arXiv:2304.04856v2 [math.ST] 5 Apr 2024

作为广义詹森不等式的图凸包界

Ilja Klebanov[1]

2024年4月9日

摘要。 詹森不等式在测度论、概率论、统计学、机器学习、信息论以及数学和数据科学的许多其他领域中无处不在。它指出，对于定义在凸定义域 $K \subseteq \mathbb{R}^d$ 上的任意凸函数 f: $K \to \mathbb{R}$，以及取值于 K 中的任意随机变量 X， $\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$。本文针对任意定义域 K 上的任意函数 f，推导出 $\mathbb{E}[f(X)]$, 的紧上界和下界，称为“图凸包界”，从而极大地推广了詹森不等式。推导这些界需要研究 f 的图的凸包，而对于复杂函数，这可能颇具挑战。另一方面，一旦确立了这些不等式，它们就和詹森不等式一样，适用于任意 取值于 K 的随机变量 X。因此，当 f 相对简单而 X 复杂或未知时，这些界尤其有用。本文还涵盖 f 的有限维和无限维定义域与陪域，以及条件期望和马尔可夫算子的类似界。

关键词。 詹森不等式 • 凸包 • 非凸函数 • 马尔可夫算子 •条件期望 • Hahn–Banach分离定理
2020年数学主题分类。 26D15 • 28B05 • 46A55 • 52A40 • 60B11 •37A30

1 引言

在许多理论和实际推导中，精确计算函数 f: $K \to \mathbb{R}$ 的均值$\mathbb{E}[f(X)]$ 颇具挑战，其中 $K \subseteq \mathbb{R}^d$和 $d \in \mathbb{N}$，作用于取值于 K的随机变量 X。因此，人们必须依赖对 $\mathbb{E}[f(X)]$的界。当 X仅以其均值 $\mathbb{E}[X]$的形式已知时，尤其如此；此时需要仅依赖于对 $\mathbb{E}[X]$和 f的了解，从而引出本文所讨论的一般问题：

问题 1.1。 在给定 $\mathbb{E}[X]$ 和 f的情况下，建立均值 $\mathbb{E}[f(X)]$ 的（紧）界。

在上述有限维设定中，詹森不等式（Jensen，1906）为定义在凸定义域 f 上的凸函数 K提供了部分答案，断言

$$\mathbb{E}[f(X)] \geq f(\mathbb{E}[X]) \tag{1.1}$$

（未指定上界）。尽管詹森不等式是测度论中的基础性结果，并广泛应用于概率论、统计学、信息论、统计物理学及许多其他研究领域，但它仍存在若干重要的局限和限制。本文旨在推导关于 $\mathbb{E}[f(X)]$, 的新界，称为图凸包界，以弥补这些不足，具体而言：

[1]柏林自由大学¨，德国柏林 14195，阿尼马勒巷 6 号（klebanov@zedat.fu-berlin.de）

- 詹森不等式对既非凸也非凹的函数 f 无能为力，而图凸包界适用于任意函数。
- 詹森不等式的一个显著局限在于，它要求函数 f 定义在 凸 定义域 K上。另一方面，我们的图凸包界不受对 f的定义域 例 3.11和图3.1中所示的此类约束限制，甚至允许非连通定义域，从而增强了其通用性。
- 詹森不等式并未提供任何形式为 $\mathbb{E}[f(X)] \leq s + f(\mathbb{E}[X])$ 或 $\mathbb{E}[f(X)] \leq cf(\mathbb{E}[X])$ 的“反向”界，其中某些 $s = s(X) \geq 0$, $c = c(X) \geq 1$ 满足相应条件。这一特性在各种数学和应用场景中可能极具价值；对于特定函数或函数类，当然可以得到此类界（Budimir et al., 2001; Dragomir, 2013; Khan et al., 2020b; Wunder et al., 2021）。满足上述性质的最小值 $s = s(X) \geq 0$ 称为詹森间隙（Abramovich and Persson, 2016; Khan et al., 2020a; Ullah et al., 2021）。图凸包界为$\mathbb{E}[f(X)]$ 提供了上界和下界。然而，与詹森间隙不同，我们的界仅通过 $\mathbb{E}[X]$ 依赖于 X （参见问题 1.1），因此适用范围更广，但只能在较弱意义上达到紧界。
- 对于定义域 K 位于无限维拓扑向量空间内 $\mathcal{X}$，詹森不等式要求 f 满足某些连续性条件（Perlman，1974年），而我们的图凸包界则不需要这一规定。

图凸包界利用了一个基本事实：点对 $(X, f(X))$ 的均值位于函数 f 的图 $\mathcal{G}(f)$ 的凸包 $\mathrm{Conv}(\mathcal{G}(f))$ 的闭包中，参见下文推论 3.3 和图 3.1。通过简单应用 Hahn–Banach 分离定理即可证明这一点——将 $\mathrm{Conv}(\mathcal{G}(f))$ 表示为半空间的交，并注意到期望值 $\mathbb{E}$ 作为一种算子会“保持”每个半空间，因为它是线性的、正的，并且对常数 Z 而言具有 $\mathbb{E}[Z] = Z$ 。尽管具体应用中的图凸包界需要对 $\mathrm{Conv}(\mathcal{G}(f))$ 进行一些研究（注意，詹森不等式也需要分析 f，即验证其凸性），但所有不等式都仅由 f 的性质推导得出，并且与詹森不等式一样，对任意随机变量 X 均成立，与其概率分布无关。这一性质至关重要，因为在实践中，函数 f 通常足够简单，至少可以通过数值方法进行分析；而随机变量 X 的分布可能复杂或未知，或者人们需要对取值于 K 的所有随机变量给出先验的统一界。

请注意，上文推导（$\mathbb{E}$，$\mathbb{E}[X]$，$\mathbb{E}[f(X)] \in$凸包（$\mathcal{G}(f)$）时提到的这三个性质都属于代数性质，与测度论或积分理论关系不大。因此，图凸包界可以扩展到期望值以外的一大类线性算子，即所谓的马尔可夫算子，也可以扩展到条件期望；这些都是詹森不等式中的众所周知的事实，参见例如（Bakry 等人，2014年，公式（1.2.1））和（Dudley，2002年，定理 10.2.7）。一旦证明了图凸包界的这三个版本，相应版本的詹森不等式便可作为直接推论得出，从而为这一著名结果提供一种新颖且简单得多的证明（参见注释 3.7）。通过采用比马尔可夫算子通常所用框架更一般的设定（注释 4.2），我们也大幅推广了相应版本的詹森不等式。

本文结构如下。第2 节介绍一般框架和符号后，我们将分别推导期望值的图凸包界（第 3 节）、马尔可夫算子（第 4 节）和条件期望（第 5 节）的图凸包界。相应的詹森不等式推论分别见各节的推论 3.6、4.11和5.7。

