arXiv:2601.05030v1 [cs.IT] 8 Jan 2026

有界海森矩阵的二阶可微凸函数的詹森不等式改进

Sambhab Mishra

摘要

詹森不等式由丹麦数学家兼工程师 Johan Jensen 提出，他因对函数理论的贡献而闻名；该不等式是凸分析中无处不在的重要结果，为凸函数的期望提供了一个基本下界。本文针对海森矩阵有界的二阶可微函数，严格推导了该不等式的改进形式。我们利用带积分余项的泰勒展开，尝试弥合经典的基于方差的界与更高精度估计之间的差距。我们还发现了由格吕斯型不等式控制的显式误差项，从而能够将偏度和峰度纳入界中。借助这些新的理论工具，我们改进了对连续分布的香农熵以及瑞利衰落信道遍历容量的现有估计，展现了这些改进在实际应用中的有效性。

1. 引言

Johan Jensen 于 1906年 首次提出的这一不等式，已渗透到现代数学的几乎所有分支。在概率形式下，詹森不等式指出，对于定义在区间 $I \subseteq \mathbb{R}$ 上的实值凸函数 ϕ，以及取值于 I 且期望有限的随机变量 X $\mathbb{E}[X]$，期望的函数值不大于函数值的期望：

$$\phi(\mathbb{E}[X]) \leq \mathbb{E}[\phi(X)]. \tag{1}$$

这一简单关系推广了一个几何事实：连接凸函数图像上任意两点的割线位于图像上方。不等式两边之差正式定义为詹森间隙，记作：

$$\mathcal{J}(\phi, X) = \mathbb{E}[\phi(X)] - \phi(\mathbb{E}[X]). \tag{2}$$

虽然 $\mathcal{J}(\phi,\ X)$ 的非负性足以确立诸如 Kullback-Leibler散度非负性或算术-几何平均不等式等基本结果，但对于要求精确量化误差的现代应用而言，这已远远不够。

- 在变分推断中，詹森间隙表示为准确近似后验分布而必须最小化的“证据下界”（ELBO）间隙。
- 在算子理论中，这一差距量化了将凸函数应用于自伴算子时非交换性所产生的影响。

- 在精算学中，这一差距代表非线性效用函数固有的风险溢价。

1.1 精化的必要性

经典不等式的主要局限在于，它对随机变量 X 的分布（均值之外的部分）以及 ϕ的局部曲率不敏感。对于线性函数，差距为零。随着曲率（二阶导数）增大，或 X 的离散程度增大，差距也会扩大。

经典的精化方法通常依赖于二阶导数的粗略全局界或简单的方差项。例如，若 ϕ 二阶可微且 $m \leq \phi''(x) \leq M$，那么众所周知，该差距有一个上界，其由与方差 σ^2 成正比、并按 $m/2$ 和$M/2$ 缩放的项构成。然而，当出现以下情况时，这些一阶精化方法无法刻画差距的变化：

- **函数具有高度非二次性：**对于对数障碍函数 $-\log x$ 或指数函数 e^x这类函数，其二阶导数在定义域内会有数量级上的变化。

- **全局上界 M 可能是无穷大，也可能过于宽松**，从而使差距的上界失去实际意义。

- **分布是不对称的：** 方差是衡量离散程度的对称指标。如果 X 高度偏斜（例如对数正态分布或帕累托分布），概率密度的质量与 ϕ 不断变化的曲率会以复杂的方式相互作用，无法用单一的方差项来刻画。

- **需要高精度控制：** 在有界域中，例如有界发射功率下的信道容量估计，标准不等式给出的误差范围通常过宽，因而难以用于系统设计。

1.2 分析框架

本文通过运用高阶分析工具，建立一个系统框架来改进詹森不等式，以应对这些局限。我们的方法基于以下三个支柱：

1. **积分余项分析：** 我们不再局限于泰勒定理的拉格朗日余项形式，而是采用积分形式 $R_1(x) = \int_\mu^x (x-t)\phi''(t)dt$。这种表示方式使我们能够运用积分不等式理论中的强大工具，特别是格吕斯不等式和切比雪夫不等式，来界定积分核与二阶导数之间的协方差。

2. **矩展开：**我们汇集了有关四阶展开的最新结果。通过将泰勒级数展开至四次，我们明确地将三阶标准化矩（偏度， γ_3）和四阶标准化矩（峰度， γ_4）引入界中。这提供了一个根据分布形状进行调整的“修正”不等式。

3. **特定应用优化：**我们研究切点经过优化的“类Jensen不等式”。我们不再围绕均值 μ展开，而是考虑围绕一个广义点 c 展开，该点可使特定分布类的误差最小化。

本文所给出的分析严谨且自洽。我们为主要定理提供详细的证明或推导，确保论证从基本原理自然推进到高级应用。随后，将所得界应用于信息论（熵估计）和通信工程（衰落信道容量）中的关键问题。

