{I}} \underline{P}(A \mid X = x_i) = \underline{P}(A)$ 且 $\sup_{i\in\mathcal{I}} \overline{P}(A \mid X = x_i) = \overline{P}(A)$。严格和弱的仅逐点收缩定义与前述类似。

要看出逐点收缩弱于一致收缩，请注意，当 $\inf_{i\in\mathcal{I}} \underline{P}(A \mid X = x_i) > \underline{P}(A)$ 且 $\sup_{i\in\mathcal{I}} \overline{P}(A \mid X = x_i) < \overline{P}(A)$ 时，我们得到严格一致收缩；若后两个不等式之一为弱不等式，则得到弱一致收缩。

[1]我们用希伯来字母 $\beth$ 表示该程序，因为希伯来语中表示程序的词נוהל（发音为$nohal$）以 $\beth$ 开头。此外，拉丁语、希腊语和西里尔语字母 p, P, π, Π——这些字母可能会让人联想到“程序”一词——通常与概率和划分相关，而希腊字母 ϖ 很容易与 ω 混淆；我们将用 ω 表示所关注的状态空间 Ω 中的一个元素。[2]在 $\beth = (B, E)$ 中，字母 B 表示广义贝叶斯更新，而 $E = X^{-1}(x)$ 是条件集。全文中，我们将广义贝叶斯更新简称为“条件化”，而将其他技术称为“更新规则”。

现在，对所有 $P \in \mathcal{P}$，定义 $\mathbf{X}_P^{A+} := \{x \in \mathbf{X} : P(A \mid X = x) > P(A)\}$ 和 $\mathbf{X}_P^{A-} := \{x \in \mathbf{X} : P(A \mid X = x) < P(A)\}$。下面的引理可由条件期望定律直接推出。

引理3。 任取 $A \in \mathcal{F}$。那么，对于所有 $P \in \mathcal{P}$， $P(\mathbf{X}_P^{A+}) > 0$ 当且仅当 $P(\mathbf{X}_P^{A-}) > 0$。

于是我们得到两个命题，它们给出了充分条件，表明当以实验的形式收集证据以学习随机变量 X 的值，并使用广义贝叶斯更新时，甚至无法实现弱收缩。

命题4。 任取 $A \in \mathcal{F}$。如果 $\mathcal{P}(A)$ 在欧几里得拓扑中是闭集，则不存在实验 $\mathbf{X}$ 使得 (B, E) 弱一致收缩 A。

命题5。 任取 $A \in \mathcal{F}$。不存在简单实验 $\mathbf{X}$ 使得 (B, E) 弱逐点收缩 A。也就是说，如果 X 是一个简单随机变量（即，如果样本空间 $\mathbf{X}$ 的指标集 $\mathcal{I}$ 是有限的，并且对于所有 $P \in \mathcal{P}$， P-概率为 1），那么 (B, E) 不会弱逐点收缩 A。

下一个例子表明，如果 $\mathcal{P}(A)$ 是开集，且 X 不是简单的，那么我们可以得到严格的纯逐点收缩。结合命题 4和 5，这就穷尽了广义贝叶斯更新的所有可能情形。

例 6. 设 A 是一个（可测）事件。对于 $0.4 < x < 0.6$，规定 $P_x(A) = x$，并令 $\mathcal{P}(A) = \{P_x(A)\}$。因此， $\underline{P}(A) = 0.4$ 且 $\overline{P}(A) = 0.6$，但 $\mathcal{P}(A)$ 是一个开集。设 f是开区间 $(0.4, 0.6)$ 内有理数的一个良序，记作集合$\mathbb{Q}_{(0.4,0.6)}$。令 $\mathbb{N} = \{1, 2, \ldots\}$ 表示自然数，则有 $f : \mathbb{Q}_{(0.4,0.6)} \leftrightarrow \mathbb{N}$是一个 $1-1$ （满射）函数。记 $q_n = f^{-1}(n)$。同样， $q_n \in \mathbb{Q}_{(0.4,0.6)}$。设 N 是一个P_X-可测随机变量，对于每个 $0.4 < x < 0.6$，其似然比满足

$$\frac{P_x(N = n \mid A)}{P_x(N = n \mid A^c)} = \frac{(1-x)q_n}{x(1-q_n)}. \tag{1}$$

请注意，(1) 对分布施加了约束 $P_x(N)$，但并未定义它。不过，(1) 是自洽的，因为对于每个 $0.4 < x < 0.6$，都存在无穷多个 q_n 值，使得似然比大于1；也存在无穷多个 q_n 值，使得该比率小于1。[3] 根据贝叶斯定理的一个简单推论，

$$\frac{P_x(A \mid N = n)}{P_x(A^c \mid N = n)} = \frac{q_n}{1-q_n},$$

它在 $\mathcal{P}(A)$ 上为常数。也就是说，相对于集合 $\mathcal{P}(A)$，给定 $N = n$ 时 A 的条件概率族是确定的，尽管 A 的无条件概率族是不确定的。因此，对于每个 $n \in \mathbb{N}$，

$$0.4 < \underline{P}(A \mid N = n) = q_n = \overline{P}(A \mid N = n) < 0.6$$

[3]在这里，我们可以看到，条件 $\mathcal{P}(A)$ 要求其为开集的必要性所在。当$P_x(A) = 0.4$ 或 $P_x(A) = 0.6$时，条件1) 不自洽。此时，对于 q_n的所有取值，似然比 (1) 的取值都只会落在1的一侧，这与全概率公式相矛盾。

而给定 $N = n$ 的贝叶斯更新仅在逐点意义上严格收缩 A。请注意，这种严格收缩在 $\mathbb{N}$ 上并非一致的，因为 $\inf_{n\in\mathbb{N}} \underline{P}(A \mid N = n) = 0.4 = \underline{P}(A)$，且 $\sup_{n\in\mathbb{N}} \overline{P}(A \mid N = n) = 0.6 = \overline{P}(A)$。

现在我们举一个扩张的例子，该例改编自 [36]。它说明，使用广义贝叶斯更新时，对于实验的每一种可能结果，事件 A 的非精确概率都会变得更加不精确。

例7。 假设我们连续两次抛掷一枚公平硬币，并且两次抛掷的结果可能相互依赖。分别用 H_i 和 T_i 表示第 $i \in \{1, 2\}$次抛掷的结果“正面”和“反面”。令

$$\mathcal{P} := \left\{ P : P(H_1) = P(H_2) = \frac{1}{2},\ P(H_1 \cap H_2) = p \right\}_{p\in\left[0,\frac{1}{2}\right]}.$$

现在，假设我们抛掷这枚硬币；我们有 $P(H_2) = 1/2$，但是

$$0 = \underline{P}^B(H_2 \mid H_1) < \underline{P}(H_2) = \frac{1}{2} = \overline{P}(H_2) < \overline{P}^B(H_2 \mid H_1)$$

and

$$0 = \underline{P}^B(H_2 \mid T_1) < \underline{P}(H_2) = \frac{1}{2} = \overline{P}(H_2) < \overline{P}^B(H_2 \mid T_1),$$

其中 $\overline{P}^B(H_2 \mid H_1) = \overline{P}^B(H_2 \mid T_1) = 1$。正如我们所见，我们起初对第二次抛掷有一个精确的信念，但无论第一次抛掷的结果如何，最终对第二次抛掷都只有空洞信念。

贝叶斯条件化规则——可以说是最广为人知的信念更新程序——可能导致扩张，这正是我们探索能够实现收缩的更新技术的动机之一。这也是我们这项工作的重点。

本文结构如下。第2 节研究在没有收集到新证据时，允许发生收缩的程序。定理10 和11 是主要结果，给出了非常一般的条件，使程序能够在没有收集到新证据时提供发生收缩的机会。第3节中，群体中的每个个体都对其他人提供的“精确”输入应用一个（凸）个人汇总规则，从而形成修正后的观点。这个过程不断迭代，直到个体观点汇合到一个不动点。由于汇总规则是凸的，这个不动点是原始观点集合的收缩。第4节研究在收集到证据且使用非贝叶斯更新规则来修正主体信念时的收缩。对于可数可加概率，给定 Ω 的一般划分 $\mathcal{E}\Omega$，条件化并不能对所有 $E \in \mathcal{E}$都实现收缩。因此，要对所有 $E \in \mathcal{E}$ 实现收缩，唯一的方法就是有意忘记与 $\mathcal{E}$相关的实验。但是，如果我们能够对 $\underline{P}$的性质作出假设，就可以给出对所有 $E \in \mathcal{E}$实现收缩的条件。第5 节总结全文。附录A中，我们研究放弃可列可加概率这一假设后实现收缩的可能性；附录B中，我们证明本文的结果。

2. 无证据情况下的约束

本节研究在未收集任何数据时，能够提供收缩机会的程序。

2.1. **精确概率的一致扩展**。 回想一下，对德·芬内蒂而言，概率测度 P 对于状态空间 Ω 中任意有限个非空子集 $\sup_{\omega\in\Omega}\sum_{i=1}^{n} c_i[I_{A_i}(\omega) - P(A_i)] \geq 0$，对所有 c_1，.....，$c_n \in \mathbb{R}$，若满足以下条件，则称其具有一致性：其中 I_{A_i} 表示集合 A_i的示性函数。德·芬内蒂概率基本定理 [11,第3节。10] 内容如下。

定理 8. 令 Ω 为所关注的状态空间。给定有限个事件 $A_1, \ldots, A_n \subseteq \Omega$ 的概率 $P(A_i)$，另一个事件 A_{n+1} 的概率 $P(A_{n+1})$

(1) 若 A_{n+1} 与 A_i's 线性相关，则该概率最终会被确定；(2) 否则，可以一致地将其赋予闭区间 $[p', p'']$ 内的任意值。

更确切地说， p' 是以下各项的最大下界（GLB）：上确界 $P(X)$ 由以下随机量给出的 $P(X)$ 从下方进行的评估， X 这些随机量与 A_i's 线性相关，且我们确定有 $X \leq A_{n+1}$。[4] 对于 p'' 也可以作同样的说明（将 上确界 替换为 下确界，最大值替换为最小值， A'_{n+1} 替换为 A''_{n+1}，并改变不等式方向等。它是从上方进行的评估的最小上界）。

请注意， $[p', p'']$ 可能是一种虚幻的限制，例如，如果 $p' = 0$ 和 $p'' = 1$。这一结果的解释如下。假设我们通过精确概率分布 P来表达对事件 A_1、.....、A_n 的主观信念。 P是精确的，这一点是德·芬内蒂主观概率理论的一项关键原则。于是，如果我们想将信念一致地扩展到一个感兴趣的新事件 A_{n+1}，那么要么当 A_{n+1} 是其他事件的线性组合时，我们可以"免费"做到这一点；要么我们会得到一个区间 $[p', p'']$，可以从中选择一个值赋给 P（A_{n+1}）。德·芬内蒂本人并未具体说明如何在 $[p', p'']$中选择一个值。其要旨似乎是："你应该足够认真地思考，从而为 P（A_{n+1}）想出一个可以赋予它的精确数值 $p \in [p', p'']$"。