2 预备知识

为尽可能在最一般的框架下表述结果，我们假设 $\mathcal{X}$ 和 $\mathcal{Y}$ 均为实的豪斯多夫局部凸拓扑向量空间，且 $f\colon K \to \mathcal{Y}$，其中 $K \subseteq \mathcal{X}$ 是任意定义域。不过，为简洁起见，建议读者将 $\mathcal{X} = \mathbb{R}^d$ 和 $\mathcal{Y} = \mathbb{R}$ 视为……此外，$(\Omega, \Sigma, \mathbb{P})$ 表示一个概率空间，而 $X\colon \Omega \to \mathcal{X}$ 表示一个函数，使得 $(X, f(X))$ 是弱可积的[1]。因此，所有期望值均按弱意义解释，这意味着 X 不一定需要可测，也就是说，不一定是随机变量。回顾一下，若存在 $e \in \mathcal{X}$，使得对于任意 $\ell \in \mathcal{X}'$，$\ell(X) \in L^1(\mathbb{P})$ 且 $\ell(e) = \int_\Omega \ell(X)\, \mathrm{d}\mathbb{P}$，则称映射 $X\colon \Omega \to \mathcal{X}$ 为弱可积；此时 X 的期望值定义为 $\mathbb{E}[X] := \int_\Omega X \, \mathrm{d}\mathbb{P} := e$ (Rudin，1991，定义 3.26)。这里，$\mathcal{X}'$ 表示 $\mathcal{X}$ 的连续对偶。特别地，每个 Pettis 可积或 Bochner可积映射都是弱可积的；为简洁起见，读者可以考虑 Bochner可积随机变量 X。在第 4 节讨论马尔可夫算子时，我们不会对 $(X, f(X))$ 施加任何可测性或可积性假设；而在第 5 节讨论条件期望时，则要求 $\mathcal{X}$ 和 $\mathcal{Y}$ 为巴拿赫空间，且 $(X, f(X))$ 为 Bochner可积。回顾一下，随机变量 $X\colon (\Omega, \Sigma, \mathbb{P}) \to \mathcal{X}$ 是 Bochner可积（或强可积）的充要条件是 X 可测且 $\int_\Omega \|X\|_{\mathcal{X}} \mathrm{d}\mathbb{P} < \infty$ (Diestel和Uhl，1977，定理 II.2.2)；此时记作 $X \in L^1(\mathbb{P}; \mathcal{X})$。相关假设将在各节分别详细说明，即假设 3.1、4.5 和 5.1。

全文中，我们使用以下记号。我们将实拓扑向量空间的子集 A 的闭包记为 $\overline{A}$，将 A 的凸包记为

$$\operatorname{Conv}(A) := \bigcap \{C \supset A \mid C \text{ convex}\} = \left\{ \sum_{j=1}^{J} w_j a_j \,\middle|\, J \in \mathbb{N},\, a_j \in A,\, w_j \geq 0,\, \sum_{j=1}^{J} w_j = 1 \right\}. \tag{2.1}$$

此外，$\mathcal{G}(f) = \{(x, f(x)) \mid x \in K\}$ 表示函数 $f\colon K \to \mathcal{Y}$ 的函数图像，并且，对于 $M \subseteq \mathcal{X} \times \mathcal{Y}$ 和 $x \in \mathcal{X}$，

$$M|_x := \{y \in \mathcal{Y} \mid (x, y) \in M\}$$

（注意，$\overline{M|_x}$ 和 $\overline{M}|_x$ 可能并不重合）。最后，尽管我们显然考虑的是关于 $\mathcal{Y}$ 上的某个全序、偏序或预序的不等式，但出于可读性考虑，大多数陈述都是针对 $\mathcal{Y}$ 上的任意二元关系提出的，我们仍将其记为 $\leq$。不过，在 $\mathcal{Y} = \mathbb{R}$ 的情况下，我们始终假设采用典范全序。在某些情况下，我们将通过添加元素 $\pm\infty$ 来扩展空间 $\mathcal{Y}$，这些元素被假定对每个 $y \in \mathcal{Y}$ 都满足 $-\infty \leq y \leq \infty$，并记为 $\overline{\mathcal{Y}} := \mathcal{Y} \cup \{\pm\infty\}$，这与扩展实数轴 $\overline{\mathbb{R}} = [-\infty, \infty] = \mathbb{R} \cup \{\pm\infty\}$ 的通常做法一致。$\mathcal{Y}$ 中的闭区间定义为 $[a, b] := \{y \in \mathcal{Y} \mid a \leq y \leq b\}$，$a, b \in \overline{\mathcal{Y}}$. 现在，我们陈述一个简单的初步引理，之后将利用它从图凸包界推出詹森不等式：

引理 *2.1.* 设 $f\colon K \to \mathbb{R}$ 是定义在凸定义域 $K \subseteq \mathbb{R}^d$ 上的凸函数。那么其上图集 $\mathit{epi}(f) := \{(x, y) \in K \times \mathbb{R} \mid y \geq f(x)\} \supseteq \mathcal{G}(f)$ 是凸的，因此，对于每个 $x \in K$，$\mathit{Conv}(\mathcal{G}(f))|_x \subseteq [f(x)\ \infty)$。如果 $K = \mathbb{R}^d$，那么对于每个 $x \in K$，$\mathit{Conv}(\mathcal{G}(f))|_x \subseteq [f(x)\ \infty)$。

证明。epi(f) 的凸性是一个众所周知的结果（Rockafellar，1997，定理 4.1）。由 (2.1) 可得 Conv($\mathcal{G}(f)$) $\subseteq$ epi(f)，且对每个 $x \in K$，Conv($\mathcal{G}(f)$)$|_x \subseteq$ epi(f)$|_x = [f(x)\ \infty)$。若 $K = \mathbb{R}^d$，则 f 连续（Borwein 和 Lewis，2006，定理 4.1.3），且 epi(f) 是闭集。因此，对每个 $x \in K$，Conv($\mathcal{G}(f)$)$|_x \subseteq$ epi(f)$|_x =$ epi(f)$|_x = [f(x)\ \infty)$。■

[1] 与随机变量 X 的通常写法一致，我们用 $f(X)$ 表示复合 $f \circ X$；并将此记号扩展到不可测的映射 X。

3 期望值的图凸包界

在本节中，我们作出以下一般假设：

假设 3.1。 $(\Omega\ \Sigma,\ \mathbb{P})$ 是一个概率空间， $\mathcal{X}$， $\mathcal{Y}$， $\mathcal{Z}$ 是实数、豪斯多夫且局部凸拓扑向量空间（配备其博雷尔 σ-代数）。此外， $\leq$ 表示 $\mathcal{Y}$ 上的任意二元关系，并规定在 $\mathcal{Y}=\mathbb{R}$ 的情况下默认为典范全序（因此采用该记号）。最后， $K\subseteq\mathcal{X}$, $f\colon K\to\mathcal{Y}$ 以及 $X\colon\Omega\to K$，使得二元组 $(X,\ f(X))$ 是弱可积的。

以下定理给出了一个众所周知且直观的结果，即随机变量的均值 $\mathbb{E}[Z]$ Z 取值于一个闭凸子集 C $\mathcal{Z}$ 时，位于该集合内[2]： $\mathbb{E}[Z]\in C$。如果 $\mathcal{Z}=\mathbb{R}^d$，则不需要闭性假设。其证明利用了 Hahn–Banach 分离定理。

定理 *3.2*（凸集的均值/重心位于其闭包内）。设假设 *3.1* 成立，并设 $Z\colon\Omega\to A$ 是一个取值于子集 $A\subseteq\mathcal{Z}$ 的弱可积映射。那么 $\mathbb{E}[Z]\in$ *Conv(A)* 如果 $\mathcal{Z}=\mathbb{R}^d$、 $d\in\mathbb{N}$，并且一般而言 $\mathbb{E}[Z]\in$ *Conv(A)*[3]。此外，对于任意 $z\in$ *Conv(A)*，存在一个概率空间 $(\Omega\ \Sigma、\ \mathbb{P})$ 以及一个弱可积映射 $Z\colon\Omega\to A$，使得 $\mathbb{E}[Z]=z$。

证明。关于一般性表述 $\mathbb{E}[Z]\in$ Conv(A)，请参见（Perlman，1974年，定理 3.1）；关于在 $\mathcal{Z}=\mathbb{R}^d$ 情况下 $\mathbb{E}[Z]\in$ Conv(A) 的证明，请参见（Dudley，2002年，定理 10.2.6.）（尽管我们的设置稍微更一般，证明方法类似）。现在令 $z\in$ Conv(A)。那么它可以写成 A 中元素的凸组合，即对于某些 $J\in\mathbb{N}$,$a_j\in A$, $w_j\geq 0$, $j=1,\ldots,J$，有 $\sum_{j=1}^{J}w_j=1$，可表示为 $z=\sum_{j=1}^{J}w_ja_j$。取 $\tilde{\Omega}=\{1,\ldots,J\}$，其分布由概率向量 $(w_1,\ldots,w_J)$ 给出，并令 $\tilde{Z}(j)=a_j$。那么 $\mathbb{E}[\tilde{Z}]=\sum_{j=1}^{J}w_ja_j=z$。