2. 数学预备知识与定义

首先，我们需要定义所研究的函数类别和概率设置。我们还将回顾作为主要分析工具的基本不等式。

2.1 凸性与广义凸性

我们始终假设 I 是 $\mathbb{R}$ 中的一个区间，且 X 是一个取值于 I 的随机变量，其概率测度为 P。

定义 2.1（凸函数）。 如果对于所有的 x、$y \in I$ 和 $\lambda \in [0,1]$，以下不等式成立，则函数 $\phi: I \to \mathbb{R}$ 是凸函数：

$$\phi(\lambda x + (1-\lambda)y) \leq \lambda\phi(x) + (1-\lambda)\phi(y).$$

如果 ϕ 在 I上二阶可微，那么凸性等价于对所有$x \in I$，条件 $\phi''(x) \geq 0$ 成立。若 $\phi''(x) > 0$，则具有严格凸性。

定义 2.2（强凸函数）。 如果函数 ϕ 满足函数 $\psi(x) = \phi(x) - \frac{m}{2}x^2$ 为凸函数，则称其为参数为 $m > 0$ 的强凸函数。对于二阶可微函数，这意味着海森矩阵具有全局下界：对所有 $x \in I$，$\phi''(x) \geq m$。强凸性是建立詹森间隙下界的一项关键性质。

定义 2.3（（m, M）-凸性）。 我们将强凸性的概念扩展为同时包含上界。若函数 ϕ 满足以下条件，则称其为（m, M）-凸函数：

$$m \leq \phi''(x) \leq M \quad \forall x \in I.$$

该条件意味着 $\phi(x) - \frac{m}{2}x^2$ 是凸函数，且 $\frac{M}{2}x^2 - \phi(x)$也是凸函数。这类函数能够对詹森间隙给出最精确的“夹逼”界，实质上将函数限制在两条曲率分别为 m 和 M的抛物线之间。

2.2 带积分余项的泰勒定理

泰勒定理是我们改进策略的核心工具。虽然多项式近似很常见，但积分余项形式对于应用函数不等式不可或缺。

定理 2.1（泰勒定理）。设 ϕ: $I \to \mathbb{R}$ 是一个函数，且 $\phi^{(n)}$ 在 I上绝对连续。那么对于任意 $a, x \in I$:

$$\phi(x) = \sum_{k=0}^{n} \frac{\phi^{(k)}(a)}{k!}(x-a)^k + R_n(x;a),$$

其中余项为:

$$R_n(x;a) = \frac{1}{n!}\int_a^x (x-t)^n \phi^{(n+1)}(t)dt.$$

在分析詹森间隙时，我们主要关注 $n=1$ 的情形，并在均值 $\mu = \mathbb{E}[X]$ 附近展开。展开式为:

$$\phi(X) = \phi(\mu) + \phi'(\mu)(X-\mu) + \int_\mu^X (X-t)\phi''(t)dt.$$

对等式两边取期望，并注意到线性项 $\mathbb{E}[\phi'(\mu)(X-\mu)]$ 因 $\mathbb{E}[X-\mu]=0$ 而消失，可得詹森间隙的精确积分表示:

$$\mathcal{J}(\phi, X) = \mathbb{E}\left[\int_\mu^X (X-t)\phi''(t)dt\right]. \tag{3}$$

这一恒等式几乎是所有现代改进的出发点。要对间隙进行界定，只需对该积分核的期望进行界定。

2.3 格吕斯不等式与切比雪夫不等式

为了估计积分余项，我们需要一些工具来界定函数乘积的积分。格吕斯不等式给出了两个有界函数协方差的界。

定义 2.4（Chebysev 泛函）。对于两个可积函数 f, g: $[a,b] \to \mathbb{R}$，Chebysev 泛函定义如下:

$$T(f,g) = \frac{1}{b-a}\int_a^b f(t)g(t)dt - \left(\frac{1}{b-a}\int_a^b f(t)dt\right)\left(\frac{1}{b-a}\int_a^b g(t)dt\right).$$

该泛函衡量积分偏离乘法性的程度。注意，如果 f和 g是均匀分布在 $[a,b]$, T上的随机变量，那么（f, g）恰好就是它们的协方差。

定理 2.2（格吕斯不等式）。设 f, g: $[a,b] \to \mathbb{R}$ 为可积函数，且存在常数 $\gamma, \Gamma, \delta, \Delta$ 满足:

$$\gamma \le f(t) \le \Gamma \quad and \quad \delta \le g(t) \le \Delta \quad for\ a.e.\ t \in [a,b].$$

则:

$$|T(f,g)| \le \frac{1}{4}(\Gamma - \gamma)(\Delta - \delta).$$

常数 $1/4$ 是最佳的。

定理 2.3 (Pre-Grüss 不等式)。 格吕斯不等式的一个强化形式通常称为 *Pre-Grüss* 不等式，它将 *Chebysev* 泛函与函数的方差联系起来：

$$|T(f,g)| \le \sqrt{T(f,f)T(g,g)}.$$

其中， $\sqrt{T(f,f)}$ 是 f 在区间 $[a,b]$ 上的标准差。

当函数 f 和 g 不仅有界，而且其质量集中在各自均值附近时，这个不等式尤为有用，因为它能给出比基于取值范围的格吕斯不等式更紧的界。在概率设置中，对于随机变量 Y 和 Z，这些不等式可转化为对 $|\mathrm{Cov}(Y,Z)|$ 的界。若 Y 表示积分核（$X-t$），而 Z 表示海森矩阵 $\phi''(t)$，则这些定理使我们能够将 X 的分布属性与 ϕ 的解析属性分离开来。