将 ɔ = deFin 记为：从 $[p', p'']$ 中任取一个值，赋给事件 A_{n+1} 的概率这一程序。则有以下结论。

定理 9。 假设——采用定理 8 中的记号—— $p' \neq p''$。那么，当 $P_{\text{deFin}}(A_{n+1}) \in (p', p'')$ 时，deFin $\nleftrightarrow A_{n+1}$；若 $P^{\text{deFin}}(A_{n+1}) \in \{p', p''\}$，则收缩是弱的。

回顾一下，对 Walley 而言，下概率测度 $\underline{P}$ 是相干的，当且仅当对于感兴趣的状态空间 Ω 的任意有限非空子集集合族$\{A_i\}_{i=0}^{n}$，都有

$$\sup_{\omega\in\Omega}\left[\sum_{i=1}^{n}\left(I_{A_i}(\omega) - \underline{P}(A_i)\right) - s\left(I_{A_0}(\omega) - \underline{P}(A_0)\right)\right] \geq 0,$$

对所有 s， $n \in \mathbb{Z}_+$。在 [39, 第3节。1] 作者给出了定理 8 的非精确概率（IP）对应版本。也就是说，Walley 提出了一种方法，可以将定义在有限集族上的相干下概率和上概率 $\underline{P}(A_i)$， $\overline{P}(A_i)$ 扩展到任何其他 $\{A_1, \ldots, A_n\} \subseteq 2^{\Omega}$ Ω $A_{n+1} \subseteq$Ω。然而，与德·芬内蒂基本定理不同，这一结果并非旨在促成收缩。非精确概率理论的核心思想是，让人们能够"自在地"使用概率集，而不必被迫从集合中选取一个精确值。

[4]这个不等式应解释为：对于所有 $\omega \in$Ω， $X(\omega) \leq I_{A_{n+1}}(\omega)$。

请注意，定理 8 和 9，以及 [39, 第3节。1], 中的结果，分别是在德·芬内蒂和 Walley 的框架下给出的。两者都基于有限可加概率。不过，由于我们考虑的是有限多个事件，这一区别无关紧要。尽管如此，重要的是要指出，德·芬内蒂和 Walley 的扩展程序也适用于从任意（可能无限的）事件集出发的情况；此时有限可加性与可列可加性的区别便至关重要，例如参见附录 A。

令$\#$ 表示基数算子，Conv(H)表示一般集合的凸包 H，ex$[K]$ 表示一般凸集 K的极点。我们可以将定理9 推广如下。

定理10。设一般程序 $\lrcorner$ 生成一个在 $(\Omega\ \mathcal{F})$ 上的概率集合 $\mathcal{P} \subseteq \Delta(\Omega\ \mathcal{F})$，满足 $\#\mathcal{P} \geq 2$，随后规定一种从 Conv$(\mathcal{P})$ 中选取一个元素 $P^\star = \underline{P}_\lrcorner = \overline{P}_\lrcorner$的方法。假设 $\emptyset \neq$ ex$[$Conv$(\mathcal{P})$ $] = \{P_j^{ex}\}_{j\in\mathcal{J}}$，其中 $\mathcal{J}$ 是一般指标集。那么，我们有

- 若 $P^\star \in$ ex$[$Conv$(\mathcal{P})]$, ，则可能存在一个集合族 $\{A\} \subseteq \mathcal{F}$，使得 $\lrcorner$ 弱收缩 $\tilde{}$ $\tilde{}$ A。此外，对所有 $A \in \mathcal{F} \setminus \{A\}$，$\lrcorner \nleftrightarrow A$;
- 如果改为 $P^\star = \sum_{j\in\mathcal{J}} \alpha_j P_j^{ex}$，那么对所有 j，$\alpha_j > 0$ 成立，$\lrcorner \nleftrightarrow A$，对所有 $A \in \mathcal{F}$。

我们也可以给出定理 10的一个拓扑版本；称 $\partial_X H$和 int$_X H$为一般集合 H 在 X中的边界和内部。

定理 11。 赋予 $[0,1]$ 欧几里得拓扑，并称 $\mathcal{B}([0,1])$ 为 $[0,1]$上的博雷尔 σ-代数。固定一个一般的 $A \in \mathcal{F}$，并假设 $\mathcal{P}(A) := \{P(A) : P \in \mathcal{P}\} \subseteq \mathcal{B}([0,1])$ 且 $\#\mathcal{P}(A) \geq 2$。那么，

- 如果 $\mathcal{P}(A)$ 在欧几里得拓扑下是闭集，且 $P^\star(A) \in \partial_{\mathcal{B}([0,1])}\mathcal{P}(A)$，那么 $\lrcorner$ 弱收缩 A;
- 若改为 $P^\star(A) \in$ int$_{\mathcal{B}([0,1])}\mathcal{P}(A)$，则 $\lrcorner \nleftrightarrow A$。

注12。 请注意，当$\mathcal{P}$ 是凸集时，假设 $\mathcal{P}(A) \subseteq \mathcal{B}([0,1])$ 得到了满足。

显而易见，定理 9 是定理 10 和定理 11 的特例。另一个符合定理 10 和定理 11 要求的程序是 Halmos 扩张 [21, 练习 48。4], [6, 第4节。13]。考虑两个一般的可测空间 $(X, \mathcal{X})$ 和 $(Y, \mathcal{Y})$。令 $\mu \in \Delta(X, \mathcal{X})$，并且对所有 $x \in X$，$\nu_x \in \Delta(Y, \mathcal{Y})$。进一步假设，对所有 $B \in \mathcal{Y}$,$\nu_\bullet(B)$: $X \to [0,1]$ 是 $\mathcal{X}$-可测的。那么，

(i) 对所有 $E \in \mathcal{X} \times \mathcal{Y}$，映射 $E \mapsto \nu_x(\{y \in Y$: $(x,y) \in E\})$ 都是 $\mathcal{X}$-可测的；(ii) 映射 π: $\mathcal{X} \times \mathcal{Y} \to [0,1]$,

$$E \mapsto \pi(E) := \int_X \nu_x\left(\{y \in Y : (x,y) \in E\}\right)\mu(\mathrm{d}x)$$

是 $\mathcal{X} \times \mathcal{Y}$ 上的概率测度。

现在假设存在一个集合 $A \subseteq Y$，其内测度与外测度不相等，即对于所有 $x \in X$， $\nu_{\star x}(A) \neq \nu_x^\star(A)$。接着，考虑 $A' = X \times A$。我们有

$$\pi_{A'} \equiv \pi(A') = \int_X \nu_x\left(\{y \in Y : (x,y) \in A'\}\right)\mu(\mathrm{d}x)\,,$$

且 $\pi_{A'} \in [\nu_{\star x}(A)\ \nu_x^\star(A)]$。因此，Halmos 扩张规定了一种方法，将 $\mathcal{Y}$ 上的可列可加概率测度扩张为$\mathcal{X} \times \mathcal{Y}$ 上的另一个可列可加概率测度，从而为一个 $\mathcal{Y}$-不可测集合 A赋予定义良好的测度。该值分别属于以 A的内测度和外测度为端点的区间。由此可立即看出，Halmos 扩张满足定理10和11中的条件。

注 13. 继续之前，我们需要指出 ɔ = 德·芬内蒂与测度论中的扩张定理（如哈尔莫斯的做法）之间一个值得注意的差异。相关的对比在于，对于基本定理（应用于概率时），德·芬内蒂将示性函数等的线性张成作为其定义域。而对于测度论学者，将概率扩张到更大的集环时，会使用示性函数的（可数）和来定义：有限情形下，这些示性函数来自一个代数；无限情形下，则来自一个 σ-代数。另有两个重要差异：（1）对于有限结构，德·芬内蒂不要求概率定义在一个代数上，而其他学者则要求；（2）对于无限结构，由于德·芬内蒂不要求可列可加性，他的内外逼近是通过示性函数的有限和得到的。相比之下，测度论学者要求概率是可列可加的，因此他们使用可数和来构造内测度和外测度的逼近。

现在我们简要 介绍三种符合定理 10 和 11的要求的程序

：(i) 凸池化 [38, 第 2], 节中， k 主体（以精确概率 P_j表示 $j \in \{1, \ldots, k\}$）的意见先以凸的方式进行池化，从而形成一个集合

$$\mathcal{P} = \left\{ P \in \Delta(\Omega, \mathcal{F}) : P = \sum_{j=1}^{k} \zeta_j P_j \right\},$$

其中，对于所有 $j \in \{1, \ldots, k\}$ 和 $\sum_{j=1}^{k} \zeta_j = 1$， $\zeta_j \geq 0$，并从中选出唯一的池化意见 $P^\star$；

(ii) *Jaynes* 最大熵 [26], ，给定一组约束条件 C 后，研究者感兴趣的概率集合为 $\mathcal{P} = \{P \in \Delta(\Omega\ \mathcal{F}) :$ P满足 C$\}$，并通过最大化 $\mathcal{P}$中的香农熵来选出 $P^\star$；

(iii) 广义信赖推断 (GFI) [22], ，通过谨慎地反演一个确定性数据生成方程（不使用贝叶斯定理），定义参数空间 Ω 上一组依赖于数据的测度，称为广义信赖分布。数学上，我们可以写作 $\mathcal{P} = \{P \in \Delta(\Omega\ \mathcal{F}) :$ P满足 [22, 方程 (2)]$\}$。正如 [22, 备注 4] $P^\star$ 所指出的，此时通过选择适当的范数来赋予样本空间范数。作者在[22, 第 1],节指出，尽管 GFI 在哲学上不同于邓普斯特-谢弗理论 [13] 和推断模型[31], ，但这三种方法得出的结果在数学上往往彼此密切相关。