■

将上述定理应用于函数 f 的凸包可得：

推论*3.3.* 在假设 *3.1* 下， $\mathbb{E}[(X,f(X))]\in$ *Conv(G(f))* 若 $\mathcal{X}=\mathbb{R}^m$, $\mathcal{Y}=\mathbb{R}^n$, $m,n\in\mathbb{N}$，且 $\mathbb{E}[(X,f(X))]\in$ *Conv(G(f))* 一般成立。此外，对于任意 $(x,y)\in$ *Conv(G(f))*，存在一个概率空间 $(\Omega\ \Sigma,\mathbb{P})$ 以及 $X\colon\Omega\to K$，使得二元组 $(X,f(X))$ 是弱可积的，并且 $\mathbb{E}[(X,f(X))]=(x,y)$。

证明。这由 定理 3.2 对 $Z\colon=(X,\ f(X))$ 以及 $A\colon=\mathcal{G}(f)\subseteq\mathcal{X}\times\mathcal{Y}=:\mathcal{Z}$（注意任何取值于 Z 的映射 $\mathcal{G}(f)$ 都具有 $(X,\ f(X))$ 的形式。■

以下结果基本上是推论 3.3 的重述。由于这是本节的主要结果，我们将其表述为一个定理：

定理 3.4 (图凸包界 $\mathbb{E}[f(X)]$)。 在 假设 *3.1* 下，令图 f 的凸包由两个“包络”函数 g_l, g_u: $\mathrm{Conv}(K)\to\mathcal{Y}$ 以如下方式“围住”：

$$\begin{aligned}&\overline{\mathrm{Conv}(\mathcal{G}(f))}\subseteq\{(x,y)\in\overline{\mathrm{Conv}(K)}\times\mathcal{Y}\mid g_l(x)\leq y\leq g_u(x)\},\\ i.e.\qquad&\overline{\mathrm{Conv}(\mathcal{G}(f))}|_x\subseteq[g_l(x),g_u(x)]\qquad for\ each\ x\in\overline{\mathrm{Conv}(K)}.\end{aligned}\tag{3.1}$$

[2]请注意，这与 Choquet 理论（尤其是 Krein–Milman 定理）所探讨的问题截然不同，相关讨论见（Phelps，2001），其中分析了有界闭凸集与其 *extreme points* 的凸包之间的关系。极点与本文无关。[3]关于一个取值于无限维空间的凸子集 Z 中的随机变量 C，且满足 $\mathbb{E}[Z]\in C\setminus C$，其示例见（Perlman，1974年，注 3.2）。

Then

$$g_l(\mathbb{E}[X]) \le \mathbb{E}[f(X)] \le g_u(\mathbb{E}[X]). \tag{3.2}$$

若 $\mathcal{X} = \mathbb{R}^m$，$\mathcal{Y} = \mathbb{R}^n$，$m$，$n \in \mathbb{N}$，则 $\overline{\mathrm{Conv}(K)}$ 和 $\mathrm{Conv}\overline{(\mathcal{G}(f))}$ 可分别替换为 $\mathrm{Conv}(K)$ 和 $\mathrm{Conv}(\mathcal{G}(f))$。此外，对于 $\mathcal{Y} = \mathbb{R}$，(3.2) 在以下意义下是尖锐的：

- 如果 $\mathcal{X} = \mathbb{R}^m$ 且 $g_l(x) = \inf(\mathrm{Conv}(\mathcal{G}(f))|_x)$ 且 $g_u(x) = \sup(\mathrm{Conv}(\mathcal{G}(f))|_x)$ 对每个 $x \in \mathrm{Conv}(K)$ 都成立，那么，对于每个 $x \in \mathrm{Conv}(K)$ 和 $y \in (g_l(x)\ g_u(x))$，存在一个概率空间 $(\Omega, \Sigma, \mathbb{P})$ 以及 $X : \Omega \to K$ 使得该对 $(X, f(X))$ 是弱可积的，且 $(\mathbb{E}[X], \mathbb{E}[f(X)]) = (x, y)$。
- 一般而言，如果对于每个 $x \in \mathrm{Conv}(K)$，都有 $g_l(x) = \inf(\mathrm{Conv}(\mathcal{G}(f))|_x)$ 且 $g_u(x) = \sup(\mathrm{Conv}(\mathcal{G}(f))|_x)$，那么对于每个 $x \in \mathrm{Conv}(K)$ 和 $y \in (g_l(x)\ g_u(x))$，以及 (x, y) 的任意邻域 U，都存在一个概率空间 $(\Omega, \Sigma, \mathbb{P})$ 和 $X : \Omega \to K$，使得对偶 $(X, f(X))$ 弱可积，且 $(\mathbb{E}[X], \mathbb{E}[f(X)]) \in U$。

证明。 由于 $\mathbb{E}[X] \in \overline{\mathrm{Conv}(K)}$ 且 $(\mathbb{E}[X],\ \mathbb{E}[f(X)]) \in \overline{\mathrm{Conv}(\mathcal{G}(f))}$，由定理 3.2 和推论 3.3，

$$\mathbb{E}[f(X)] \in \overline{\mathrm{Conv}(\mathcal{G}(f))}|_{\mathbb{E}[X]} \subseteq [g_l(\mathbb{E}[X]), g_u(\mathbb{E}[X])],$$

参见 图 3.1。在有限维情形下 $\mathcal{X} = \mathbb{R}^m$，$\mathcal{Y} = \mathbb{R}^n$，根据定理 3.2 和推论 3.3，我们可以省去所有闭包。界的锐性主张也可直接由推论 3.3得出。■

注 3.5。 我们再次强调，与詹森不等式一样，界 g_l 和 g_u 完全由函数 f 决定，并且对满足最小可积性假设的任意 K值映射 X 都成立。

直接应用定理 3.4 中的图凸包界，即可得到詹森不等式：

推论 3.6（詹森不等式）。 设假设 *3.1* 成立，且 $f\colon K \to \mathbb{R}$ 是定义在凸定义域 $K \subseteq \mathbb{R}^d$上的凸函数。则 $\mathbb{E}[f(X)] \ge f(\mathbb{E}[X])$。

证明。 这直接由 引理 2.1 和 定理 3.4 结合 $g_l := f$得出。 ■

备注 3.7。 值得注意的是，这种证明詹森不等式的新颖策略比传统方法简单得多。例如，在证明（Dudley， 2002年，定理 10.2.6）时，作者概述了：

- 首先论证 $\mathbb{E}[X] \in K$，类似于 定理 3.2，以确保 $f(\mathbb{E}[X])$ 定义良好，
- 其次，进行一个类似于 引理 2.1的简单论证，并且
- 第三，进行更复杂的论证，以证明不等式本身。

然而，通过第二次应用第一个论证，这次针对一对（X， $f(X)$），即可完全绕过后一论证，得到（$\mathbb{E}[X]$， $\mathbb{E}[f(X)]$） $\in \mathrm{Conv}(\mathcal{G}(f))$，从而直接证明詹森不等式。这种优雅的方法似乎此前一直未被注意到，其在非凸函数和定义域上的应用也同样如此，正如定理 3.4所示。

请注意，在我们的设定中，詹森不等式（ $f(\mathbb{E}[X])$ 中关于 $\mathbb{E}[f(X)]$ 的下界是多余的，甚至可能没有良好定义（因为 $\mathbb{E}[X]$ 不保证属于 K，参见例 3.11）。然而，如果比较 $\mathbb{E}[f(X)]$ 和 $f(\mathbb{E}[X])$ 至关重要，那么只要 $f(\mathbb{E}[X])$ 存在，所述结果便会导出以下推论：