3. 有界海森矩阵下基于方差的改进

我们首先推导与（m,M）-凸函数相关的基本界。这些结果确立了间隙估计的基准表现，并说明了其对随机变量方差的直接依赖关系。

3.1 标准方差界

定理 3.1。 设 $\phi: I \to \mathbb{R}$ 是二阶可微函数，且对所有 $x \in I$ 均有 $m \le \phi''(x) \le M$。设 X 是一个取值于 I 的随机变量，其均值为 μ 且方差有限 σ^2。则詹森间隙满足：

$$\frac{m}{2}\sigma^2 \le \mathbb{E}[\phi(X)] - \phi(\mathbb{E}[X]) \le \frac{M}{2}\sigma^2. \tag{4}$$

证明。 我们使用泰勒余项的拉格朗日形式。存在一个随机变量 ξ，其取值介于 μ 和 X 之间，使得：

$$\phi(X) = \phi(\mu) + \phi'(\mu)(X-\mu) + \frac{1}{2}\phi''(\xi)(X-\mu)^2.$$

由于对所有 $x \in I$，$m \le \phi''(x) \le M$，因此必然有 $m \le \phi''(\xi) \le M$。由此可对二次项进行界定：

$$\frac{m}{2}(X-\mu)^2 \le \frac{1}{2}\phi''(\xi)(X-\mu)^2 \le \frac{M}{2}(X-\mu)^2.$$

对整个不等式链两边取期望：

$$\mathbb{E}\left[\phi(\mu) + \phi'(\mu)(X-\mu) + \frac{m}{2}(X-\mu)^2\right] \le \mathbb{E}[\phi(X)] \le \mathbb{E}\left[\phi(\mu) + \phi'(\mu)(X-\mu) + \frac{M}{2}(X-\mu)^2\right].$$

利用期望的线性性，$\mathbb{E}[\phi(\mu)] = \phi(\mu)$ 且 $\mathbb{E}[\phi'(\mu)(X-\mu)] = \phi'(\mu)(\mathbb{E}[X]-\mu) = 0$。偏差平方的期望即为方差，$\mathbb{E}[(X-\mu)^2] = \sigma^2$。代入这些结果可得：

$$\phi(\mu) + \frac{m}{2}\sigma^2 \le \mathbb{E}[\phi(X)] \le \phi(\mu) + \frac{M}{2}\sigma^2.$$

从两边同时减去 $\phi(\mu)$，即可得到詹森间隙的所述上下界 . ■

该定理对詹森间隙提供了一个有力的解释：在一阶近似下，它是波动（σ^2）的能量乘以函数的平均凸性。对于强凸函数（$m > 0$），下界 $\frac{m}{2}\sigma^2$ 严格为正，从而量化了不确定性的“代价”。

3.2 通过划分区域进行改进

如果区间 I 较大，定理 3.1 中的界可能会比较松，因为 M 必须界定整个定义域上的曲率。对于像 $\phi(x) = e^x$ 或 $\phi(x) = 1/x$ 这样的函数，二阶导数会呈指数或几何级数变化。为了解决这个问题，我们可以对定义域进行划分。

定理 3.2 (分区方差界)。 设 $\{I_k\}_{k=1}^K$ 是定义域 I的一个划分，使得 $\cup I_k = I$ 且 $I_j \cap I_k = \emptyset$。令 $p_k = P(X \in I_k)$，并令 σ_k^2和 μ_k分别为给定 $X \in I_k$时 X 的条件方差和均值。令 $m_k = \inf_{x\in I_k} \phi''(x)$ 和 $M_k = \sup_{x\in I_k} \phi''(x)$。则：

$$\sum_{k=1}^{K} p_k \left(\frac{m_k}{2}\sigma_k^2 + \mathcal{J}(\phi, \mu_k)\right) \le \mathcal{J}(\phi, X) \le \sum_{k=1}^{K} p_k \left(\frac{M_k}{2}\sigma_k^2 + \mathcal{J}(\phi, \mu_k)\right).$$

这里，$\mathcal{J}(\phi, \mu_k)$ 表示来自“组间”方差的差距贡献，具体为 $\phi(\mu_k) - \phi(\mu)$。

请注意，该定理实质上将总詹森间隙分解为“分区内”间隙（由局部方差和局部曲率界定）以及“分区间”间隙。

通过将界限 m_k 和 M_k局部化，我们可以避免分布尾部中 ϕ'' 的极端值放宽高概率区域内的界限。例如，在为正态分布的 $\mathbb{E}[e^X]$ 求界时，当 x 较大时，曲率 e^x 会非常大。分区可以隔离尾部的贡献，使中心区域（其中 e^x 较小）的部分得到更紧的界 [1]。

4. 泰勒余项的 Grüss 型改进

虽然基于方差的界限具有稳健性，但它们实质上将 ϕ'' 视为常数（或有界区间）。通过考虑积分核与变化的二阶导数之间的相关性，我们可以获得更高的精度。这便引出了 Grüss 型改进。

4.1 将格吕斯不等式应用于余项

回顾间隙的积分表示：

$$\mathcal{J}(\phi, X) = \mathbb{E}\left[\int_{\mu}^{X} (X - t)\phi''(t)dt\right].$$