3. 基于凸池化的约束

作为一种允许主体收集新证据、但不通过条件化来更新主体信念的程序示例，我们介绍 [12] 中著名的模型。该模型中，作者假设有 k 个体，每个个体都有自己的主观概率

分布 F_i，用于表示某个未知参数 $\omega \in \Omega$ 的取值。[5] 对于主体 i 而言，所有其他 $k-1$ 主体的观点都构成了新证据。主体不会以这些观点为条件进行条件化，而是i 将自己的观点与其他主体的观点汇总起来。DeGroot 指出，对每个主体 i重复这一过程后，群体最终会（渐近地）就一个共同的主观概率分布达成一致。更新观点后，群体中每个成员的概率分布都属于集合

$$\mathcal{P} = \left\{ F = \sum_{j=1}^{k} \zeta_j F_j \right\}, \tag{2}$$

其中 $\zeta_j \in [0,1]$，对所有 $j \in \{1,\dots,k\}$，且 $\sum_{j=1}^{k} \zeta_j = 1$。特别地，对所有 $i \in \{1,\dots,k\}$，我们记第一次（汇聚）迭代之后主体i更新后的概率测度为 F_{i1}，其定义为 $F_{i1} = \sum_{j=1}^{k} p_{ij} F_j$。这意味着个体 i 通过系数 p_{i1}、……、p_{ik} 来权衡所有主体（包括自身）的意见，这些系数表示主体 i 赋予群体其他成员意见的相对重要性。由于这对所有主体都成立，我们可以用线性代数符号表示这一更新过程。将 $\mathbf{P}$ 称为 $k \times k$ 随机矩阵，其各行由概率向量 $(p_{i1}, \dots, p_{ik})$ $i \in \{1,\dots,k\}$构成。再令 $\mathbf{F} = (F_1, \dots, F_k)^\top$；则有 $\mathbf{F}^{(1)} = \mathbf{PF}$，其中 $\mathbf{F}^{(1)} := (F_{11}, \dots, F_{k1})^\top$。当然，这对所有迭代都成立，因此进而有 $\mathbf{F}^{(n)} = \mathbf{P}\mathbf{F}^{(n-1)} = \mathbf{P}^n\mathbf{F}$，对所有 $n \in \mathbb{N}$均成立。成员会无限期地持续进行这些修正，或者直到对某个 $N \in \mathbb{N}$，有 $\mathbf{F}^{(n)} = \mathbf{F}^{(n-1)}$，对所有 $n \geq N$ 均成立；此时进一步修正不会改变成员的意见。以下是 [12] 的主要结果。

定理 14. 若存在 $n \in \mathbb{N}$，使得 $\mathbf{P}^n$ 至少有一列中的每个元素均为正数，则可以达成共识。

也就是说，若满足定理 14 中的条件，则存在一个 $k \times 1$ 维向量 $\boldsymbol{\pi} = (\pi_1, \dots, \pi_k)$（根据 [12, 定理 3]，该向量是唯一的），其元素均非负且总和为 1，并且满足 $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$。进而，这意味着若将行向量均相同且等于 $\boldsymbol{\pi}$ 的随机矩阵称为 $\boldsymbol{\Pi}$，则有 $\mathbf{F}^\star = \boldsymbol{\Pi}\mathbf{F}$，其中 $\mathbf{F}^\star = (F_1^\star, \dots, F_k^\star)^\top$，满足 $F^\star = F_1^\star = \cdots F_k^\star = \sum_{j=1}^{k} \pi_j F_j$；这里的 $F^\star$ 是共识中形成的共同主观分布。注意，$F^\star$ 属于式 (2) 中的 $\mathcal{P}$。

对所有 $n \in \mathbb{N}$，称 $\mathcal{P}_n := \mathsf{Conv}(F_{1n}, \dots, F_{kn})$。在此例中，对所有 $n \in \mathbb{Z}_+$，我们有 $\mathcal{P}_n \subseteq \mathcal{P}_{n-1}$，其中 $\mathcal{P}_0$ 是在方程 (2) 中设定的集合 $\mathcal{P}$。这意味着序列 $(\mathcal{P}_n)$ 的极限是集合$\cap_{n\in\mathbb{Z}_+}\mathcal{P}_n$。如果满足定理 14 中的条件，那么 $\cap_{n\in\mathbb{Z}_+}\mathcal{P}_n = \{F^\star\}$。给定一个一般集合 $A \in \mathcal{F}$，DeGroot 程序 $\beta = \mathsf{DeGr}$ 通常只能弱收缩 A；例如，当 $F^\star$ 属于 $\mathcal{P} \equiv \mathcal{P}_0$ 的极值时便是如此。也就是说，$F^\star(A) \geq \underline{F}(A)$ 且 $F^\star(A) < \overline{F}(A)$，或者$F^\star(A) > \underline{F}(A)$ 且 $F^\star(A) \leq \overline{F}(A)$。然而，可能存在 $r, s \in \mathbb{Z}_+, r < s$，使得$\underline{F}_s(A) > \underline{F}_r(A)$ 且 $\overline{F}_s(A) < \overline{F}_r(A)$，其中 $\underline{F}_s(A) = \inf_{F\in\mathcal{P}_s} F(A)$，$\overline{F}_s(A) = \sup_{F\in\mathcal{P}_s} F(A)$，而 $\mathcal{P}_r$ 也类似。

[5] 通常，参数空间 Θ 中的元素用 θ表示，而状态空间 Ω 中的元素用 ω表示。由于 DeGroot 模型只关注参数空间，我们仅在本节中使用$\omega \in \Omega$ 来表示参数空间，以便与其他章节的记号保持一致。

德格鲁特模型是这样一种可能的例子：主体收集证据后，使用不同于对所收集数据进行条件化的规则来修正其初始意见。我们证明，至少存在一种这样的程序：如果使用了所有收集到的信息，就可能发生收缩。

请注意，与 DeGroot 的共识模型不同，根据命题4，在奥曼重要的“同意彼此意见不合”模型 [3] 中不存在收缩的机会。这里概述一下原因。将群体关于所关注事件 A 的初始概率IP集$\mathcal{P}(A)$，定义为每个主体获知其“私人”信息后，对 A 得出的精确个体意见的最小值和最大值所对应的下概率和上概率 $\underline{P}(A)$ 和 $\overline{P}(A)$。也就是说，在奥曼过程中，第1轮得到 $\mathcal{P}(A)$。由于群体中的主体数量有限，这个概率IP集是闭集。在奥曼模型中，主体随后反复分享各自精确的个体概率，所针对的是 A。从第一轮之后的每一轮起，他们都会运用贝叶斯更新，根据新共享的证据——即了解到的其他主体对 A 的概率——修正自己对 A 的个体概率。这个程序相当于针对闭的概率IP集 $\mathcal{P}(A)$，根据事件 A 的更新后个体精确概率，反复运用广义贝叶斯更新。根据奥曼定理，经过有限轮后，该过程会到达一个不动点，并形成共识 $P^{\star}(A)$。但是，如果 $\underline{P}(A) < \overline{P}(A)$，那么命题4表明，共识意见 $P^{\star}(A)$ 不可能始终满足 $\underline{P}(A) < P^{\star}(A) < \overline{P}(A)$。

4. 基于非贝叶斯更新的收缩

假设某个实验的结果诱导出感兴趣的状态空间 Ω 的一个划分 $\mathcal{E} = \{E_j\}$ 。那么，如果我们保留概率测度可数可加这一假设，对 E_j 进行条件化就无法实现收缩，对所有 $E_j \in \mathcal{E}$皆如此。因此，一般而言，对所有 $E_j \in \mathcal{E}$ 进行收缩只有在我们有意忘掉诱导出划分 $\mathcal{E}$的整个实验时才可能发生。但如果我们能够对与概率集合 $\mathcal{P}$ 相关联的下概率 $\underline{P}$的性质作出假设，且该概率集合表示主体的信念，并且考虑不同于贝叶斯条件化的更新程序，那么我们就有机会实现收缩。

4.1. 背景。 本节介绍一些背景概念，以便更好地理解接下来的结果。下概率和上概率（分别简称 LP 和 UP）是一种特殊类型的肖凯容量。

定义15。 给定一个可测空间 $(\Omega, \mathcal{F})$，其中 $\Omega \neq \emptyset$，我们称集函数$\nu: \mathcal{F} \to [0,1]$ 为肖凯容量，若 $\nu(\emptyset) = 0$、 $\nu(\Omega) = 1$，且对于所有满足 $A \subseteq B$ 的 A、$B \in \mathcal{F}$，均有 $\nu(A) \leq \nu(B)$。

记 $\mathcal{M} := \{P \in \Delta(\Omega, \mathcal{F}): \ P(A) \geq \underline{P}(A), \ \forall A \in \mathcal{F}\}$ 与 相容的 （可数可加）概率测度 $\underline{P}$ [19] 所构成的集合族，并假设该集合非空且相对紧。另请注意， $\mathcal{M}$ 是凸的。以下是下概率的特例。

定义16。线性规划 $\underline{P}$ 是一个阶 *Choquet* 容量 k，或称为 k-单调容量，如果对于任意集合族 $\{A, A_1, \dots, A_k\} \subseteq \mathcal{F}$满足 $A_i \subseteq A$，对所有 $i \in \{1, \dots, k\}$，有

$$\underline{P}(A) \geq \sum_{\emptyset \neq \mathcal{I} \subseteq \{1,\dots,k\}} (-1)^{\#\mathcal{I}-1} \underline{P}(\cap_{i \in \mathcal{I}} A_i). \tag{3}$$

其共轭 UP $\overline{P}$ 称为 k-交替容量，因为它满足：对于任意集合族 $\{A, A_1, \dots, A_k\} \subseteq \mathcal{F}$，只要 $A \subseteq A_i$，对所有 $i \in \{1, \dots, k\}$，

$$\overline{P}(A) \leq \sum_{\emptyset \neq \mathcal{I} \subseteq \{1,\dots,k\}} (-1)^{\#\mathcal{I}-1} \overline{P}(\cup_{i \in \mathcal{I}} A_i), \tag{4}$$

本文后续将使用的线性规划的一种特殊情形是凸线性规划，即二阶肖凯容量；它们满足 $\underline{P}(A \cup B) \geq \underline{P}(A) + \underline{P}(B) - \underline{P}(A \cap B)$，对所有$A$、$B \in \mathcal{F}$均成立。我们还将使用线性规划的另一种特殊情形：信念函数。

定义17。线性规划 $\underline{P}$ 称为信念函数，如果它是一个 ∞阶肖凯容量，即（3）对每个 k都成立。

信念函数独特之处在于，它可以直观地解释为一个随机集对象，并以 Ω 的子集形式实现自身。

定义18。如果 $\underline{P}$ 是一个信念函数，其对应的质量函数 是非负集函数 m： $\mathcal{F} \to [0, 1]$,

$$A \mapsto m(A) := \sum_{B \subseteq A} (-1)^{\#(A-B)} \underline{P}(B), \tag{5}$$

其中 $A - B \equiv A \cap B^c$，且 A 的子集 B 也必须属于 $\mathcal{F}$。

质量函数 m 具有以下性质

(a) $m(\emptyset) = 0$; (b) $\sum_{B \subseteq \Omega} m(B) = 1$; (c) $\underline{P}(A) = \sum_{B \subseteq A} m(B)$，且对 $\underline{P}$是唯一的。