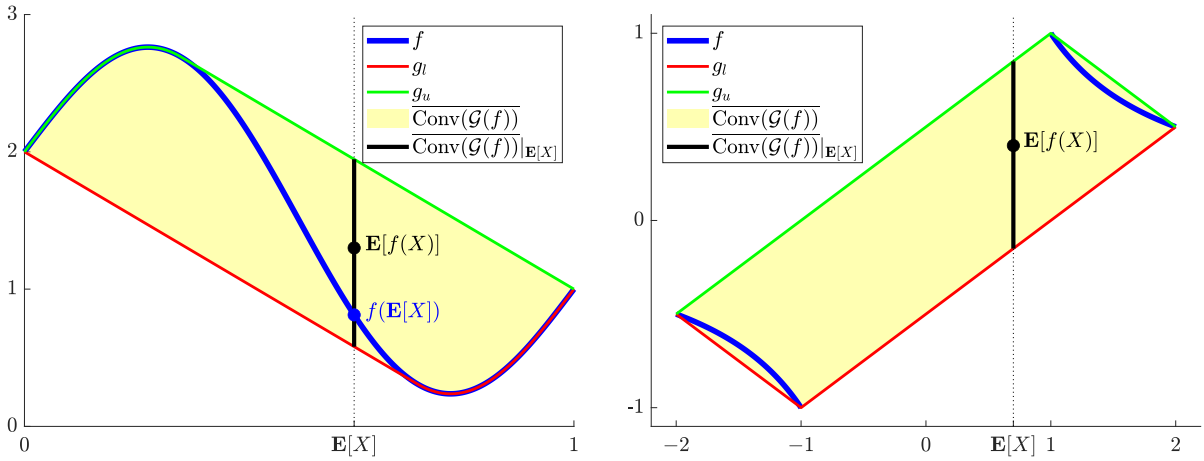

图 3.1：例 3.10 和 3.11 的示意图。由于点对 $(X, f(X))$ 的均值 $\mathbb{E}[(X, f(X))] = (\mathbb{E}[X], \mathbb{E}[f(X)])$（黑点）位于函数 f 的图的凸包的闭包 $\overline{\mathrm{Conv}}(\mathcal{G}(f))$ 中，因此 $\mathbb{E}[f(X)]$ 的取值受限于粗黑线 $\overline{\mathrm{Conv}}^{\mathbb{E}[X]}$，从而给出定理 3.4 中的界。请注意，在第二个例子中，f 的定义域 $K = [-2,-1] \cup [1,2]$ 不连通，并且当 $\mathbb{E}[X] \in (-1, 1)$ 时，$f(\mathbb{E}[X])$ 甚至没有定义，但这并不影响定理 3.4。

推论 3.8. 设 定理 *3.4* 中的假设成立，且令 $\mathcal{Y} = \mathbb{R}$。那么

(a) $c_l(\mathbb{E}[X])\ f(\mathbb{E}[X]) \le \mathbb{E}[f(X)] \le c_u(\mathbb{E}[X])\ f(\mathbb{E}[X])$,

(b) $(\text{下确界}_{x\in K} f(x) \quad f(\mathbb{E}[X]) - s_u) \le \mathbb{E}[f(X)] \le (\text{上确界}_{x\in K} f(x) \quad f(\mathbb{E}[X]) + s_l)$,

(c) $\hat{c}_l\, f(\mathbb{E}[X]) \le \mathbb{E}[f(X)] \le \hat{c}_u\, f(\mathbb{E}[X])$,

只要这些量

$$c_l(x) := \tfrac{g_l(x)}{f(x)}, \qquad s_l := \|f - g_l\|_\infty, \qquad \hat{c}_l := \inf_{x\in \overline{\mathrm{Conv}(K)}} \tfrac{g_l(x)}{f(x)},$$
$$c_u(x) := \tfrac{g_u(x)}{f(x)}, \qquad s_u := \|f - g_u\|_\infty, \qquad \hat{c}_u := \sup_{x\in \overline{\mathrm{Conv}(K)}} \tfrac{g_u(x)}{f(x)},$$

以及 $f(\mathbb{E}[X])$ 都有定义。如果 $\mathcal{X} = \mathbb{R}^m$、$\mathcal{Y} = \mathbb{R}^n$、$m$、$n \in \mathbb{N}$、$\overline{\mathrm{Conv}(K)}$，在定义 $\hat{c}_l$ 和 $\hat{c}_u$ 时，可以替换为 $\mathrm{Conv}(K)$。

证明。 (a) 和 (b) 直接由定理 3.4得出，而 (c)由(a)得出。 ■

注 3.9. 虽然定理 3.4 和推论 3.8 (a) 中的界需要了解 $\mathbb{E}[X]$（参见问题 1.1），但 (b) 和 (c) 比较 $\mathbb{E}[f(X)]$ 与 $f(\mathbb{E}[X])$，而不依赖这些信息。特别是，(b) 和 (c) 给出了先验界，适用于所有取值于 K 的随机变量 X。因此，从 (a) 过渡到 (c) 时，下界 $c_l(\mathbb{E}[X])$ 必须用其最保守的估计替代，即对 x 的所有可能取值，取 $c_l(x)$ 的下确界（对于上界则取上确界）。然而，这些界仍比显然的不等式更精确（只要它们有定义）

$$\inf_{x\in K} f(x) \le \mathbb{E}[f(X)] \le \sup_{x\in K} f(x), \qquad f(\mathbb{E}[X])\ \inf_{x,y\in \overline{K}} \tfrac{f(x)}{f(y)} \le \mathbb{E}[f(X)] \le f(\mathbb{E}[X])\ \sup_{x,y\in \overline{K}} \tfrac{f(x)}{f(y)}.$$

例 3.10。 设 $K=[0,1]$ 和 $f\colon\ K\to\mathbb{R}$, $f(x)=2-x+\sin(2\pi x)$，如 图 3.1（左）所示，它既非凸也非凹。则 Conv($\mathcal{G}(f)$ 的“包络”函数为

$$g_l(x)=\begin{cases}2-ax & \text{if } x\le x_*,\\ f(x) & \text{otherwise,}\end{cases}\qquad g_u(x)=\begin{cases}f(x) & \text{if } x\le 1-x_*,\\ 1+a-ax & \text{otherwise,}\end{cases}$$

其中， $x_*\approx 0.715$ 是方程 $2\pi\cos(2\pi x)\,x=\sin(2\pi x)$ 和 $a=1-2\pi\cos(2\pi x_*)$ 的解。注意，推论 3.8 中的常数 $\hat{}c_l\approx 0.5$ 和 $\hat{}c_u\approx 6.5$，与注 3.5 中显而易见的界 $\inf_{x,y\in K}{}^{f}(x)\,f(y)\approx 0.09$ 和 $\sup_{x,y\in K}{}^{f}(x)\,f(y)\approx 11.6$. 相比，效果更好（很容易看出如何修改函数 f，使其相较于这些界的改进幅度任意增大）。

例 3.11。 令 $K=[-2,-1]\cup[1,2]$ 为函数 f 的非连通 $K\to\mathbb{R}$, $f(x)=x^{-1}$ 定义域，见 图 3.1 （右图）；该函数同样既非凸也非凹。同样，来自Conv($\mathcal{G}(f)$) 的“包络”函数由定理 3.4给出，即分段线性函数

$$g_l(x)=\begin{cases}(-3-x)/2 & \text{if } x\in[-2,-1],\\ (x-1)/2 & \text{if } x\in[-1,2],\end{cases}\qquad g_u(x)=\begin{cases}(x+1)/2 & \text{if } x\in[-2,1],\\ (3-x)/2 & \text{if } x\in[1,2],\end{cases}$$