让我们分析在固定的 X 实现下的内层积分。若将积分变量 t 视为在 $[\mu, X]$上均匀分布，我们便可对函数 $f(t) = X - t$与 $g(t) = \phi''(t)$ 的乘积应用格吕斯不等式。然而，Dragomir 提出了一种更有力的方法，即直接对概率期望应用格吕斯不等式。我们来定义这些映射。实际上，我们计算的是随机核 $K(t)$ 与导数 $\phi''(t)$ 的协方差。直接应用 Pre-Grüss 不等式可得相应界，其取决于二阶导数的 L_p 范数。

定理 4.1 (Dragomir 的 Grüss 型精化)。 设 $\phi:\ I \to \mathbb{R}$ 满足 ϕ'' 绝对连续，且 $\phi''' \in L_\infty(I)$ （即 ϕ''' 本质有界，且上界为 $\|\phi'''\|_\infty$）。则：

$$\left|\mathcal{J}(\phi, X) - \frac{\phi''(\mu)}{2}\sigma^2\right| \le \frac{\|\phi'''\|_\infty}{6}\mathbb{E}[|X-\mu|^3]. \tag{5}$$

证明： 我们从 $\phi(X)$ 在 μ 处截至二阶导数项的泰勒展开开始：

$$\phi(X) = \phi(\mu) + \phi'(\mu)(X-\mu) + \frac{1}{2}\phi''(\mu)(X-\mu)^2 + R_2(X).$$

余项 $R_2(X)$ 涉及三阶导数：

$$R_2(X) = \frac{1}{2}\int_\mu^X (X-t)^2\phi'''(t)dt.$$

取期望后，一阶项消失。二阶项给出近似 $\frac{\phi''(\mu)}{2}\sigma^2$。该近似的误差为 $|\mathbb{E}[R_2(X)]|$。我们对其进行界定：

$$|R_2(X)| \le \frac{1}{2}\left|\int_\mu^X |X-t|^2|\phi'''(t)|dt\right|.$$

由于 $|\phi'''(t)| \le \|\phi'''\|_\infty$，我们将其从积分中移除：

$$|R_2(X)| \le \frac{\|\phi'''\|_\infty}{2}\left|\int_\mu^X (X-t)^2 dt\right| = \frac{\|\phi'''\|_\infty}{2}\frac{|X-\mu|^3}{3}.$$

取期望可得 $\frac{\|\phi'''\|_\infty}{6}\mathbb{E}[|X-\mu|^3]$。

这一结果通过加入一个与三阶绝对中心矩成正比的修正项，改进了方差界。这意味着，如果分布是对称的，且 ϕ'''有界，那么“简单”的方差近似 $\frac{\phi''(\mu)}{2}\sigma^2$ 的精度可达偏度的阶数。如果 X 是高斯分布，则 $\mathbb{E}[|X-\mu|^3] = 0$ 不成立（它是绝对矩），因此这是一个误差界，而不是修正项。对称分布的三阶代数矩为零，这引出了下一节的四阶展开。

4.2 利用格林函数得到的界

一种十分独特而优雅的方法是使用二阶微分算子的格林函数。该方法可以将詹森间隙精确地表示为二阶导数的加权积分。

定理 4.2 (格林函数表示)。 设 $\phi \in C^2[a,b]$。那么，对于任意随机变量 $X \in [a,b]$，詹森间隙可以表示为：

$$\mathcal{J}(\phi, X) = \int_a^b G(t)\phi''(t)dt,$$

其中，$G(t)$ 是由以下公式定义的格林函数：

$$G(t) = \int_a^t F_X(u)du - \frac{t-a}{b-a}\int_a^b F_X(u)du,$$

假设特定的边界条件，或者更简单地说，与累积分布函数的凸组合 $F_X(t)$ 相关。近期文献中推导出的一种更直接的概率形式 [2] 为：

$$\mathcal{J}(\phi, X) = \frac{1}{2}\int_a^b \phi''(t)\left[\mathbb{E}|X-t| - |\mathbb{E}[X]-t|\right]dt. \tag{6}$$

由于 $|\mathbb{E}[X]-t| = |\mu - t|$，核变为 $K(t) = \frac{1}{2}(\mathbb{E}|X-t| - |\mu-t|)$。注意，$K(t)$ 始终非负，这与凸函数 $\phi(\phi'' \geq 0)$的詹森间隙非负一致。

精化： 对这一表示应用 Chebysev 泛函：

$$\int_a^b K(t)\phi''(t)dt = (b-a)\overline{K}\cdot\overline{\phi''} + (b-a)\mathrm{Cov}_{uniform}(K, \phi''),$$

其中，$\overline{K}$ 是核在 $[a,b]$ 上的均值。项 $(b-a)\,K\cdot\phi''$ 利用平均曲率近似该间隙。格吕斯不等式给出了协方差项的界：

$$\left|\mathcal{J}(\phi, X) - (b-a)\overline{K}\cdot\overline{\phi''}\right| \leq \frac{b-a}{4}(\max K - \min K)(M-m).$$

这提供了一种精化，将凸性的“平均”行为与概率质量和曲率变化之间的具体相互作用区分开来。

5. 四阶矩精化

在许多实际场景中，例如金融风险建模或物理学中的湍流，分布会呈现显著的偏度和峰度（重尾）。基于二阶（方差）的方法所得的界在此并不充分。我们提出一种严格的四阶精化，该精化源自近期进展 [3]。

