arXiv:2304.03063v1 [cs.IT] 6 Apr 2023

若干类詹森不等式及其在信息论中的应用

内里·梅尔哈夫

安德鲁与厄娜·维特比电气与计算机工程学院，以色列理工学院，理工学院城，海法 32000，以色列 电子邮件：merhav@technion.ac.il

谨以此文纪念雅各布·齐夫，

他是信息论天空中一颗璀璨的明星

多年来一直激励着我们许多人，今后亦将如此。

摘要

众所周知，传统的詹森不等式是通过用经过点（$\boldsymbol{E}\{X\}$， $f(\boldsymbol{E}\{X\})$）的切向仿射函数，对给定的凸函数 $f(x)$进行下界估计来证明的，其中 $\boldsymbol{E}\{X\}$ 是随机变量 X 的期望。尽管在所有由与 f 相切的仿射函数诱导的下界中，该切向仿射函数给出了最紧的下界，但事实证明，当函数 f 只是某个更复杂表达式的一部分，而我们需要对该表达式的期望进行界定时，最紧的下界可能属于经过不同于（$\boldsymbol{E}\{X\}$， $f(\boldsymbol{E}\{X\})$）的某一点的切向仿射函数。在本文中，我们利用这一观察结果，在多种情况下针对给定的具体表达式优化切点，由此推导出若干族不等式，以下将其称为“类詹森”不等式；据作者所知，这些不等式都是新的。通过若干与信息论相关的应用实例，展示了这些不等式的紧致性程度及其潜在实用性。

1 引言

众所周知，詹森不等式是信息论等多个领域中最基础、最有用的数学工具之一。有趣的是，它还将许多其他广为人知且本身十分重要的不等式包含为其特例。众多例子中，

我们提及施瓦茨–柯西不等式（它又为不确定性原理和克拉美–罗下界提供支持）、刘维尔不等式、霍尔德不等式，以及调和平均数、几何平均数和算术平均数之间的不等式。在信息论领域，詹森不等式是信息不等式（即相对熵的非负性）、数据处理不等式（它又引出法诺不等式）以及条件熵与无条件熵之间不等式的基础。此外，它在推导香农理论中的单字母公式以及矩约束下的最大熵理论中发挥着核心作用（例如，参见 [1] 的第 12 章）。

在过去二十年中，围绕詹森不等式开展了大量研究工作，包括改进 [2], [3], [4], [5], 变体 [6], [7], [8], 完善 [9], [10], [11], 和扩展 [12], 等，不一而足。此外，关于反向詹森不等式也有许多推导，例如参见 [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23] 和 [24], ，其中列举了部分相关文献。在大多数研究中，所推导的不等式都通过许多应用加以说明，例如算术平均数与几何平均数之间的有用关系、熵的逆向界、相对熵以及更一般的 f-散度、H¨older 不等式的逆向形式等。在许多此类工作中，主要结果都表示为对差值的上界，即 $\boldsymbol{E}\{f(X)\} - f(\boldsymbol{E}\{X\})$，其中 f 是凸函数， $\boldsymbol{E}\{\cdot\}$是期望算子， X 是随机变量。然而，这些界主要取决于与 f相关的全局参数，例如其值域和定义域，而并不特别依赖于 X所对应的基础概率函数（连续情形下为概率密度函数，离散情形下为概率质量函数）。一方面，反向詹森不等式的一个理想性质应当是：当 X 高度集中在其均值附近时，该不等式应当是紧的，这正如普通詹森不等式所具有的众所周知的性质一样。在 [24], 中，作者尝试处理这一问题。

本文从另一个角度重新审视詹森不等式。它基于以下简单观察，而这一观察根植于詹森不等式的证明：给定的凸函数 $f(x)$，有切向仿射函数 $\ell(x) = f(a) + f'(a)(x-a)$ 作为其下界，其中 a 是 x 定义域中的任意数， $f'(a)$ 是 f 在 $x = a$ 处的导数（前提是 f 在 $x = a$ 处可导）。通过选取 $a = \boldsymbol{E}\{X\}$ 并对不等式两边取期望， $f(X) \geq \ell(X)$，即可轻易证明詹森不等式。需要记住的是

在这里， $a_* = \boldsymbol{E}\{X\}$ 是 a 的最优选择，即在 a 的所有可能取值中使 $\boldsymbol{E}\{\ell(X)\}$ 最大化，从而在这类关于 $\boldsymbol{E}\{f(X)\}$ 的下界中得到最紧的下界。然而，当函数 $f(X)$ 只是某个更复杂表达式的一部分，而我们要为该表达式的期望求下界时， a 的最优选择可能不同于 $\boldsymbol{E}\{X\}$ 。例如，我们可能想为 $\boldsymbol{E}\{g[f(X)]\}$ 求下界，其中 g 是单调非递减函数；或者为 $\boldsymbol{E}\{f(X)g(X)\}$ 求下界，其中 g 是正函数和／或凸函数；也可能是两者的组合，等等。

为了说明这一事实，考虑以下示例（将在第 2 节中详细讨论）：对 $\boldsymbol{E}\{f(X)g(X)\}$ 求下界，其中 g 是正函数。在这种情况下，

$$\boldsymbol{E}\{f(X)g(X)\} \geq \boldsymbol{E}\{[f(a) + f'(a)(X-a)]g(X)\}, \tag{1}$$