为 $\mathbb{E}[f(X)]$ 提供紧界。在这种情况下，可能无法比较 $\mathbb{E}[f(X)]$ 与 $f(\mathbb{E}[X])$，因为当 $\mathbb{E}[X]\in(-1,1)$ 时，后者没有定义。

注3.12。 虽然在上述示例中，包络函数 g_l 和 g_u 是通过解析方法确定的（除了对 x_*的近似之外），但即使对于复杂函数 f，通常也可以高精度且以很小的计算量进行数值计算，只要其定义域 K 是有界的。

4 马尔可夫算子的图凸包界

请注意，用于在定理 3.2 中证明 $\mathbb{E}[Z]\in$ Conv(A) 的期望值 $\mathbb{E}$ 的唯一性质是其线性性、对于每个 $\ell\in\mathcal{Z}'$ 而言 $\ell(Z)$ 的正性（意味着 $\ell(Z)\ge 0$ implies$\mathbb{E}[\ell(Z)]\ge 0$），以及对于每个常值随机变量 $Z\in\mathcal{Z}$ 都满足 $\mathbb{E}[Z]=Z$ 的性质（詹森不等式同样如此）。因此，詹森不等式和图凸包界可以被解释为代数结果，而不是测度论的结果，并且可以推广到不一定涉及积分、而仅满足上述三个性质的算子。这一推广正是本节的目的，也将使我们能够在后续章节中讨论条件期望。这类算子称为马尔可夫算子，在时间演化现象和动力系统的分析中发挥着关键作用（Bakry 等，2014）。众所周知，它们满足詹森不等式（Bakry 等，2014，公式 (1.2.1)），但我们显著拓宽了其适用的背景，参见备注 4.2。

定义 4.1。 设 Ω_1,Ω_2 为集合， $\mathcal{S}_1,\mathcal{S}_2$ 为 Ω_1 和 Ω_2 上的函数向量空间，且包含单位常值函数 $\mathbb{1}_{\Omega_1}\colon \Omega_1\to\mathbb{R}, t\mapsto 1$ 和 $\mathbb{1}_{\Omega_2}$，例如 $\mathcal{S}_1=L^1(\mathbb{P}_1)$，如果 $(\Omega_1,\Sigma_1,\mathbb{P}_1)$ 是一个概率空间。线性算子 $\mathcal{E}\colon \mathcal{S}_1\to\mathcal{S}_2$ 称为马尔可夫算子，如果

(i) $\mathcal{E}(\mathbb{1}_{\Omega_1})=\mathbb{1}_{\Omega_2}$, (ii) $\mathcal{E}(\varphi)\ge 0$ 对于每个满足 $\varphi\ge 0$ 的 $\varphi\in\mathcal{S}_1$。

此外，设 $\mathcal{Z}$ 是一个实拓扑向量空间，并且

$$\mathcal{S}_1^{\mathcal{Z}} \subseteq \{Z_1 : \Omega_1 \to \mathcal{Z} \mid \ell \in \mathcal{Z}' : \ell(Z_1) \in \mathcal{S}_1\}, \quad \mathcal{S}_2^{\mathcal{Z}} \subseteq \{Z_2 : \Omega_2 \to \mathcal{Z} \mid \ell \in \mathcal{Z}' : \ell(Z_2) \in \mathcal{S}_2\},$$

例如 $\mathcal{S}_1^{\mathcal{Z}} = L^1(\mathbb{P}_1; \mathcal{Z})$在 $\mathcal{S}_1 = L^1(\mathbb{P}_1)$ 的情况下。我们称 $\mathcal{E}^{\mathcal{Z}}: \mathcal{S}_1^{\mathcal{Z}} \to \mathcal{S}_2^{\mathcal{Z}}$为一个 $\mathcal{Z}$-值的马尔可夫算子，它与马尔可夫算子 $\mathcal{E}$ 相关联，若对每个 $\ell \in \mathcal{Z}'$ 和 $Z_1 \in \mathcal{S}_1^{\mathcal{Z}}$ 都有 $\ell(\mathcal{E}^{\mathcal{Z}}(Z_1)) = \mathcal{E}(\ell(Z_1))$，即对每个 $\ell \in \mathcal{Z}'$，以下图表都可交换。

$$\begin{array}{ccc} \mathcal{S}_1^{\mathcal{Z}} & \xrightarrow{\mathcal{E}^{\mathcal{Z}}} & \mathcal{S}_2^{\mathcal{Z}} \\ \downarrow{\scriptstyle \ell\circ} & & \downarrow{\scriptstyle \ell\circ} \\ \mathcal{S}_1 & \xrightarrow{\mathcal{E}} & \mathcal{S}_2 \end{array}$$

注 4.2。通常假设 Ω_1,Ω_2 是可测空间或测度空间，并且 $\mathcal{S}_1$、$\mathcal{S}_2$ 中的元素是可测的，或相对于某些测度属于 L^1（Bakry 等，2014；Eisner 等，2015）。此外，一些作者要求马尔可夫算子是自同态，即 $\mathcal{S}_1 = \mathcal{S}_2$（Bakry 等，2014）；是 L^1-压缩（Lin，1979；Rudnicki，2000）；或保持 L^1-范数（Eisner 等，2015）。通过采用定义 4.1 中的最低要求，并引入全新的 $\mathcal{Z}$ 值马尔可夫算子这一概念，我们由此显著拓展了目前已知的马尔可夫算子詹森不等式版本的适用范围（Bakry 等，2014，公式（1.2.1））。

例 4.3（参见 第 3 节。设 $(\Omega_1, \Sigma_1, \mathbb{P}_1) = (\Omega_2, \Sigma_2, \mathbb{P}_2)$ 是一个概率空间，且 $\mathcal{S}_1 = L^1(\mathbb{P}_1)$ 和 $\mathcal{S}_2 = \operatorname{span}(\mathbb{1}_{\Omega_2})$。此外，设 $\mathcal{Z}$ 是一个实的、豪斯多夫的局部凸拓扑向量空间，且 $\mathcal{S}_1^{\mathcal{Z}} = \{Z_1 : \Omega_1 \to \mathcal{Z} \mid Z_1$ 弱可积$\}$, $\mathcal{S}_2^{\mathcal{Z}} = \{Z_2 : \Omega_2 \to \mathcal{Z} \mid Z_2$ 常值$\}$。那么，期望值 $\mathbb{E}: \mathcal{S}_1^{\mathcal{Z}} \to \mathcal{S}_2^{\mathcal{Z}}$是一个取值于 $\mathcal{Z}$ 的马尔可夫算子，与马尔可夫算子 $\mathbb{E}: \mathcal{S}_1 \to \mathcal{S}_2$相关联，其中 $\mathcal{S}_2^{\mathcal{Z}}$被视为 $\mathcal{Z}$，而 $\mathcal{S}_2$ 被视为 $\mathbb{R}$。请注意，按照概率论中的惯例，两个算子都使用相同的记号 $\mathbb{E}$。