5.1 四阶展开定理

定理5.1. 设 $\phi \in C^4(I)$，并假设 $\phi^{(5)}$ 存在且在 I上有界。设 X 为均值为 μ、方差为 σ^2、偏度为 $\gamma_3 = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma^3}\right)^3\right]$、峰度为 $\gamma_4 = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma^4}\right)^4\right]$ 的随机变量。那么，期望 $\mathbb{E}[\phi(X)]$ 可展开为：

$$\mathbb{E}[\phi(X)] = \phi(\mu) + \frac{\phi''(\mu)}{2}\sigma^2 + \frac{\phi'''(\mu)}{6}\sigma^3\gamma_3 + \frac{\phi^{(4)}(\mu)}{24}\sigma^4\gamma_4 + \mathcal{R}_4, \tag{7}$$

余项有如下界：

$$|\mathcal{R}_4| \leq \frac{\|\phi^{(5)}\|_\infty}{120}\mathbb{E}[|X-\mu|^5].$$

推导： 我们对 $\phi(x)$ 在 μ 处进行四阶泰勒展开：

$$\phi(x) = \phi(\mu) + \phi'(\mu)(x-\mu) + \frac{\phi''(\mu)}{2}(x-\mu)^2 + \frac{\phi'''(\mu)}{6}(x-\mu)^3 + \frac{\phi^{(4)}(\mu)}{24}(x-\mu)^4 + R_4(x).$$

对每一项取期望:

$$\begin{aligned}
\mathbb{E}[\phi(\mu)] &= \phi(\mu).\\
\mathbb{E}[\phi'(\mu)(X-\mu)] &= 0.\\
\mathbb{E}\left[\frac{\phi''(\mu)}{2}(X-\mu)^2\right] &= \frac{\phi''(\mu)}{2}\sigma^2.\\
\mathbb{E}\left[\frac{\phi'''(\mu)}{6}(X-\mu)^3\right] &= \frac{\phi'''(\mu)}{6}\mathbb{E}[(X-\mu)^3] = \frac{\phi'''(\mu)}{6}\sigma^3\gamma_3.\\
\mathbb{E}\left[\frac{\phi^{(4)}(\mu)}{24}(X-\mu)^4\right] &= \frac{\phi^{(4)}(\mu)}{24}\mathbb{E}[(X-\mu)^4] = \frac{\phi^{(4)}(\mu)}{24}\sigma^4\gamma_4.
\end{aligned}$$

余项 $\mathcal{R}_4 = \mathbb{E}[R_4(X)]$ 使用拉格朗日形式 $R_4(x) = \frac{\phi^{(5)}(\xi)}{120}(x-\mu)^5$进行界定。~~因此，~~ $|\mathbb{E}[R_4(X)]| \le \frac{\sup|\phi^{(5)}|}{120}\mathbb{E}[|X-\mu|^5]$。

5.2 “修正后的”Jensen不等式

该定理使我们能够表述一个“修正后的”Jensen不等式。对于具有正偏度（$\gamma_3 > 0$）的分布和三阶导数为正的函数（例如 e^x），标准的詹森下界 $\phi(\mu)$ 乃至方差精化 $\phi(\mu) + \frac{\phi''}{2}(\mu)\sigma^2$ 都会持续低估真实期望。

推论 5.2 (带符号的精化). 若 $\phi'''(\mu)\gamma_3 \ge 0$ 且 $\phi^{(4)}(\mu)\gamma_4 \ge 0$ （并且 $\phi^{(5)}$ 可忽略或受到控制），则:

$$\mathbb{E}[\phi(X)] \ge \phi(\mu) + \frac{\phi''(\mu)}{2}\sigma^2 + \frac{\phi'''(\mu)}{6}\sigma^3\gamma_3.$$

这一改进对保险定价至关重要。如果 ϕ 表示指数效用函数（风险厌恶），而 X 表示损失（右偏），那么偏度项为正。忽略这一项会导致风险溢价定价过低。对于重尾损失，四阶项（峰度）会进一步推高保费，从而为应对极端事件提供有数学依据的缓冲 [3]。

6. 詹森–默瑟不等式与切点优化

虽然泰勒展开以均值 μ为中心进行分析，但这并不总是近似的最优点。Simic 等人近期关于詹森–默瑟不等式的研究表明，可以对切点进行优化 [5, 6]。

6.1 类詹森不等式

标准詹森不等式使用 μ处的切线: $L(x) = \phi(\mu) + \phi'(\mu)(x-\mu)$。凸性意味着 $\phi(x) \ge L(x)$。考虑在任意点 c处作切线。定义 $L_c(x) = \phi(c) + \phi'(c)(x-c)$。那么 $\mathbb{E}[\phi(X)] \ge \mathbb{E}[L_c(X)] = \phi(c) + \phi'(c)(\mu - c)$。

我们定义函数 $h(c) = \phi(c) + \phi'(c)(\mu - c)$。为了找到最佳下界，我们最大化 $h(c)$ [4]。

$$h'(c) = \phi'(c) + \phi''(c)(\mu - c) - \phi'(c) = \phi''(c)(\mu - c).$$

令 $h'(c) = 0$，则有 $\phi''(c) = 0$ 或 $c = \mu$。由于 ϕ 是凸函数，$\phi'' \geq 0$。若为严格凸函数，则唯一的临界点是 $c = \mu$。因此，对于基本仿射下界，均值 μ 确实是最优的。