公式（5）称为 $\underline{P}$ [40]的莫比乌斯变换。质量函数 m 在 $\mathcal{F}$上诱导出一个精确概率分布，即一个随机集的分布。这些概念将在[19]中进一步研究。另请注意，定义18 仅适用于Ω 为有限集的情形。关于无限集上的信念函数，一般定义及其莫比乌斯刻画见[34]。

给定集合 $E \in \mathcal{F}$，对一组概率进行更新 $\mathcal{P}$，就是将集函数 $\underline{P}$ 替换为条件集函数 $\underline{P}^{\times}(\cdot \mid E)$的某个版本。定义 $\underline{P}^{\times}$ 正是更新规则要解决的问题。回想一下，我们在第1节介绍了广义贝叶斯条件化规则：$= (B, E)$。

现在，我们给出下概率和上概率的另外三种更新规则的形式化定义。广义贝叶斯规则、几何规则和邓普斯特规则是文献中最常用、研究最多的规则，而 Gärdenfors规则则是一种非常通用的更新机制，涵盖了许多其他信念修正规则。关于主体为何选择其中一种而非另一种， [19, 37]。

定义19。设 $\mathcal{P} \subseteq \Delta(\Omega\ \mathcal{F})$ 为闭凸集。那么，根据几何规则得到的条件下确界概率和上确界概率是集合函数 $\underline{P}^G, \overline{P}^G$，且对所有 $A, E \in \mathcal{F}$

$$\begin{aligned} \underline{P}^{(G,E)}(A) \equiv \underline{P}^G(A \mid E) &:= \frac{\underline{P}(A \cap E)}{\underline{P}(E)} \quad \text{and} \\ \overline{P}^{(G,E)}(A) \equiv \overline{P}^G(A \mid E) &= 1 - \underline{P}^{(G,E)}(A^c), \end{aligned} \tag{6}$$

前提是 $\underline{P}(E) > 0$。

因此，广义贝叶斯更新程序与几何更新程序的主要区别在于，前者考虑 $P(A \cap E)$ 与 $P(E)$ 之比的下确界，而后者考虑下确界之比。接下来我们介绍邓普斯特更新规则。

定义 20。称 $E \in \mathcal{F}$ 为收集到的证据。假设 $\underline{P}$ 是一个信念函数，其质量函数为 m，且满足 $\underline{P}(E) > 0$。设 $\underline{P}_0$ 为一个独立信念函数，其对应的质量函数 m_0 满足 $m_0(E) = 1$。条件信念函数 $\underline{P}^D(\cdot \mid E)$ 定义为

$$\underline{P}^{(D,E)}(A) \equiv \underline{P}^D(A \mid E) := \underline{P}(A) \oplus \underline{P}_0(E), \quad \forall A \in \mathcal{F},$$

其中组合 运算符 $\oplus$ 表示与 $\underline{P}^D(\cdot \mid E)$ 相关联的质量函数 s

$$m^D(A \mid E) = \frac{\sum_{C \cap E = A} m(C)}{\sum_{C' \cap E \neq \emptyset} m(C')}, \quad \forall A \in \mathcal{F}.$$

因此，邓普斯特更新规则得出如下结论。若 $\mathcal{P} = \mathcal{M}$，则根据 邓普斯特更新规则 得到的线性规划和上确界概率是集合函数 $\underline{P}^D, \overline{P}^D$，使得对所有 $A, E \in \mathcal{F}$

$$\begin{aligned} \overline{P}^{(D,E)}(A) \equiv \overline{P}^D(A \mid E) &:= \frac{\overline{P}(A \cap E)}{\overline{P}(E)} \quad \text{and} \\ \underline{P}^{(D,E)}(A) \equiv \underline{P}^D(A \mid E) &= 1 - \overline{P}^{(D,E)}(A^c), \end{aligned} \tag{7}$$

前提是 $\overline{P}(E) > 0$。

若 $\underline{P}$ 是信念函数，且 $\mathcal{P} = \mathcal{M}$，那么几何规则似乎是邓普斯特规则的自然对偶。但在实际操作中，两者有所不同，正如[19, 第2节2]所指出的。主要区别在于，邓普斯特规则要求 $\underline{P}$ 是信念函数，而几何规则则不作此要求。关于几何规则与邓普斯特规则的详细比较，见 [15,17]。此外，文献2单调/2交替容量上的邓普斯特条件化规则公理化扩展见[33]。

最后，我们引入如下内容。

定义21。将 $E \in \mathcal{F}$ 称为收集到的证据。设 $\underline{P}$ 是一个信念函数，其质量函数为 m。考虑函数 $\mathfrak{f}\colon \mathcal{F} \times \mathcal{F} \to [0,1]$，它满足约束条件

(a) 对所有 $X \in \mathcal{F}$，$\sum_{B \in \mathcal{F}} \mathfrak{f}(B, X) = 1$；

(b) $B \subseteq E^c \implies \mathfrak{f}(B, X) = 0$；(c) 对所有 $X \in \mathcal{F}$，$\mathfrak{f}(\emptyset, X) = 0$。

那么，根据*Gärdenfors*广义成像更新规则 （GGI）得到的信念函数 $\underline{P}_{(I,E)}(\cdot) \equiv \underline{P}^I(\cdot \mid E)$，其对应的质量函数为

$$m^I(A \mid E) = \sum_{X \in \mathcal{F}} \mathfrak{f}(A, X) m(X), \quad \forall A \in \mathcal{F}. \tag{8}$$

由于 E 是收集到的信息，我们有 $E \sqcup E^c = \Omega$，其中 $\sqcup$ 表示不交并。方程（8）告诉我们，当得知收集到的证据不在 E^c中时， m 赋予 E^c 中各（子）事件的概率，会根据函数 $\mathfrak{f}$转移到 E 中“最接近”的（子）事件。约束（a）用于确保转移的“概率份额”随后经过归一化。约束（b）对应于闭世界假设，可以表述为：“如果收集到的证据不在 E^c 中，那么它必定属于 E。”Smets [37, 第 C 节6] 舍弃了要求（b），因为他采用开放世界假设；该假设否定了前述陈述，以表示主体可能错误地指定了其所使用的状态空间。因此，如果收集到的证据不在 E^c中，它可能属于 E，也可能属于主体在实验开始时未考虑的 E 的某个超集。约束（c）只是一项合理性检查。函数 $\mathfrak{f}$ 是条件概率的一种形式；它最初由 [20]引入，随后由 [37]推广。选择函数 $\mathfrak{f}$ 说明，收集到证据 E会如何影响我们围绕 A的信念变化。GGI 涵盖了许多其他现有的更新规则[37]。

在本文余下部分中，我们用 $\beth_\times(E)$, $\times \in \{B, G, D, I\}$ 分别表示在给定收集到的证据 E 时的广义贝叶斯、几何、德姆普斯特和加德纳斯更新规则。若给定证据 E 时，规则 $\times$ （严格）收缩 A，则记为 $(\times, E) \not\hookrightarrow A$ ；若规则 $\times$ 对划分 $\mathcal{E}$ 的所有元素都（严格）收缩 A，则记为 $(\times, \mathcal{E}) \not\hookrightarrow A$。

4.2. 有意遗忘。本节将阐述以下主张：如果我们不愿假设 $\underline{P}$ 至少是凸集——更不用说信念函数了——也就是说，如果我们只能使用广义贝叶斯规则和几何规则来更新信念，那么对于表示某个感兴趣实验结果的划分 $\mathcal{E}$ 的所有元素 E，我们就无法实现收缩。在这个非常一般的情形下，我们需要遗忘才能实现收缩。我们仍保留 $\mathcal{P}$ 是闭集且为凸集这一假设。

引理 22。 设 $\mathcal{E}$ 是 Ω 的可测可数分割，并设 $\times \in \{B, G\}$。那么，对于任意 $A \in \mathcal{F}$，我们有

$$\inf_{E \in \mathcal{E}} \underline{P}^\times(A \mid E) \leq \underline{P}(A) \quad \text{and} \quad \sup_{E \in \mathcal{E}} \overline{P}^\times(A \mid E) \geq \overline{P}(A).$$

引理 22 的一个直接推论如下。

定理 23。 设 $\mathcal{E}$ 是 Ω 的可测可数分割。那么对于 $\times \in \{B, G\}$，我们有：对所有 $A \in \mathcal{F}$，存在 $E \in \mathcal{E}$，使得 $\beth_\times(E)$ 不会弱收缩A。

这一结果告诉我们，对于任何感兴趣的事件 $A \in \mathcal{F}$，我们都不可能找到 $\mathcal{E}$ 中能够（哪怕是弱地）收缩 A 的元素。因此，这意味着存在 E_1、$E_2 \in \mathcal{E}$、E_1，它们可能不同于 E_2，使得 $[\underline{P}(A)\ \overline{P}(A)] \subset [\underline{P}^B(A \mid E_1)\ \overline{P}^B(A \mid E_1)]$ 且

$[\underline{P}(A)\ \overline{P}(A)] \subset [\underline{P}^G(A \mid E_2)\ \overline{P}^G(A \mid E_2)]$。显然，如果主体不愿意对 $\mathcal{P}$ 的性质作出任何额外假设，那么就没有机会使划分 $\mathcal{E}$ 中的所有元素都发生收缩。

在 [36, 定理 2.3]中，作者给出了使所有 $E \in \mathcal{E}$发生扩张的充分条件。于是，有意遗忘整个导致 A 扩张的实验，似乎是实现收缩的唯一可行选择。顾名思义，有意遗忘是指主体自愿忘记部分信息，例如这些信息冗余、有害，或主体受命这样做。收集证据 E 之后，如果我们当前的信念由下概率和上概率$\underline{P}^{\times}(\cdot \mid E)$和$\overline{P}^{\times}(\cdot \mid E)$分别表示，那么遗忘就是指逆转学习过程，使“更新后”的下概率和上概率恢复为原先的下“先验概率”和上“先验概率”，即 $\underline{P}(\cdot)$和 $\overline{P}(\cdot)$。

遗忘这一主题在统计学中有所研究。在[32], 中，例如，遗忘是指稳定遗忘；具体来说，假设主体所处的环境可能发生变化。那么，主体对意外事件的反应取决于其对环境发生变化可能性的信念。如果环境具有波动性，那么单个意外事件就会触发对过去信念的遗忘，并促使主体重新学习一种新的应急情境。因此，在时间 t，主体收集证据 E_t；主体利用这些证据推断环境是否发生了变化。如果环境发生了变化，主体就会抹去对过去事件的记忆，并将先验信念重置为最初的先验知识。如果没有发生变化，主体则可以基于先前的信念，学习关于环境结构的新后验信念。有限记忆程序则是另一个例子。在[4, 5], 中，作者研究了基于有限记忆的多臂老虎机问题：由于工作记忆受限，过于久远的数据会被遗忘。