例 4.4（参见第 5 节）。设 $(\Omega 1, \Sigma_1, \mathbb{P}_1) = (\Omega 2, \Sigma_2, \mathbb{P}_2)$ 是一个概率空间，且 $\mathcal{F} \subseteq \Sigma_1$是 Σ_1 的一个子-σ-代数。设 $\mathcal{S}_1 = L^1(\Sigma_1, \mathbb{P}_1)$ 和 $\mathcal{S}_2 = L^1(\mathcal{F}, \mathbb{P}_1|_{\mathcal{F}})$。此外，设 $\mathcal{Z}$是一个巴拿赫空间，$\mathcal{S}_1^{\mathcal{Z}} = L^1(\Sigma_1, \mathbb{P}_1; \mathcal{Z})$ 和 $\mathcal{S}_2^{\mathcal{Z}} = L^1(\mathcal{F}, \mathbb{P}|_{\mathcal{F}}; \mathcal{Z})$。那么，条件期望 $\mathbb{E}[\cdot|\mathcal{F}]: \mathcal{S}_1^{\mathcal{Z}} \to \mathcal{S}_2^{\mathcal{Z}}$是一个取值于 $\mathcal{Z}$ 的马尔可夫算子，与马尔可夫算子$\mathbb{E}[\cdot|\mathcal{F}]: \mathcal{S}_1 \to \mathcal{S}_2$相关联（例如，这可由（Diestel和Uhl，1977，定理 II.2.6）直接推出）。同样，按照惯例，两个算子都使用相同的记号 $\mathbb{E}[\cdot|\mathcal{F}]$。

为将 第 3节中的研究结果扩展到马尔可夫算子，我们为本节作出以下假设：

假设 4.5。 $\mathcal{X}$、$\mathcal{Y}$ 是两个实数、豪斯多夫且局部凸的拓扑向量空间，Ω_1, $\Omega 2$、$\mathcal{S}_1$、$\mathcal{S}_2$、$\mathcal{S}_1^{\mathcal{X}}$、$\mathcal{S}_2^{\mathcal{X}}$、$\mathcal{S}_1^{\mathcal{Y}}$、$\mathcal{S}_2^{\mathcal{Y}}$ 与定义 4.1 中相同，只是分别将 $\mathcal{Z}$ 替换为 $\mathcal{X}$ 和 $\mathcal{Y}$。此外，$\mathcal{E}^{\mathcal{X}}: \mathcal{S}_1^{\mathcal{X}} \to \mathcal{S}_2^{\mathcal{X}}$ 和 $\mathcal{E}^{\mathcal{Y}}: \mathcal{S}_1^{\mathcal{Y}} \to \mathcal{S}_2^{\mathcal{Y}}$ 分别是取值于 $\mathcal{X}$ 和 $\mathcal{Y}$ 的马尔可夫算子，且它们与同一个马尔可夫算子 $\mathcal{E}: \mathcal{S}_1 \to \mathcal{S}_2$ 相关联。$\leq$ 是 $\mathcal{Y}$ 上的任意二元关系；当 $\mathcal{Y} = \mathbb{R}$ 时，我们假定它是典范全序。最后，$K \subseteq \mathcal{X}$、$f: K \to \mathcal{Y}$ 以及 $X: \Omega_1 \to K$，使得有序对 $(X, f(X)) \in \mathcal{S}_1^{\mathcal{X}} \times \mathcal{S}_1^{\mathcal{Y}}$。

引理 4.6。 在 假设 *4.5*下，算子 $\mathcal{E}^{\mathcal{X}\times\mathcal{Y}}: \mathcal{S}_1^{\mathcal{X}} \times \mathcal{S}_1^{\mathcal{Y}} \to \mathcal{S}_2^{\mathcal{X}} \times \mathcal{S}_2^{\mathcal{Y}}$, $(\tilde{X}, \tilde{Y}) \mapsto (\mathcal{E}^{\mathcal{X}}(\tilde{X})\ \mathcal{E}^{\mathcal{Y}}(\tilde{Y}))$ 是一个 $(\mathcal{X}\times\mathcal{Y})$值马尔可夫算子，与同一个马尔可夫算子相关联$\mathcal{E}$。

证明。 设 $\tilde{Z} = (\tilde{X}, \tilde{Y}) \in \mathcal{S}_1^{\mathcal{X}} \times \mathcal{S}_1^{\mathcal{Y}}$和 $\ell \in (\mathcal{X}\times\mathcal{Y})'$。那么 ℓ可写成如下形式：$\ell(x, y) = \ell^{\mathcal{X}}(x) + \ell^{\mathcal{Y}}(y)$, 其中 $\ell^{\mathcal{X}} \in \mathcal{X}'$, $\ell^{\mathcal{Y}} \in \mathcal{Y}'$。因此，$\ell(\tilde{Z}) = \ell^{\mathcal{X}}(\tilde{X}) + \ell^{\mathcal{Y}}(\tilde{Y}) \in \mathcal{S}$（因为 $\mathcal{S}$是一个向量空间），证毕（$\mathcal{S}_2^{\mathcal{X}} \times \mathcal{S}_2^{\mathcal{Y}}$的情形也类似）。此外，

$$\ell(\mathcal{E}^{\mathcal{X}\times\mathcal{Y}}(\tilde{Z})) = \ell^{\mathcal{X}}(\mathcal{E}^{\mathcal{X}}(\tilde{X}) + \ell^{\mathcal{Y}}(\mathcal{E}^{\mathcal{Y}}(\tilde{Y}) = \mathcal{E}(\ell^{\mathcal{X}}(\tilde{X})) + \mathcal{E}(\ell^{\mathcal{Y}}(\tilde{Y})) = \mathcal{E}(\ell(\tilde{Z})).$$

■

现在我们来表述 定理 3.2 关于马尔可夫算子的对应结论：

定理 *4.7*（马尔可夫算子遵守闭凸集约束）。设 Ω_1,$\Omega 2$, $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{E}$,$\mathcal{Z}$, $\mathcal{S}_1^{\mathcal{Z}}$, $\mathcal{S}_2^{\mathcal{Z}}$, $\mathcal{E}^{\mathcal{Z}}$ 如定义 *4.1* 所述，并假设 $\mathcal{Z}$ 是豪斯多夫且局部凸的。此外，设 $Z \in \mathcal{S}_1$ 满足 $Z(\Omega_1) \subseteq A$，其中 $A \subseteq \mathcal{Z}$是某个子集。那么 $(\mathcal{E}^{\mathcal{Z}}(Z))(\Omega_2) \subseteq \mathrm{Conv}(A)$。

证明。 根据 Hahn–Banach分离定理，由于 Conv(A) 是凸集，它等于所有包含它的闭半空间 $\mathcal{H}_{\ell,s} := \{z \in \mathcal{Z} \mid \ell(z) \geq s\} \supseteq$Conv($A$) 的交，其中$\ell \in \mathcal{Z}' \setminus \{0\}$ 且 $s \in \mathbb{R}$。由于对于 $\omega \in \Omega$,

$$\begin{aligned} Z(\Omega_1) \subseteq \mathcal{H}_{\ell,s} &\implies \ell(Z) \geq s \\ &\implies \mathcal{E}(\ell(Z)) \geq s \\ &\implies \ell(\mathcal{E}^{\mathcal{Z}}(Z) \geq s \\ &\implies \mathcal{E}^{\mathcal{Z}}(Z)(\Omega_2) \subseteq \mathcal{H}_{\ell,s}, \end{aligned}$$

$(\mathcal{E}^{\mathcal{Z}}(Z))(\Omega_2)$ 属于每个包含 Conv(A) 的半空间 $\mathcal{H}_{\ell,s}$，从而证明了该结论。 ■

上述定理应用于函数图像的凸包 f 可得：

推论 4.8。 在 假设 *5.1*下，$(\mathcal{E}^{\mathcal{X}}(X)\ \mathcal{E}^{\mathcal{Y}}(f(X)))(\Omega_2) \subseteq \mathrm{Conv}(\mathcal{G}(f))$。

证明。 根据 引理 4.6，由 定理 4.7 可得 $Z := (X,\ f(X))$ 以及 $A := \mathcal{G}(f) \subseteq \mathcal{X} \times \mathcal{Y} =: \mathcal{Z}$。 ■