然而，如果我们采用二次下界（抛物线型下估函数），或采用涉及其他函数 $g(x)$ 的界（例如 $\phi(x) \geq g(x)$，其中 g 更简单但非线性），最优参数可能会发生偏移。Simić [5] 探讨了如下形式的不等式：

$$\mathbb{E}[\phi(X)] \geq \phi(c) + \phi'(c)(\mu - c) + \frac{m}{2}(\sigma^2 + (\mu - c)^2).$$

这里，下界由仿射项加上一个源自强凸性的二次修正项构成。对 c 求最大值可得到一个更精细的界。如果 m 接近平均曲率，那么对于非对称分布，将 c 移向分布的众数（而非均值）有时能得到更紧的结果。

6.2 詹森-默瑟不等式

詹森-默瑟不等式给出了一个涉及区间端点的间隙上界。

定理 6.1. 如果 ϕ 在 $[a, b]$ 上是凸函数，则：

$$\phi\left(a + b - \sum w_i x_i\right) \leq \phi(a) + \phi(b) - \sum w_i \phi(x_i).$$

利用格林函数方法，人们对该不等式进行了改进，通过方差和高阶项有效地界定“默瑟间隙”。当随机变量受限于有限区间时，这些改进尤为相关，例如在贝塔分布分析中 [5, 6]。

7. 概率与统计学中的应用

7.1 协方差界

随机变量的两个函数的协方差 $\mathrm{Cov}(f(X)\ g(X))$，在统计物理学和经济学中经常出现。如果 f 和 g 的单调方向相同，切比雪夫代数不等式指出协方差非负。利用我们基于导数的界：

$$\mathrm{Cov}(f(X), g(X)) \approx \mathbb{E}[f'(X)g'(X)]\mathrm{Var}(X).$$

如果 $f(x) = x$ 且 $g(x) = \phi'(x)$，这近似于 $\mathrm{Cov}(X,\ \phi'(X)) \approx \mathbb{E}[\phi''(X)]\sigma^2$。将其与基于方差的詹森间隙 $\frac{1}{2}\mathbb{E}[\phi''(X)]\sigma^2$ 相比较，我们发现两者相差一个因子 2。利用这一关系，我们可以用协方差来界定詹森间隙：

$$\frac{1}{2}\inf \phi''\mathrm{Var}(X) \leq \mathcal{J}(\phi, X) \leq \mathrm{Cov}(X, \phi'(X)).$$

这个上界 当 ϕ' 比 ϕ 本身更容易分析时，r 界尤其有用 [7] .

7.2 矩母函数

估计矩母函数（MGF） $M_X(t) = \mathbb{E}[e^{tX}]$是一个经典应用。设 $\phi(x) = e^{tx}$。 $\phi''(x) = t^2 e^{tx}$。直接应用定理 3.1 需要对 e^{tx} 进行界定。在 $(-\infty, \infty)$ 上，它无界。不过，对于有界随机变量 $X \in [a, b]$:

$$e^{t\mu} + \frac{t^2 e^{ta}}{2}\sigma^2 \leq \mathbb{E}[e^{tX}] \leq e^{t\mu} + \frac{t^2 e^{tb}}{2}\sigma^2.$$

这为有界随机变量的矩母函数提供了明确且可计算的界，在大偏差理论中很有用。

8. Applications in Information Theory

信息论建立在凸函数之上，例如 $\log x$ 和 $x\log x$。对詹森不等式的改进可直接转化为对数据压缩和传输[4, 9]的改进界限。

8.1 香农熵的改进界

对于概率密度函数为 $f(x)$的连续随机变量 X ，其微分熵为 $h(X) = -\int f(x)\log f(x)dx$。已有研究推导出混合分布的微分熵相关的詹森型界；参见 [12]。

设 $Y = f(X)$。则 $h(X) = -\mathbb{E}[\log Y]$。由于 $-\log$是凸函数，因此 $\mathbb{E}[-\log Y] \geq -\log \mathbb{E}[Y]$。

$$\mathbb{E}[Y] = \int f(x)^2 dx = E,$$

信息 能量。因此， $h(X) \geq -\log E$，即阶数为 2.

精化的Jensen不等式给出了误差项：

$$h(X) = -\log E + \mathcal{R}.$$

对 $\phi(y) = -\log y$ 使用二阶精化：

$$\mathcal{R} \approx \frac{1}{2}\mathbb{E}\left[\frac{1}{Y^2}\right]\mathrm{Var}(Y) = \frac{1}{2}\int \frac{1}{f(x)^2}\mathrm{Var}(f(X))dx.$$

该表达式将《熵》亏损直接与概率密度函数本身的变异性联系起来。对于均匀分布，$\mathrm{Var}(f(X)) = 0$，且该界是紧的。对于高度尖峰的分布，方差较大，而Rényi 熵给出的界较为松[8]。

贝塔分布熵： 对于 $X \sim$ 贝塔分布(α, β)，熵的表达式涉及二伽马函数 $\psi(\alpha)$。使用 $\log X$ 的四阶展开，我们可以用涉及方差和偏度的有理函数高精度地近似 $\psi(\alpha)$ 项，从而在数值计算程序 [10] 中无需计算特殊函数。