机器学习（ML）领域也在研究遗忘。例如，在 [7], 中，作者提出了一种算法，其中包含一个遗忘因子，用于平衡新数据和过去数据的相对重要性；当检测到概念漂移时，该算法会调整模型，使其更加关注新数据。在这一框架中，遗忘是指随着新证据不断收集，过去数据的重要性逐渐降低。机器学习中另一个遗忘至关重要的例子是持续学习：由于数据会被丢弃且生命周期有限，忘记不重要的信息、保留对未来有用的信息，是持续学习要解决和关注的主要问题 [27]。

最后，也并不令人意外的是，心理学家和认知科学家对遗忘现象进行了深入研究。研究有意遗忘的参考教科书是 [18]。作者考察了在各种情境下，要求人们忘记某些内容的指令对记忆的影响。他们指出，如今可获取的信息数量极其庞大，因此有必要在线遗忘某些信息；而且，用新信息替换已有信息也常常是必不可少的（比如一个人更换电话号码）。有意遗忘的研究源于定向遗忘这一现象：遵循将某些信息视为“需要遗忘”的指令，我们就能更有效地处理大量信息（例如，法庭上呈示的证据若不予采纳，便会被要求忽略）。这样一来，干扰就会减少，我们也能将全部资源用于其余需要记住的

信息。很容易看出，稳定遗忘是有意遗忘的一种特殊情形，机器学习文献中使用的遗忘因子方法也是如此。

考虑这样一个问题：主体通过概率集合 $\mathcal{P} = \mathcal{M}$ 表达其在 $(\Omega, \mathcal{F})$ 上的初始信念。为方便起见，我们记作 $\mathcal{P} \equiv \mathcal{P}_{E_0}$。随着数据逐渐可得，主体会对 $\mathcal{P}$ 中的每个元素使用贝叶斯条件化规则来更新其信念。具体来说，假设我们收集到形式为 $E_1 \subseteq \mathcal{F}$ 的证据；随后，我们更新 $\mathcal{P}$ 中的元素，以得到

$$\begin{aligned}\mathcal{P}_{E_1} := \Bigg\{ & P_{E_1} \in \Delta(\Omega, \mathcal{F}) : P_{E_1}(A) \equiv P(A \mid E_1) \\ & = \frac{P(E_1 \mid A) P(A)}{P(E_1)} \propto P(E_1 \mid A) P(A), \forall A \in \mathcal{F} \Bigg\},\end{aligned}$$

其中，$P \in \mathcal{P}$ 表示先验，$P(E_1 \mid \cdot)$ 表示似然。更一般地，设截至时间 $t > 0$ 收集到的证据由集合族 $\{E_k\}_{k=1}^{t} \subseteq \mathcal{F}$ 表示。那么，主体更新后的意见由集合给出

$$\begin{aligned}\mathcal{P}_{E_1 \cdots E_t} := \Bigg\{ & P_{E_1 \cdots E_t} : P_{E_1 \cdots E_t}(A) \equiv P_{E_1 \cdots E_{t-1}}(A \mid E_t) \\ & \propto P(E_t \mid A, E_1, \ldots, E_{t-1}) P_{E_1 \cdots E_{t-1}}(A), \forall A \in \mathcal{F} \Bigg\},\end{aligned}$$

其中，$\mathcal{P}_{E_1 \cdots E_{t-1}} \ni P_{E_1 \cdots E_{t-1}}(\cdot) \equiv P(\cdot \mid E_1 \cdots E_{t-1})$ 表示“修订后”的先验（即在时间 $t-1$ 计算出的后验），而 $P(E_t \mid \cdot)$ 表示似然。

任取满足 $1 \leq k \leq t$ 的 $k \in \mathbb{N}$。假设 $\mathcal{P}_{E_1 \cdots E_{t-k}}$ 和 $\mathcal{P}_{E_1 \cdots E_{t-k} \cdots E_t}$ 都是凸闭集。固定一个感兴趣的事件 $A \in \mathcal{F}$，并定义 $\mathcal{P}_{\star E_1 \cdots E_{t-k}}(A) := \{P_{E_1 \cdots E_{t-k}} \in \mathcal{P}_{E_1 \cdots E_{t-k}} : P_{E_1 \cdots E_{t-k}}(A) = \underline{P}_{E_1 \cdots E_{t-k}}(A)\}$ 以及 $\mathcal{P}^{\star}{}_{E_1 \cdots E_{t-k}}(A) := \{P_{E_1 \cdots E_{t-k}} \in \mathcal{P}_{E_1 \cdots E_{t-k}} : P_{E_1 \cdots E_{t-k}}(A) = \overline{P}_{E_1 \cdots E_{t-k}}(A)\}$。对于一般的 $P \in \Delta(\Omega, \mathcal{F})$，以及一般的 A、$B \in \mathcal{F}$，定义依赖关系的概念 d_P，并定义由其取值诱导的集合如下：

$$\begin{aligned} d_P(A, B) &:= P(A \cap B) - P(A) P(B), \\ \Sigma^{+}(A, B) &:= \{P \in \Delta(\Omega, \mathcal{F}) : d_P(A, B) > 0\}, \\ \Sigma^{-}(A, B) &:= \{P \in \Delta(\Omega, \mathcal{F}) : d_P(A, B) < 0\}. \end{aligned}$$

现在将 $\urcorner = (\text{IF}_{\times}, \mathscr{E})$ 称为有意遗忘证据 $\mathscr{E}$ 的程序（其中 $\mathscr{E}$ 可以是 $\mathcal{F}$ 的一个元素、状态空间 Ω 的一个完整划分 $\mathcal{E}$，或由划分构成的集合族 $\{\mathcal{E}\}$），此程序在认可规则 $\times$ 更新后执行。接下来给出有意遗忘诱导收缩的充分条件。

定理24。固定一个感兴趣的事件 $A \in \mathcal{F}$，并令主体认可任意规则 $\times \in \{B, G\}$。如果

$$\mathcal{P}_{\star E_1 \cdots E_{t-k}}(A) \cap \Sigma^{-}(A, E_{t-k+1} \cap \cdots \cap E_t) \neq \emptyset$$

and

$$\mathcal{P}^{\star}{}_{E_1 \cdots E_{t-k}}(A) \cap \Sigma^{+}(A, E_{t-k+1} \cap \cdots \cap E_t) \neq \emptyset,$$

那么遗忘 $E_t \quad {}_{-k+1} \cap \cdots \cap E_t$ 严格收缩 A，符号表示为 $(\text{IF}_{\times}, E_{t-k+1} \cap \cdots \cap E_t) \leftarrow\!\!\!\!\!\!\cdot \; A$.

如果这对划分 $\mathcal{E}_s$, $s \in \{t-k+1, \ldots, t\}$ 的所有元素 E_s 都成立，我们记作

$$(\mathrm{IF}_\times, \mathcal{E}_{t-k+1}, \ldots, \mathcal{E}_t) \nleftarrow A,$$

因此，我们可以遗忘时间 $t-k$之后发生的所有实验。若令 $k=t$，便可得到[32]中的稳定遗忘。注意，在稳定遗忘中，定理$E_1 \cdots E_{t-k}$ 中的下标24 被 E_0替换。我们还有以下结论。

推论25。固定一个感兴趣的事件 $A \in \mathcal{F}$，并令主体认可任意规则 $\times \in \{B, G\}$。如果

$$\mathcal{P}_{\star E_k \cdots E_t}(A) \cap \Sigma^-(A, E_1 \cap \cdots \cap E_{k-1}) \neq \emptyset$$

and

$$\mathcal{P}^\star{}_{E_k \cdots E_t}(A) \cap \Sigma^+(A, E_1 \cap \cdots \cap E_{k-1}) \neq \emptyset,$$

那么遗忘 $E_1 \cap \cdots \cap E_{k-1}$ 会严格收缩 A，记作 $(\mathrm{IF}_\times, E_1 \cap \cdots \cap E_{k-1}) \nleftarrow A$。

在这种情况下，我们得到了机器学习版本的遗忘[7]。主体会有意遗忘时间 k之前收集的数据。更形式化地说，作者会为每个时间 E_t收集的证据 t 赋予一个系数，该系数取决于 t，且随着时间0 与当前时间 t 的距离增大而趋近于 T；也就是说，随着0 增至 $|t-T|$ 无穷大，该系数趋近于零。足够久远的证据会被大幅折扣，以至于在实际应用中，我们可以将其视为已被遗忘，因此推论25 适用。

请注意，只要所选的更新规则先导致扩张，有意遗忘始终是实现收缩的可行方式。本节聚焦于 $\times \in \{B, G\}$，因为广义贝叶斯规则和几何规则是我们介绍过的最一般规则（它们不要求所关注的下概率是信念函数）。

4.2.1. *Levi*中立性。 一种特定类型的遗忘源自Levi关于可修正的无误性的研究，参见例如[28]。假设在时间 t，一个主体拥有一个信念体 $\mathcal{K}_t$，涉及 $\mathcal{F}$中的事件；也就是说，这是一个逻辑谓词集合族，用以描述主体在时间 t的信念。[6] 为研究收缩，我们聚焦于集合 $\mathcal{P}_{\mathcal{K}_t}$，它是相对紧且凸的概率测度集合，表示主体在时间 t 由 $\mathcal{K}_t$所诱导的信念。[7] 考虑一个任意事件 $H \subseteq \Omega$； $P_{\mathcal{K}_t}$ 中每个元素 $\mathcal{P}_{\mathcal{K}_t}$都具有以下两个性质：

- 若 $\mathcal{K}_t$ 排除 H，也就是说，若根据时间 t 的信念，事件 H 被认为不可能——记作 $\mathcal{K}_t \mapsto \neg H$ ——那么，对于所有 $A \in \mathcal{F}$， $P_{\mathcal{K}_t}(A \mid H)$ 都没有定义；
- 反之，若 $\mathcal{K}_t$ 没有排除 H，那么对于所有 $A \in \mathcal{F}$， $P_{\mathcal{K}_t}(A \mid H)$ 都有定义，并表示主体关于在 H 发生的情况下 A 之可信度的信念。

集合 $\mathcal{P}_{\mathcal{K}_t}$ 与下概率 $\underline{P}_{\mathcal{K}_t}$相容。假设主体在时间 $t+1$ 收集到证据 E。令 $\mathcal{K}_{t+1}$ 为新的信念体；为简化记法，我们写作$\mathcal{K}_t \cup \{E\}$。接着，假设主体认可广义贝叶斯规则或几何规则 $\times \in \{B, G\}$；主体的更新后信念由 $\underline{P}^\times_{\mathcal{K}_t}(\cdot \mid E)$ 概括。若（根据 $\times$） E 使感兴趣的事件 A' 发生扩张，那么主体可以忽略 E 以实现收缩；他们执行“逆条件化”。我们称主体对E持 Levi中性态度。其不可错信念由 $\mathcal{K}_{t+1}$概括，可能会被修正，并且我们有