如同第 3 节一样，上述推论的另一种表述形式给出了马尔可夫算子的图凸包界，可替代期望值：

定理 4.9 ($\mathcal{E}^{\mathcal{Y}}(f(X))$)。 在 假设 *5.1* 下，设 f 的图的凸包被两个“包络”函数 g_l, g_u: $\mathrm{Conv}(K) \to \mathcal{Y}$所“包围”，并满足(3.1)。那么

$$g_l(\mathcal{E}^{\mathcal{X}}(X)) \leq \mathcal{E}^{\mathcal{Y}}(f(X)) \leq g_u(\mathcal{E}^{\mathcal{X}}(X)). \tag{4.1}$$

注意，与定理 *3.4*不同，(4.1) 中的所有对象都是定义在Ω_2上的函数。

证明。 由于（$\mathcal{E}^{\mathcal{X}}$（$X$）$(\Omega_2)$ $\subseteq \overline{\mathrm{Conv}(K)}$）且（$\mathcal{E}^{\mathcal{X}}$（$X$） $\mathcal{E}^{\mathcal{Y}}$（$f$（$X$））$(\Omega_2)$ $\subseteq \mathrm{Conv}(\mathcal{G}$（$f$）） 根据定理 4.7 和推论 4.8,

$$(\mathcal{E}^{\mathcal{Y}}(f(X)))(\Omega_2) \subseteq \overline{\mathrm{Conv}(\mathcal{G}(f))}|_{\mathcal{E}^{\mathcal{X}}(X)} \subseteq [g_l(\mathcal{E}^{\mathcal{X}}(X)), g_u(\mathcal{E}^{\mathcal{X}}(X))].$$

■

推论 4.10。 设定理 *4.9* 的假设成立，并设 $\mathcal{Y} = \mathbb{R}$。那么，使用推论 *3.8*中的记号，

(*a*) $c_l(\mathcal{E}^{\mathcal{X}}(X))\ f(\mathcal{E}^{\mathcal{X}}(X)) \leq \mathcal{E}^{\mathcal{Y}}(f(X)) \leq c_u(\mathcal{E}^{\mathcal{X}}(X))\ f(\mathcal{E}^{\mathcal{X}}(X))$,
(*b*) $\max\left(\inf_{x\in K} f(x)\quad f(\mathcal{E}^{\mathcal{X}}(X)) - s_u\right) \leq \mathcal{E}^{\mathcal{Y}}(f(X)) \leq \min\left(\sup_{x\in K} f(x)\quad f(\mathcal{E}^{\mathcal{X}}(X)) + s_l\right)$,

(*c*) $\hat{c}_l\, f(\mathcal{E}^{\mathcal{X}}(X)) \leq \mathcal{E}^{\mathcal{Y}}(f(X)) \leq \hat{c}_u\, f(\mathcal{E}^{\mathcal{X}}(X))$,

只要这些量有定义。

证明。 (a) 和 (b) 直接由定理 4.9得出，而 (c)由(a)得出。 ■

我们通过将詹森不等式推导为定理4.9的一个直接推论来结束本节。据作者所知，关于马尔可夫算子的詹森不等式尚未在如此一般的层面上得到确立，参见注释 4.2。

推论 4.11 （马尔可夫算子的詹森不等式）。 设 假设 *4.5* 成立，且 $f\colon \mathbb{R}^d \to \mathbb{R}$ 是凸函数。则 $\mathcal{E}^{\mathcal{Y}}(f(X)) \geq f(\mathcal{E}^{\mathcal{X}}(X))$。

证明。 这直接由 定理 4.9 和引理 2.1得出。 ■

5 条件期望的图凸包界

众所周知，詹森不等式对条件期望（Dudley，2002年，定理 10.2.7）同样成立。本节将建立条件期望的图凸包界，这是第 4 节（参见例 4.4）所述更广泛框架的一个具体实例，并直接由其中的结果推导得出。由于条件期望通常是在巴拿赫空间上针对 Bochner 可积随机变量定义的（Diestel和Uhl，1977，第 V.1 节），因此需要稍强一些的假设：

假设 5.1。 $(\Omega, \Sigma, \mathbb{P})$ 是一个概率空间，$\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ 是实巴拿赫空间，且 $K \subseteq \mathcal{X}$ 是 $\mathcal{X}$ 的子集。此外，$\leq$ 是 $\mathcal{Y}$ 上的任意二元关系；在 $\mathcal{Y} = \mathbb{R}$ 的情况下，我们假定它是典范全序；$\mathcal{F} \subseteq \Sigma$ 是一个子-σ-代数，$\mathbb{E}[\cdot|\mathcal{F}]$ 表示相应的条件期望。最后，$K \subseteq \mathcal{X}$, $f\colon K \to \mathbb{R}$ 且 $X\colon \Omega \to K$ 是一个随机变量，使得二元组 $(X, f(X))$ Bochner可积。

定理 5.2 (条件期望遵循闭凸集约束)。 设 $Z\colon (\Omega, \Sigma, \mathbb{P}) \to A$ 是一个取值于实数巴拿赫空间的子集 $A \subseteq \mathcal{Z}$ 中的 *Bochner*可积随机变量，并且 $\mathcal{F} \subseteq \Sigma$ 是一个子-σ-代数。则 $\mathbb{E}[Z|\mathcal{F}] \in \overline{\mathrm{Conv}(A)}$ $\mathbb{P}$-几乎处处。

证明。注意到，正如条件期望中通常的情况一样，对于 $\mathbb{E}[\cdot|\mathcal{F}]\colon L^1(\Sigma, \mathbb{P}) \to L^1(\mathcal{F}, \mathbb{P}|_{\mathcal{F}})$ 和 $\mathbb{E}[\cdot|\mathcal{F}]\colon L^1(\Sigma, \mathbb{P}; \mathcal{Z}) \to L^1(\mathcal{F}, \mathbb{P}|_{\mathcal{F}}; \mathcal{Z})$，记号上并无区别（分别对应于定义 4.1中的 $\mathcal{E}$ 和 $\mathcal{E}_{\mathcal{Z}}$）。由于对于每个 $\ell \in \mathcal{Z}'$，$\mathbb{E}[\ell(Z)|\mathcal{F}] = \ell\mathbb{E}[Z|\mathcal{F}]$（这可由例如（Diestel和Uhl，1977，定理 II.2.6）轻易推出），并且根据（Kallenberg，2021，定理 8.1(ii)），$W \geq s$ 意味着对于每个取实数值的随机变量 $W\colon (\Omega, \Sigma, \mathbb{P})$，$\mathbb{E}[W|\mathcal{F}] \geq s$ $\mathbb{P}$-几乎处处成立，因此该结论直接由定理 4.7 得出。■

注 5.3。 与第 4 节不同，L^1 中的所有结论都只能 $\mathbb{P}$-几乎必然成立。因此，在从定理 4.7 的证明得出定理 5.2的结论时，考虑所有包含 $\mathrm{Conv}(A)$ 的半空间（可能不可数多个）的交集时必须谨慎。不过，按照条件期望的构造方法（Diestel和Uhl，1977，定理 V.1.4），存在 $\mathbb{E}[Z|\mathcal{F}]$ 的一个代表元，同时属于所有这些半空间。

该定理应用于函数 f 的图像的凸包，得到：

推论 5.4。 在 假设 *5.1* 下，$\mathbb{E}[(X, f(X))|\mathcal{F}] \in \overline{\mathrm{Conv}(\mathcal{G}(f))}$ $\mathbb{P}$-几乎必然。

Pro的。 这由 定理 5.2 关于 $Z\colon = (X, f(X))$ 以及 $A\colon = \mathcal{G}(f) \subseteq \mathcal{X} \times \mathcal{Y} =\colon \mathcal{Z}$. ■

定理 5.5（关于 $\mathbb{E}[f(X)|\mathcal{F}]$）。在假设 *5.1* 下，设图 f 的凸包被两个“包络”函数 g_l，g_u: $\mathrm{Conv}(K) \to \mathcal{Y}$ “包围”，并满足（3.1）。则

$$g_l(\mathbb{E}[X|\mathcal{F}]) \leq \mathbb{E}[f(X)|\mathcal{F}] \leq g_u(\mathbb{E}[X|\mathcal{F}]) \ \mathbb{P}\text{-}almost\ surely. \tag{5.1}$$