8.2 Kullback-Leibler散度（反向Pinsker不等式）

KL散度 $D_{KL}(P||Q) = \mathbb{E}_P[\log \frac{P(X)}{Q(X)}]$ 非负。令 $Z = \frac{Q(X)}{P(X)}$。则 $D_{KL} = \mathbb{E}_P[-\log Z]$。注意 $\mathbb{E}_P[Z] = 1$。Jensen不等式： $\mathbb{E}[-\log Z] \geq -\log \mathbb{E}[Z] = -\log 1 = 0$。

改进（定理 3.1）：

$$D_{KL}(P||Q) \geq \frac{1}{2} \inf \left(\frac{1}{Z^2} \right) \mathrm{Var}_P(Z).$$

$\inf(1/Z^2) = \inf(P/Q)^2$. $\mathrm{Var}_P(Z) = \mathbb{E}_P[(Q/P)^2] - (\mathbb{E}_P[Q/P])^2 = \chi^2(Q||P)$.

因此，我们推导出 KL 散度与卡方散度之间更精细的关系：

$$D_{KL}(P||Q) \geq \frac{1}{2} \left(\inf \frac{P(x)}{Q(x)} \right)^2 \chi^2(Q||P).$$

这是一种“反向Pinsker不等式”，建立了比标准Pinsker不等式更强的拓扑结构；标准Pinsker不等式通过全变差距离 [11] 对 KL 散度进行界定。

9. Applications in Wireless Communications

在无线衰落信道分析中，容量是一个依赖于信道增益的随机变量。遍历容量是其期望。对于 MIMO信道，已有研究分析了容量近似中 Jensen型界的紧致性 [14]。

9.1 瑞利衰落容量

容量为 $C = B\log_2(1 + \text{信噪比} \cdot |h|^2)$。设 $\gamma = |h|^2$。对于瑞利衰落， $\gamma \sim$ 服从Exp(1)。$\phi(\gamma) = \log_2(1 + \rho\gamma)$，其中 $\rho =$ 信噪比。 ϕ 是凹函数。

Jensen不等式标准上界： $\mathbb{E}[C] \leq \log_2(1 + \rho)$。该上界在渐近意义下较松（差值随 $\log \rho$ 增大）。

应用四阶精化（定理5.1）：对于 $\gamma \sim$ 服从Exp(1)， $\mu = 1, \sigma^2 = 1, \gamma_3 = 2, \gamma_4 = 9$。 $\phi(x)$ 的导数为 $= \ln(1 + \rho x)$（以纳特为单位）：

$$\phi'(1) = \frac{\rho}{1+\rho}.$$
$$\phi''(1) = \frac{-\rho^2}{(1+\rho)^2}.$$
$$\phi'''(1) = \frac{2\rho^3}{(1+\rho)^3}.$$
$$\phi^{(4)}(1) = \frac{-6\rho^4}{(1+\rho)^4}.$$

展开：

$$\mathbb{E}[\ln(1 + \rho\gamma)] \approx \ln(1+\rho) - \frac{1}{2}\frac{\rho^2}{(1+\rho)^2}(1) + \frac{1}{6}\frac{2\rho^3}{(1+\rho)^3}(2) - \frac{1}{24}\frac{6\rho^4}{(1+\rho)^4}(9).$$

化简：

$$C_{approx} \approx \ln(1+\rho) - \frac{1}{2}\left(\frac{\rho}{1+\rho}\right)^2 + \frac{2}{3}\left(\frac{\rho}{1+\rho}\right)^3 - \frac{9}{4}\left(\frac{\rho}{1+\rho}\right)^4.$$

在高信噪比（$\frac{\rho}{1+\rho} \to 1$）下，修正项为 $-0.5 + 0.66 - 2.25 \approx -2.09$。实际差距是与欧拉–马歇罗尼常数相关的项。这个展开表明，纳入偏度（正项）和峰度（负项）旨在捕捉对数函数在零附近（即"深衰落"区域）的复杂行为。与简单的方差惩罚相比，这一改进准确得多；简单的方差惩罚仅为 -0.5，会严重低估衰落造成的容量损失 [13]。

10. 数值验证

为展示所提出下界的有效性，我们针对函数 $\phi(x) = e^{-x}$（其中 $X \sim$ 服从均匀分布）给出数值比较。

- 真实值：$\mathbb{E}[e^{-X}] = \frac{1-e^{-2}}{2} \approx 0.432$。
- 詹森下界：$e^{-1} \approx 0.368$。（误差：0.064）。

表 1：$\mathbb{E}[e^{-X}]$、$X \sim U$ 的界限比较

方法	公式	数值	相对误差
精确	$\int_0^2 0.5e^{-x}dx$	0.4323	0%
Jensen不等式（经典）	$e^{-\mu}$	0.3679	-14.9%
方差精化	$e^{-\mu} + \frac{e^{-\mu}}{2}\sigma^2$	$0.3679 + \frac{0.3679}{2}(0.333)$	0.4292
四阶	展开（含峰度）	0.4325	+0.04%

方差精化将误差从 15% 大幅降低至 0.7%。对于这个光滑函数和有界分布，四阶展开几乎消除了误差。这证实了理论预测：纳入高阶矩可使解析函数的逼近精度得到指数级提升。