[6]不要将其与信念函数混淆。[7]具体而言，我们假设 $\mathcal{P}_{\mathcal{K}_t} = \mathcal{M}_{\mathcal{K}_t} := \{P \in \Delta(\Omega\ \mathcal{F}) :\ P(A) \geq \underline{P}_{\mathcal{K}_t}(A),\ \ \forall A \in \mathcal{F}\}$。

$\mathcal{K}_{t+2} = \mathcal{K}_t$。信念体收缩： $\mathcal{K}_{t+1}$ 失去元素 E，并恢复到时间 t时的状态。我们写作 $(\text{LN}_\times, E) \hookleftarrow \wp A'$，表示在更新并认可规则 $\times$ 之后，对E 持 Levi中性态度会使 A'收缩。若这对划分 $\mathcal{E}$ 的所有元素都成立，其中该划分是表示感兴趣实验可能结果的 Ω 的划分，我们写作 $(\text{LN}_\times, \mathcal{E}) \hookleftarrow \wp A'$。

4.3. **关于 $\underline{P}$ 的性质假设**。如果主体愿意对代表其信念的下概率 $\underline{P}$的类型作出一些假设，那么我们就可以在不诉诸有意遗忘的情况下，对所有 $E \in \mathcal{E}$ 进行收缩。俗话说，天下没有免费的午餐。以下是[19]中的定理 5.9。

定理 26。令 $\mathcal{E} = \{E, E^c\}$ 为与所关注实验结果相关联的划分。假设 $\underline{P}$是一个信念函数，满足 $\underline{P}(E)$ $\underline{P}(E^c) > 0$，并考虑任意事件 $A \in \mathcal{F}$。那么，如果 $\mathcal{E}$ 在几何规则下使 A 扩张，那么它必定在邓普斯特规则下使 A收缩。同样，如果 $\mathcal{E}$ 在邓普斯特规则下使 A 扩张，那么它必定在几何规则下使A 收缩。

定理 26 的证明只要求 $\underline{P}$是凸的，但我们需要假设$\underline{P}$ 实际上是一个信念函数，否则就无法使用邓普斯特规则（见定义 20）。由此可见，邓普斯特规则与几何规则彼此矛盾。

假设 $\underline{P}$ 是信念函数，就能让我们使用原本无法使用的更新规则。

定理 27。设 E 是主体收集到的证据，并假设 $\underline{P}$ 是一个信念函数，其质量函数为 m，且满足 $\underline{P}(E) > 0$。考虑任意事件 $A \in \mathcal{F}$。我们有 $(I, E) \hookleftarrow \wp A$ 当且仅当

$$\sum_{B \subseteq A} \left[\sum_{X \in \mathcal{F}} \mathfrak{f}(B, X) m(X) - m(B) \right] > 0 \quad \text{and}$$

$$\sum_{B \subseteq A^c} \left[\sum_{X \in \mathcal{F}} \mathfrak{f}(B, X) m(X) - m(B) \right] > 0.$$

如果条件 在定理 27 中对划分 $\mathcal{E}$的每个元素都成立，我们记作 $(I, \mathcal{E}) \hookleftarrow \wp A$.

本节的要点是，如果我们愿意对与表示我们信念的集合 $\mathcal{P}$ 相关联的下概率 $\underline{P}$ 的性质作出假设，那么就能通过使用通常不允许对所有 $E \in \mathcal{E}$ 进行收缩的更新规则（例如几何规则）来实现收缩。我们还可以使用全新的更新技术，例如邓普斯特规则或 GGI，而这些技术原本并不适用。

5. 结论

本文表明，在更新主体的观点时，收缩概率集（代表其信念）至少存在三种情形：不依据证据进行信念修正、基于凸池化进行信念修正，以及基于非贝叶斯更新进行信念修正。此外，我们还为每种此类框架提供了程序示例。

这只是深入研究收缩现象的第一步，我们将在不久的将来继续开展这项研究。具体而言，我们计划找出更多可以进行收缩的情形，并寻找连接这些情形的共同纽带。

致谢

我们感谢三位匿名审稿人提供了有益且富有建设性的意见。Michele Caprio 谨此感谢 ARO MURI W911NF2010080 提供的资助。

附录A. Dubins-deFinetti 条件化

本节将展示，如果我们愿意背离经典的概率柯尔莫哥洛夫范式，就能获得更多实现收缩的机会。具体而言，正如我们将看到的，如果允许概率仅为有限可加，贝叶斯法则便可能导致收缩。

假设我们采用 Dubins-deFinetti 条件化（DdFC）框架；关于 DdFC 的深入介绍可见 [10, 14, 35]。另请注意， [9], 作者在 DdFC 框架下比较了信念函数的广义贝叶斯法则、几何规则和邓普斯特规则。就本文而言，该框架与柯尔莫哥洛夫框架的两个主要区别是：概率不必是可数可加的；条件化不是针对 σ-域进行，而是针对事件进行，这些事件是状态空间划分中的成员。以下陈述均成立；我们将为第一条提供示例，第二条也可类似证明。

(1) 如果概率测度是有限可加但非可数可加的，那么收缩可能发生在可数划分的所有元素上；(2) 请回忆，若一个概率 P 在一组 P零测事件的可测并集为 P零测时，就称其为完全可加。如果概率测度是可数可加但非完全可加的，那么收缩可能发生在不可数划分的所有元素上。

还需注意，如果概率测度是完全可加的，那么它们必定是离散的。

定义 28. 如果对每个事件 A，只要对所有 $E \in \mathcal{E}$， $P(A \mid E)$都有定义，并且对任意常数 k_1、k_2，若对所有 $E \in \mathcal{E}$，$k_1 \leq P(A \mid E) \leq k_2$，则 $k_1 \leq P(A) \leq k_2$，我们就称概率测度 P 在划分 $\mathcal{E}$ 中可聚合。

定义 28 指出，对于每个事件 A，如果划分 $\mathcal{E}$ 上的所有条件概率都被两个量 k_1 和 k_2 所界定，那么该事件的无条件概率也同样被这两个量 [25]所界定。德·芬内蒂 [10] 证明了有限可加概率测度（FAPMs）在可数划分上不具可聚合性。

假设我们不再要求 $\mathcal{P}$ 是一组可数可加概率，而允许它是一组有限可加概率测度。那么，由于有限可加概率测度具有非聚合性，通过贝叶斯更新 $\mathcal{P}$ 可能会出现弱收缩和严格收缩。本节给出的两个例子都基于 [14, 页 92], 中的例子，我们在此陈述该例，以说明这两个例子的构造方式。

例29。（Dubins）设 $\Omega = \{A, B\} \times \{N = 1, 2, \ldots\}$。规定 t

- $P(A) = P(B) = 1/2$,
- $P(N = n \mid A) = 2^{-n}$，对 $n \in \{1, 2, \ldots\}$ 而言，是一个可数可加条件概率,
- $P(N = n \mid B) = 0$，对 $n \in \{1, 2, \ldots\}$ 而言，是一个强有限可加条件概率。[8]

那么，$P(N = n) = 2^{-(n+1)} > 0$，对 $n \in \{1, 2, \ldots\}$ 而言，并且（边际上） P在由划分 $\mathcal{E}_N = \{\{N = 1\}, \{N = 2\}, \ldots\}$ 生成的子代数上仅为有限可加。 P展示了划分 $\mathcal{E}_N = \{\{N = 1\}, \{N = 2\}, \ldots\}$ 中事件 A 的非聚合性，因为对 $n \in \{1, 2, \ldots\}$ 而言， $P(A) = 1/2$ 且 $P(A \mid N = n) = 1$。

例示1（弱收缩）。按如下方式使用例子29 。考虑 Ω 上的一个概率集 $\mathcal{P}$ ，满足Ω $= \{A, B\} \times \{N = 1, 2, \ldots\}$，使得 $\mathcal{P} = \{P_\alpha, 0 < \alpha \leq 1\}$，其中

- $P_\alpha(A) = \alpha$,
  对于 $n \in \{1, 2, \ldots\}$, • $P_\alpha(N = n \mid A) = 2^{-n}$ 是一个可列可加条件概率,
  对于 $n \in \{1, 2, \ldots\}$, • $P_\alpha(N = n \mid B) = 0$ 是一个强有限可加条件概率。

请注意， $P_\alpha(N = n \mid A)$ 和 $P_\alpha(N = n \mid B)$ 都不依赖于 α。关于 $\mathcal{P}$，我们有 $0 < P_\alpha(A) \leq 1$。对于每个 $0 < \alpha < 1$，事件 A 在划分 $\mathcal{E}_N$ 中的 P_α 都具有非聚合性，因为对于 $n \in \{1, 2, \ldots\}$， $P_\alpha(A \mid N = n) = 1$。于是可见，对于 $n \in \{1, 2, \ldots\}$，$1 = P_1(A) = P_1(A \mid N = n)$。因此，利用划分 $\mathcal{E}_N$ 中的信息$\{N = n\}$ 进行贝叶斯更新，会弱收缩 $\mathcal{P} = \{P_\alpha, 0 < \alpha \leq 1\}$。

例示2（严格收缩）。将例29 修改如下。考虑一个 $\mathcal{P}$ 上的概率集合Ω $= \{A, B\} \times \{N = 1, 2, \ldots\}$，使得 $\mathcal{P} = \{P_\alpha, 0 < \alpha < 1\}$，其中

- $P_\alpha(A) = \alpha$，因此相对于 $\mathcal{P}$，我们有 $0 < P_\alpha(A) < 1$,
- $P_\alpha(N = n \mid A) = (1 - \alpha)2^{-n}$，对于 $n \in \{1, 2, \ldots\}$,
- $P_\alpha(N = n \mid B) = \alpha 2^{-n}$，对于 $n \in \{1, 2, \ldots\}$。

请注意，这两个条件概率各自都是定义在 N 上、依赖于 α 的有限可加概率分布。此外，注意到 $P(N = n) = 2\alpha(1-\alpha)2^{-n} > 0$，这对于每个 α 而言，也是定义在 N 上的有限可加概率分布。

通过贝叶斯更新，对于每个 $0 < \alpha < 1$和每个 $n \in \{1, 2, \ldots\}$， $P_\alpha(A \mid N = n) = 1/2$。也就是说，对于每个满足 $0 < \alpha < 1$的 $\alpha \neq 1/2$，划分 $\mathcal{E}_N$ 中的事件 A 上， P_α 具有非聚合性。然而， $1/2 = P_{1/2}(A) = P_{1/2}(A \mid N = n)$。因此，利用来自划分 $\mathcal{E}_N$ 的信息 $\{N = n\}$ 进行贝叶斯更新，会严格收缩 $\mathcal{P} = \{P_\alpha, 0 < \alpha < 1\}$。

注30。 这两个例子有助于解释为什么命题4 和 5 仅限于可列可加概率。

附录B. 证明

命题4的证明。设 $\underline{P}$是满足 $\underline{P}(A) = \min_{P \in \mathcal{P}} P(A)$的概率函数。由引理3可知，若 $\underline{P}(\mathrm{X}_{\underline{P}}^{A+}) > 0$, 则 $\underline{P}(\mathrm{X}_{\underline{P}}^{A-}) > 0$，且$(B, E)$不会严格一致收缩 A。也就是说，对每个 $P_{1i}(A) \in \mathcal{P}(A)$， $\underline{P}(A) \leq P_{1i}(A)$，因此当 $\underline{P}(A) = P_{1i}(A)$时，$(B, E)$不会严格一致收缩 A。因此，若(B, E)弱一致收缩 A，则有 $\underline{P}(\mathrm{X}_{\underline{P}}^{A+}) = 0$，进而 $\underline{P}(\{x \in \mathrm{X} : \underline{P}^B(A \mid X = x) = \underline{P}(A)\}) = 1$。现在令

[8]回顾 t 若有限可加概率测度允许由零测集构成的可数划分 [，则称其为强有限可加 1].