注意，与 定理 *3.4* 不同，(5.1) 中的所有对象都是随机变量。

证明。 由于 $\mathbb{E}[X|\mathcal{F}] \in$ 凸包(K) $\mathbb{P}$-几乎处处成立，并且根据 $\mathbb{E}[X|\mathcal{F}], \mathbb{E}[f(X)|\mathcal{F}]) \in$ 凸包$(\mathcal{G}(f))$ $\mathbb{P}$-几乎处处成立，定理 5.2 和 推论 5.4，

$$\mathbb{E}[f(X)|\mathcal{F}] \subseteq \overline{\mathrm{Conv}(\mathcal{G}(f))}|_{\mathbb{E}[X|\mathcal{F}]} \subseteq [g_l(\mathbb{E}[X|\mathcal{F}]), g_u(\mathbb{E}[X|\mathcal{F}])] \ \mathbb{P}\text{-almost surely.}$$

■

推论 5.6。 设 定理 *5.5* 的假设成立，且 $\mathcal{Y} = \mathbb{R}$。那么，采用 推论 *3.8* 中的记号，

(a) $c_l(\mathbb{E}[X|\mathcal{F}])\ f(\mathbb{E}[X|\mathcal{F}]) \leq \mathbb{E}[f(X)] \leq c_u(\mathbb{E}[X|\mathcal{F}])\ f(\mathbb{E}[X|\mathcal{F}])$,

(b) $\text{最大}\left(\text{下确界}_{x\in K} f(x) \quad f(\mathbb{E}[X|\mathcal{F}]) - s_u\right) \leq \mathbb{E}[f(X)] \leq \text{最小}\left(\text{上确界}_{x\in K} f(x) \quad f(\mathbb{E}[X|\mathcal{F}]) + s_l\right)$,

(c) $\hat{c}_l\, f(\mathbb{E}[X|\mathcal{F}]) \leq \mathbb{E}[f(X)] \leq \hat{c}_u\, f(\mathbb{E}[X|\mathcal{F}])$,

$\mathbb{P}$-几乎必然地，只要这些量有定义。

证明。 （a）和（b）直接由定理 5.5得出，而（c）由（a）得出。 ■

同样，条件詹森不等式几乎可直接由定理 5.5 :

推论 5.7（条件詹森不等式）。设假设 *5.1* 成立，且 f: $\mathbb{R}^d \to \mathbb{R}$ 是一个凸函数。那么 $\mathbb{E}[f(X)|\mathcal{F}] \geq f(\mathbb{E}[X|\mathcal{F}])$ $\mathbb{P}$-几乎必然。

证明。 这直接由定理 5.5和引理 2.1得出。 ■

致谢

本研究由德国研究联合会（DFG）通过德国卓越战略（EXC-2046/1，项目编号 390685689）资助，资助渠道为柏林数学研究中心 MATH+ 的 EF1-10 项目。作者感谢 Tim Sullivan 仔细校对本稿并提出宝贵建议。

参考文献

S. Abramovich 和 L.-E. Persson。《“詹森差距”的一些新估计》。*J. Inequal. Appl.*，第 39 号论文，9 页，2016 年。doi:10.1186/s13660-016-0985-4。D. Bakry、I. Gentil 和 M. Ledoux。《马尔可夫扩散算子的分析与几何》，收入 *Grundlehren der mathematischen Wissenschaften [Fundamental Principles ofMathematical Sciences]*第 348 卷。Springer, Cham，2014 年。doi:10.1007/978-3-319-00227-9。J. M. Borwein 和 A. S. Lewis。《凸分析与非线性优化》，收入 *CMSBooks in Mathematics/Ouvrages de Math´ematiques de la SMC*第 3 卷。Springer，纽约，第二版，2006 年。doi:10.1007/978-0-387-31256-9。理论与例子。I. Budimir、S. S. Dragomir 和 J. Pe˘cari´c。詹森离散不等式的进一步逆向结果及其在信息论中的应用。*JIPAM. J. Inequal. Pure Appl. Math.*，2(1)：第 5 号论文，14 页，2001 年。

J. Diestel 和 J. J. Uhl。《向量测度》，收入 Mathematical Surveys 第 15 卷。美国数学学会，普罗维登斯，罗德岛州，1977 年。doi:10.1090/surv/015。S. S. Dragomir。《詹森不等式的一些逆向形式及其应用》。Bull. Aust. Math. Soc.，87(2)：177–194，2013 年。doi:10.1017/S0004972712001098。R. M. Dudley。《实分析与概率论》，收入 Cambridge Studies in Advanced Mathematics 第 74 卷。剑桥大学出版社，剑桥，2002年。doi:10.1017/CBO9780511755347。1989 年原版的修订重印本。T. Eisner、B. Farkas、M. Haase 和 R. Nagel。《遍历理论的算子理论方面》，收入 Graduate Texts in Mathematics 第 272 卷。Springer，Cham，2015 年。doi:10.1007/978-3- 319-16898-2。J. L. W. V. Jensen。《关于凸函数及平均值之间的不 ′等式 ′》。Acta Math.，30(1)：175–193，1906 年。doi:10.1007/BF02418571。O. Kallenberg。《现代概率论基础》，收入 Probability Theory and Stochas- tic Modelling 第 99 卷。Springer，Cham，2021 年。doi:10.1007/978-3-030-61871-1。第三版。S. Khan、M. Adil Khan、S. I. Butt 和 Y.-M. Chu。《关于二阶可微函数的詹森差距的新界及其应用》。Adv. Difference Equ.，第 333 号论文，11 页，2020a 年。doi:10.1186/s13662-020-02794-8。S. Khan、M. Adil Khan 和 Y.-M. Chu。《由格林函数导出的詹森不等式逆命题及其在信息论中的应用》。Math. Methods Appl. Sci.，43(5)：2577–2587，2020b 年。doi:10.1002/mma.6066。M. Lin。《无有限不变测度的马尔可夫算子半群的弱混合》。载于 Ergodic theory（会议论文集，数学研究所，Oberwolfach，1978），收入 Lecture Notes in Math. 第 729 卷，第 89–92 页。Springer，柏林，1979 年。doi:10.1007/BFb0063286。M. D. Perlman。《无限维空间上凸向量值函数的詹森不等式》。J. Multivariate Anal.，4：52–65，1974年。doi:10.1016/0047-259X(74)90005-0。R. R. Phelps。《Choquet 定理讲义》，收入 Lecture Notes in Mathematics 第 1757 卷。Springer-Verlag，柏林，第二版，2001 年。doi:10.1007/b76887。R. Rockafellar。《凸分析》。Princeton Landmarks in Mathematics and Physics。普林斯顿大学出版社，1997 年。ISBN 9780691015866。W. Rudin。《泛函分析》。International Series in Pure and Applied Mathematics。McGraw-Hill, Inc.，纽约，第二版，1991 年。R. Rudnicki。《马尔可夫算子：在扩散过程与种群动力学中的应用》。第 27 卷，第 67–79 页。2000 年。doi:10.4064/am-27-1-67-79。谨以此纪念 Wies-law Szlenk。H. Ullah、M. Adil Khan 和 T. Saeed。《詹森差距界的确定及其应用》。Mathematics，9(23)，2021 年。doi:10.3390/math9233132。G. Wunder、B. Groß、R. Fritschek 和 R. F. Schaefer。《一个詹森不等式逆向结果及其在互信息估计中的应用》。载于 2021 IEEE Information Theory Workshop (ITW)，第 1–6 页，2021 年。doi:10.1109/ITW48936.2021.9611449。