11. 未来工作

未来，我们希望将本文提出的四阶詹森间隙精化从标量情形推广至更广的范围，并研究其在算子、多元及学习理论背景中的影响。这些方向是当前工作的自然延伸，并引出了以下研究计划。

- **算子詹森不等式：** 将四阶精化推广至希尔伯特空间上自伴算子的情形。算子的非交换性会引入"交换子"项，可采用类似技术对其进行界定 [15, 16]。

- **多元扩展：** 为随机向量的凸函数构建显式的基于张量的精化 $\mathbf{X} \in \mathbb{R}^n$。一个核心技术挑战是控制高阶混合中心矩，这些矩自然会归纳为共偏度和共峰度

张量；四矩界的显式多元类比将有助于多元不确定性量化和随机优化。

- **深度学习损失函数：**将改进后的詹森间隙界应用于变分目标，例如变分自编码器中的证据下界（ELBO）。对詹森间隙进行更严格的解析控制，有望产生更优的替代目标并缩小变分间隙，从而为现代变分推断中的近似误差提供有原则的控制方法。 [17]。

总之，本文提出的改进旨在弥合经典不等式理论与现代随机建模对高精度的要求之间的差距；我们期望上述拓展能进一步加强二者之间的联系。

12. 致谢

作者谨向 Govind Chandra Sarangi 博士和 Ashok Kumar Panigrahi 博士致谢，感谢他们给予的支持和宝贵意见。

参考文献

[1] Liao, J. 和 Arthur Berg. “詹森不等式的强化。”《美国统计学家》，第 73 卷，2017 年 7 月，https://doi.org/10.1080/00031305.2017.1419145.[2] Adil Khan, Muhammad、Shahid Khan 等。“利用 S-凸性对詹森间隙的新估计及其应用。”《物理学前沿》，第 8 卷，2020 年，https://www.frontiersin.org/journals/ physics/articles/10.3389/fphy.2020.00313.[3] Zeghdoudi, Halim. 2025. “詹森不等式的四阶矩细化：严谨性、余项界及其在精算中的应用。” Journal of Time Scales Analysis 1 (2): 55–65. https://doi.org/10. 64721/8s4t0t84.[4] Merhav, Neri. “若干类詹森不等式及其在信息论中的应用。”《熵》，第 25 卷，第 5 期，2023 年，第 752 页，https://doi.org/10.3390/e25050752.[5] Simic, Slavko. 詹森不等式的若干推广。2020, https://doi.org/10.48550/arXiv. 2011.10746.[6] Adil Khan, Muhammad、Slavica Ivelić Bradanović 等。“强凸函数的詹森–默瑟不等式的新改进及其应用。”《公理》，第 13 卷，第 8 期，2024 年，第 553 页，https://doi.org/10.3390/axioms13080553.[7] Hössjer, Ola 和 Arvid Sjölander。“有界随机变量协方差的精确上下界。”《统计学与概率通讯》，第 182 卷，2021 年 12 月，第 109323 页，https://doi.org/10.1016/j.spl.2021.109323.[8] Saeed, Tareq 等。“詹森不等式的细化及其应用。”《AIMS 数学》，第 7 卷，第 4 期，2022 年，第 5328–46 页，https://doi.org/10.3934/math.2022297.[9] Xiao, Lei 和 Guoxiang Lu.“詹森不等式的一种新细化及其在信息论中的应用。”《开放数学》，第 18 卷，2020 年 12 月，第 1748–59 页，https://doi.org/10.1515/ math-2020-0123.[10] Lu, Guoxi.“詹森不等式的新细化与熵上界。” Journal of Mathe- matical Inequalities，第 12 卷，2018 年 6 月，第 403–21 页，https://doi.org/10.7153/jmi-2018-12-30.[11] Dragomir, Silvestru.“詹森不等式的一种细化及其在 f-散度测度中的应用。” TAIWANESE JOURNAL OF MATHEMATICS，第 14 卷，2010 年 2 月，第 153–64 页，https://doi.org/10. 11650/twjm/1500405733。

[12] J. Melbourne 等。“混合分布的微分熵：新的界及其应用。” IEEE 信息论汇刊，第 68 卷，第 4 期，2022 年 4 月，第 2123–46 页，https://doi.org/10. 1109/TIT.2022.3140661.[13] Allanki, Sanyasi Rao. 瑞利平坦衰落信道容量。第 13 卷，2021 年 12 月，第 1890–98 页.[14] Yuan, J、Matthaiou, M、Jin, S 和 Gao F，2017 年，“詹森界的紧致性及其在 MIMO 通信中的应用”，IEEE 通信汇刊，第 65 卷，第 2 期，第 579-593 页。https: //doi.org/10.1109/TCOMM.2016.2623945[15] Dragomir，“ 希尔伯特空间中自伴算子可微对数凸函数的新詹森型不等式。” 萨拉热窝数学杂志，第 7 卷，2024 年 6 月，第 67–80 页，https: //doi.org/10.5644/SJM.07.1.07.[16] Khosravi, Maryam 等。“蔡－戴维斯－詹森不等式的细化。” 数学分析与应用通报，第 3 卷，2011 年 1 月，第 127–33 页。[17] 詹森间隙的紧致界：一种用于生成式建模的新方法，https: //doi.org/10.48550/arXiv.2502.03988