$\overline{P}$是满足 $\overline{P}(A) = \max_{P\in\mathcal{P}} P(A)$的概率函数。依同样的推理，$\overline{P}(\{x \in \mathbf{X}: \overline{P}^{B}(A \mid X = x) = \overline{P}(A)\}) = 1$，于是 (B, E) 也不会弱一致收缩 A。□

命题5的证明。假设（采用归谬证明），在一个 $\mathcal{P}$-测度为1的 X 值集合上（即对所有 $P \in \mathcal{P}$, 其 P-概率为1），对每个 x_i，都存在 $P_{1i}(A)$ 和$P_{2i}(A)$ 属于 $\mathcal{P}(A)$，使得对每个 $P(A \mid x_i) \in \mathcal{P}(A \mid x_i)$，要么 $P_{1i}(A) < P(A \mid x_i) \leq P_{2i}(A)$，要么 $P_{1i}(A) \leq P(A \mid x_i) < P_{2i}(A)$。由于 X 是一个简单随机变量，定义 $P_1(A) := \min_i P_{1i}(A)$，以及 $P_2(A) := \max_i P_{2i}(A)$。根据命题4，假设 $\mathcal{P}(A)$ 不是闭集。不失一般性，假设它向下开（若 $\mathcal{P}(A)$ 向上开，推理过程类似）。因此，$\underline{P}(A) < P_1(A)$。于是，存在 $P_0 \in \mathcal{P}$，使得 $\underline{P}(A) < P_0(A) < P_1(A)$。由于对每个 i, $P_{1i}(A) < P_0(A \mid x_i)$ 或 $P_{1i}(A) \leq P_0(A \mid x_i)$，我们还可得，对每个 i，$P_0(A) < P_1(A) \leq P_0(A \mid x_i)$。但这样一来，$P_0(\mathbf{X}_{P_0}^{A+}) = 1$，这与引理3矛盾。□

定理的证明 *9*。 由定义 1 和定理 8可立即得出。 □

定理的证明 *10*。 我们首先证明，$\inf_{P\in\mathrm{Conv}(\mathcal{P})} P(\cdot)$ 凸包的下概率（线性规划） $\mathcal{P}$与$\inf_{P\in\mathrm{ex}[\mathrm{Conv}(\mathcal{P})]} P(\cdot)$ 凸包极值的下概率 $\mathcal{P}$ 相同。为此，任取 $A \in \mathcal{F}$。由于 $\mathrm{ex}[\mathrm{Conv}(\mathcal{P})] \subseteq \mathrm{Conv}(\mathcal{P})$，因此有

$$\inf_{P\in\mathrm{Conv}(\mathcal{P})} P(A) \leq \inf_{P\in\mathrm{ex}[\mathrm{Conv}(\mathcal{P})]} P(A). \tag{9}$$

接着，令 $\emptyset \neq \mathrm{ex}[\mathrm{Conv}(\mathcal{P})] = \{P_j^{ex}\}_{j\in\mathcal{J}}$。对于所有 $P \in \mathrm{Conv}(\mathcal{P})$ 和所有 $A \in \mathcal{F}$，有

$$\begin{aligned} P(A) = \sum_{j\in\mathcal{J}} \alpha_j P_j^{ex}(A) &\geq \sum_{j\in\mathcal{J}} \alpha_j \underline{P}^{ex}(A) \\ &= \underline{P}^{ex}(A) := \inf_{P\in\mathrm{ex}[\mathrm{Conv}(\mathcal{P})]} P(A), \end{aligned}$$

其中 $\{\alpha_j\}_{j\in\mathcal{J}}$ 是一个由正实数组成的集合族，满足 $\sum_{j\in\mathcal{J}} \alpha_j = 1$，这意味着

$$\inf_{P\in\mathrm{Conv}(\mathcal{P})} P(A) \geq \inf_{P\in\mathrm{ex}[\mathrm{Conv}(\mathcal{P})]} P(A). \tag{10}$$

将式（9）与式（10）结合，即得所需的等式。

现在，若 $P^\star \in \mathrm{ex}[\mathrm{Conv}(\mathcal{P})]$,，则可能存在一个集合族 $\{A\} \subseteq \mathcal{F}$，使得 $P^\star(A) = \underline{P}(A)$ 或 $P^\star(A) = P(A)$，因此对于该集合族中的元素，收缩较弱；而 $P^\star(A) > \underline{P}(A)$ 且 $P^\star(A) < P(A)$，对所有 $A \in \mathcal{F} \setminus \{A\}$ 均成立。反之，若 $P^\star = \sum_{j\in\mathcal{J}} \alpha_j P_j^{ex}$, $\alpha_j > 0$对所有 j 均成立，则对所有 $A \in \mathcal{F}$，有 $P^\star(A) \in (\underline{P}(A)\ \overline{P}(A))$，因此对所有 $A \in \mathcal{F}$，有 $\not\supset \not\hookrightarrow A$。□

定理 11 的证明。如果 $\mathcal{P}(A)$ 在欧几里得拓扑中是闭集，且 $P^\star(A) \in \partial_{\mathcal{B}([0,1])}\mathcal{P}(A)$，那么 $P^\star(A) = \underline{P}(A)$ 或 $P^\star(A) = \overline{P}(A)$，因此该收缩是弱的。反之，如果 $P^\star(A) \in \mathrm{int}_{\mathcal{B}([0,1])}\mathcal{P}(A)$，那么 $P^\star(A) \in (\underline{P}(A)\ \overline{P}(A))$，所以我们有 $\not\supset \not\hookrightarrow A$。□

引理 *22*的证明。 此证明借鉴了 [19, 引理 51]。任取 $A \in \mathcal{F}$。由于$\mathcal{P}$ 是闭集，存在 $P_{(A)} \in \mathcal{P}$ 使得 $P_{(A)}(A) = \underline{P}(A)$。注意，下标$(A)$

提醒我们，这个概率测度会随 A 的选择而变化。因此，我们有

$$\begin{aligned}\underline{P}(A) = P_{(A)}(A) &= \sum_{E\in\mathcal{E}} P_{(A)}(A \mid E)P_{(A)}(E)\\ &\geq \sum_{E\in\mathcal{E}} \underline{P}^{\times}(A \mid E)P_{(A)}(E)\\ &\geq \sum_{E\in\mathcal{E}} \inf_{E\in\mathcal{E}} \underline{P}^{\times}(A \mid E)P_{(A)}(E)\\ &= \inf_{E\in\mathcal{E}} \underline{P}^{\times}(A \mid E)\sum_{E\in\mathcal{E}} P_{(A)}(E)\\ &= \inf_{E\in\mathcal{E}} \underline{P}^{\times}(A \mid E).\end{aligned}$$

A的上概率也适用同样的论证，也就是说，如果我们选取 $P'_{(A)} \in \mathcal{P}$，使得 $P'_{(A)}(A) = \overline{P}(A)$，$P'_{(A)}$ 可能不同于 $P_{(A)}$，那么

$$\overline{P}(A) \leq \sum_{E\in\mathcal{E}} \overline{P}^{\times}(A \mid E)P'_{(A)}(E) \leq \sup_{E\in\mathcal{E}} \overline{P}^{\times}(A \mid E).$$

□

定理*23*的证明。 由引理22可直接得出。 □

定理24的证明。此证明源自 [36, 定理2.3]的证明。固定一个感兴趣的 $A \in \mathcal{F}$ 事件，并令 $\times \in \{B, G\}$。任取 $P_{E_1\cdots E_{t-k}} \in \mathcal{P}_{\star E_1\cdots E_{t-k}}(A) \cap \Sigma^{-}(A, E_{t-k+1}\cap\cdots\cap E_t)$。于是，我们有 $P_{E_1\cdots E_{t-k}}(A) = \underline{P}_{E_1\cdots E_{t-k}}(A)$，因为 $P_{E_1\cdots E_{t-k}} \in \mathcal{P}_{\star E_1\cdots E_{t-k}}(A)$，并且$P_{E_1\cdots E_{t-k}}(A\cap E_{t-k+1}\cap\cdots\cap E_t) < P_{E_1\cdots E_{t-k}}(A)P_{E_1\cdots E_{t-k}}(E_{t-k+1}\cap\cdots\cap E_t)$，因为 $P_{E_1\cdots E_{t-k}} \in \Sigma^{-}(A, E_{t-k+1}\cap\cdots\cap E_t)$。于是，

$$\begin{aligned}\underline{P}_{E_1\cdots E_{t-k}}(A) &= P_{E_1\cdots E_{t-k}}(A)\\ &> \frac{P_{E_1\cdots E_{t-k}}(A\cap E_{t-k+1}\cap\cdots\cap E_t)}{P_{E_1\cdots E_{t-k}}(E_{t-k+1}\cap\cdots\cap E_t)}\\ &= P_{E_1\cdots E_{t-k}}(A \mid E_{t-k+1}\cap\cdots\cap E_t)\\ &\geq \underline{P}^{\times}_{E_1\cdots E_{t-k}}(A \mid E_{t-k+1}\cap\cdots\cap E_t).\end{aligned}$$

类似的论证可得 $\overline{P}_{E_1\cdots E_{t-k}}(A) < \overline{P}^{\times}_{E_1\cdots E_{t-k}}(A \mid E_{t-k+1}\cap\cdots\cap E_t)$。因此，无论主体 $\times \in \{B, G\}$ 认可哪种更新规则，$E_{t-k+1}\cap\cdots\cap E_t$ 都会扩张 A。反过来，遗忘 $E_{t-k+1}\cap\cdots\cap E_t$ 会收缩 A，记作 $(\text{IF}^{\times}, E_{t-k+1}\cap\cdots\cap E_t) \nleftrightarrow A$。□

推论 *25*的证明。 与定理 24的证明类似。 □

定理 27 的证明。任取 $A \in \mathcal{F}$。回顾一下，根据 (8)，对所有 $A \in \mathcal{F}$，我们有 $m^I(A \mid E) = \sum_{X\in\mathcal{F}} \mathfrak{f}(A, X)m(X)$。此外，根据与线性规划 $\underline{P}$ 相关联的质量函数 m 的性质 (c)（见定义 5），我们有 $\underline{P}(A) = \sum_{B\subseteq A} m(B)$。因此，

$$\underline{P}^I(A \mid E) = \sum_{B\subseteq A} m^I(B \mid E) = \sum_{B\subseteq A}\sum_{X\in\mathcal{F}} \mathfrak{f}(B, X)m(X).$$

仅当 假设 $(I, E) \leftrightarrow A$。那么，根据收缩的定义，$\underline{P}^I(A \mid E) > \underline{P}(A)$ 且 $\overline{P}^I(A \mid E) < \overline{P}(A)$。当且仅当

$$\sum_{B\subseteq A}\sum_{X\in\mathcal{F}} \mathfrak{f}(B,X)m(X) > \sum_{B\subseteq A} m(B)$$

$$\iff \sum_{B\subseteq A}\left[\sum_{X\in\mathcal{F}} \mathfrak{f}(B,X)m(X) - m(B)\right] > 0$$

and

$$\sum_{B\subseteq A^c}\sum_{X\in\mathcal{F}} \mathfrak{f}(B,X)m(X) > \sum_{B\subseteq A^c} m(B)$$

$$\iff \sum_{B\subseteq A^c}\left[\sum_{X\in\mathcal{F}} \mathfrak{f}(B,X)m(X) - m(B)\right] > 0.$$

后者成立，因为 $\overline{P}^I(A \mid E) < \overline{P}(A) \iff 1 - \underline{P}^I(A^c \mid E) < 1 - \underline{P}(A^c) \iff \underline{P}^I(A^c \mid E) > \underline{P}(A^c)$。**如果** 假设

$$\sum_{B\subseteq A}\left[\sum_{X\in\mathcal{F}} \mathfrak{f}(B,X)m(X) - m(B)\right] > 0 \quad \text{and}$$

$$\sum_{B\subseteq A^c}\left[\sum_{X\in\mathcal{F}} \mathfrak{f}(B,X)m(X) - m(B)\right] > 0.$$

于是，我们有

$$\underline{P}^I(A \mid E) = \sum_{B\subseteq A}\sum_{X\in\mathcal{F}} \mathfrak{f}(B,X)m(X) > \sum_{B\subseteq A} m(B) = \underline{P}(A)$$

and

$$\underline{P}^I(A^c \mid E) = \sum_{B\subseteq A^c}\sum_{X\in\mathcal{F}} \mathfrak{f}(B,X)m(X) > \sum_{B\subseteq A^c} m(B) = \underline{P}(A^c),$$

这意味着 $\overline{P}^I(A \mid E) < \overline{P}(A)$。证明至此完成。 □

Author Contributions

两位作者对本文的贡献相同。

参考文献

[1] Thomas E. Armstrong。有限可加测度的强奇异性、不相交性与强有限可加性。数学分析与应用杂志，131(2):565–587，1988。[2]Thomas Augustin、Frank P. A. Coolen、Gert de Cooman 和 Matthias C. M. Troffaes（编）。不精确概率导论。威利概率与统计丛书。英国奇切斯特：Wiley，2014。[3] Robert J. Aumann。同意彼此意见不合。《统计年鉴》，4(6):1236–1239，1976。

[4] Dorian Baudry、Yoan Russac 和 Olivier Cappé。多臂老虎机的有限记忆子采样策略。载于 Marina Meila 和 Tong Zhang（编），ICML 2021：第38届国际机器学习会议论文集，第727–737页。PMLR，2021。[5] Donald A. Berry 和 Bert Fristedt。多臂老虎机问题——实验的序贯分配。统计学与应用概率专著。多德雷赫特：Springer，1985。[6] Patrick Billingsley。概率与测度。纽约：Wiley，第2版，1995。[7] Weipeng Cao、Zhong Ming、Zhiwu Xu、Jiyong Zhang 和 Qiang Wang。带动态遗忘因子的在线序贯极限学习机。IEEE Access，7:179746–179757，2019。[8] Michele Caprio 和 Ruobin Gong。动态精确与非精确概率运动学。可见于 arXiv:2110.04382，2022。[9] Giulianella Coletti、Davide Petturiti 和 Barbara Vantaggi。有限情形下作为条件概率下包络的条件信念函数。Information Sciences，339:64–84，2016。[10] Bruno de Finetti。概率、归纳与统计。纽约：Wiley，1972。[11] Bruno de Finetti。概率论，第1卷。纽约：Wiley，1974。[12] Morris H. DeGroot。达成共识。美国统计学会会刊，69(345):118–121，1974。[13]Arthur P. Dempster。面向统计学家的Dempster–Shafer演算。International Journal of Approximate Reasoning，48(2):365–377，2008。[14] Lester E. Dubins。有限可加条件概率、可聚合性与分解。《概率年刊》，3(1):89–99，1975。[15] Didier Dubois 和 Henri Prade。基于信念函数、序条件函数和可能性测度的更新。载于 Piero P. Bonissone、Max Henrion、Laveen N. Kanal 和 John F. Lemmer（编），UAI '90：第六届人工智能中的不确定性年度会议论文集，美国马萨诸塞州剑桥麻省理工学院，1990年7月27—29日，第311–330页。Elsevier，1990。[16]Daniel Ellsberg。风险、模糊性与Savage公理。The Quarterly Journal of Economics，75(4):643–669，1961。[17]Itzhak Gilboa 和 David Schmeidler。更新模糊信念。Journal of Economic Theory，59(1):33–49，1993。[18] Jonathan M. Golding 和 Colin M. MacLeod（编）。有意遗忘：跨学科视角。新泽西州马瓦：Lawrence Erlbaum Associates，1998。[19] Ruobin Gong 和 Xiao-Li Meng。审慎判断遭遇令人不安的更新：扩张、必然损失与辛普森悖论。Statistical Science，36(2):169–190，2021。[20] Peter Gärdenfors。流变中的知识：认知状态动态模型。马萨诸塞州波士顿：麻省理工学院出版社，1988。[21] Paul R. Halmos。测度论。数学研究生教材。纽约：Springer，1950。[22] Jan Hannig、Hari Iyer、Randy C. S. Lai 和 Thomas C. M. Lee。广义信赖推断：综述与新结果。美国统计学会会刊，111(515):1346–1361，2016。[23] Sven Ove Hansson。信念修正逻辑。载于 Edward N. Zalta（编），斯坦福哲学百科全书。加利福尼亚州帕洛阿尔托：斯坦福大学形而上学研究实验室，2022年春季版，2022。[24] Timothy Herron、Teddy Seidenfeld 和 Larry Wasserman。分裂式条件化：关于扩张的更多结果。Philosophy of Science，64(3):411–444，1997。[25]Joseph B. Kadane、Mark J. Schervish 和 Teddy Seidenfeld。有限可加概率的统计学含义，第211–232页。剑桥概率、归纳与决策理论研究。英国剑桥：剑桥大学出版社，1999。[26] Hiremagalur K. Kesavan。Jaynes最大熵原理。载于 Christodoulos A. Floudas 和 Panos M. Pardalos（编），优化百科全书。马萨诸塞州波士顿：Springer，2008。[27] Timothée Lesort、Vincenzo Lomonaco、Andrei Stoian、Davide Maltoni、David Filliat 和 Natalia Díaz-Rodríguez。机器人学的持续学习：定义、框架、学习策略、机遇与挑战。Information Fusion，58:52–68，2020。[28] Isaac Levi。为什么不确定概率是理性的。Journal of Applied Logic，7(4):364–376，2009。

[29]Isaac Levi. 如何实现虽不可能出错却可修正的完全信念。载于 Horacio Arló-Costa、Vincent F. Hendricks 和 Johan van Benthem（编），形式认识论读本：资料集，第 247– 267 页。瑞士沙姆：施普林格国际出版公司，2016 年。[30] Massimo Marinacci 和 Luigi Montrucchio。模糊性数学导论。《经济理论中的不确定性》。美国纽约：劳特利奇，2004 年。[31]Ryan Martin 和 Chuanhai Liu。推断模型：不确定性推理，统计学与应用概率专著第 145 卷。美国佛罗里达州博卡拉顿：CRC出版社，2015 年。[32] Vincent Moens 和 Alexandre Zénon。运用强化贝叶斯变化检测进行学习与遗忘。《PLoS计算生物学》，4(15):1–41，2019 年。[33] Davide Petturiti 和 Barbara Vantaggi。条件次模Choquet期望值与条件相干风险度量。《International Journal of Approximate Reasoning》，113:14–38，2019 年。[34] Davide Petturiti 和 Barbara Vantaggi。如何评估相干信念：*Dempster-Shafer*证据理论中不同相干性概念的比较，第 161–185 页。施普林格国际出版公司，沙姆，2022 年。[35] Eugenio Regazzini。德·芬内蒂的相干性与统计推断。《统计年鉴》，15(2):845– 864，1987 年。[36] Teddy Seidenfeld 和 Larry Wasserman。概率集的扩张。《统计年鉴》，21(3):1139–1154，1993 年。[37] Philippe Smets。关于更新。载于第七届人工智能不确定性国际会议论文集，第 378–385 页，1991 年。[38] Rush T. Stewart 和 Ignacio Ojea Quintana。基于非精确概率的概率意见汇聚。《哲学逻辑杂志》，47(1):17–45，2018 年。[39] Peter Walley。非精确概率的统计推理，统计学与应用概率专著第 42 卷。英国伦敦：Chapman and Hall出版社，1991 年。[40] Roland R. Yager 和 Liping Liu（编）。*Dempster-Shafer*信念函数理论经典著作，模糊性与软计算研究第 39 卷。美国纽约：Springer，2008 年。宾夕法尼亚大学 PRECISE中心、计算机与信息科学系

ania,

美国宾夕法尼亚州费城核桃街3330号，邮编19104电子邮箱： caprio@seas.upenn.edu网址：https://precise.seas.upenn.edu/people/doctoral-students卡内基梅隆大学哲学与统计学系，宾夕法尼亚州匹兹堡福布斯大道5000号，邮编15213电子邮箱： teddy@stat.cmu.edu网址：https://www.cmu.edu/dietrich/philosophy/people/faculty/seidenfeld.html