并通过对右端（r.h.s.）关于 a 取最大值，我们很容易得到，此处的最优选择为 a $a_* = \boldsymbol{E}\{Xg(X)\}/\boldsymbol{E}\{g(X)\}$，从而得到该不等式，

$$\boldsymbol{E}\{f(X)g(X)\} \geq f\left(\frac{\boldsymbol{E}\{Xg(X)\}}{\boldsymbol{E}\{g(X)\}}\right) \cdot \boldsymbol{E}\{g(X)\}, \tag{2}$$

只要 g 满足我们能够轻松计算 $\boldsymbol{E}\{g(X)\}$ 和 $\boldsymbol{E}\{Xg(X)\}$，这就很有用。虽然这个特定的不等式也可以通过对密度 $\tilde{p}(x) = p(x)g(x)/\int_{-\infty}^{\infty} p(x')g(x')\mathrm{d}x'$ 应用（普通）詹森不等式 $\boldsymbol{E}\{f(X)\} \geq f(\boldsymbol{E}\{X\})$ 得到，但我们将在下文看到，还有各种像这样没有明显简单解释的不等式。此后，我们将这类不等式称为类詹森不等式，因为它们的推导采用了与经典詹森不等式证明所依据的相同一般思路。我们还将展示这些不等式在信息论中的用途。

我们的贡献在本项工作中具有以下特点：

1. 在许多情况下（如上例），参数的最优取值（例如上述讨论中的参数a）可以用闭式求得。在其他情况下，所得表达式可能不适合进行闭式优化，此时我们有两种选择：(i) 进行数值优化；(ii) 任意选择 a，从而得到一个有效的下界，同时记住，经过合理推测有可能得到较好的下界。

2. 我们的不等式提供两类界：(i) 需要计算 X 的前两个矩（或等价地，前两个累积量）的界；(ii) 需要计算 X 的矩母函数（MGF）及其导数，或等价地，计算 X 的累积量生成函数（CGF）及其导数的界。所有这些类型的矩通常都很容易以闭式计算，尤其是在 X 由独立同分布（i.i.d.）随机变量之和给出的情况下，这种情况在信息论应用中十分常见。

3. 我们的大多数推导都可推广到多变量凸函数 .

4.我们所考虑的类詹森不等式类别具有足够的灵活性，可以推导出不一定是凸函数、甚至某些凹函数的下界，从而为反向詹森不等式开辟另一条途径。这可以通过将给定函数表示为所讨论类别中的一种形式来实现（例如，凸函数与正函数的乘积、两个正凸函数的乘积、单调函数与凸函数的复合等）。

5. 我们通过几个与信息论相关的示例展示了类詹森不等式的实用性。我们还给出了数值结果，以说明这些界的紧致性程度。

6. 我们的类詹森不等式具有理想的性质：随着 X 越来越集中于其均值附近，它们会变得更加严格，就像普通詹森不等式一样。

7. 在全文中，我们仅讨论包含凸函数 f 的表达式的期望下界，但应当理解，如果 f 为凹函数且不等式方向相反，这些结论仍然适用。

8. 应当理解，我们在本文中推导出的类詹森不等式只是一些示例，旨在说明一项基本思路：针对具体表达式，优化给定凸函数的切点。可以设想，同样的思路还可应用于许多在理论和实践上更具意义的情形。

最后，有必要作一条技术说明：在下文的所有推导中，我们都假设凸函数 f 至少在最优点 $x = a_*$处可微。应当理解，

不过，这一假设主要是为了方便，并非出于实际需要。众所周知，凸函数在任意点都至少存在单侧导数，因此$f'(a_*)$可以取 f 在$x = a_*$处左导数与右导数之间的任意值。为了证明下界（关于 a）的导数为零的点确实使下界达到最大值（而非最小值等），我们还需要进一步假设 f二阶可微。不过，这一假设不会限制所述下界的适用性，因为该下界适用于 a的任意取值，包括导数为零的点，即使使用标准方法无法证明该点使下界达到最大值也是如此。当下界依赖于多个参数时，也有类似的说明。

本文余下部分的每一节都将探讨一类不同的詹森型不等式，它们对应于包含凸函数 f 的表达式的不同形式。

2 凸函数与正函数的乘积

本节重点讨论如何为形如 $\boldsymbol{E}\{f(X)g(X)\}$ 的表达式求下界，其中 f 是凸函数， g 是非负函数。事实上，设 $f : \mathrm{IR} \to \mathrm{IR}$ 为凸函数， $g : \mathrm{IR} \to \mathrm{IR}^+$ 为非负函数。那么，对于任意 $a \in \mathrm{IR}$,

$$\begin{aligned} \boldsymbol{E}\{f(X)g(X)\} &\geq \boldsymbol{E}\{[f(a) + f'(a)(X - a)]g(X)\} && (3) \\ &= [f(a) - af'(a)]\boldsymbol{E}\{g(X)\} + f'(a)\boldsymbol{E}\{Xg(X)\}. && (4) \end{aligned}$$

为求 a 的值，使右侧达到最大，我们令导数等于零，并得到 t:

$$[f'(a) - f'(a) - af''(a)]\boldsymbol{E}\{g(X)\} + f''(a)\boldsymbol{E}\{Xg(X)\} = 0 \quad (5)$$

或者等价地,

$$f''(a)[\boldsymbol{E}\{Xg(X)\} - a\boldsymbol{E}\{g(X)\}] = 0, \quad (6)$$

由此得到

$$a_* = \frac{\boldsymbol{E}\{Xg(X)\}}{\boldsymbol{E}\{g(X)\}}, \quad (7)$$

并且容易验证，在 $a = a_*$ 处的二阶导数为 $-f''(a_*)\boldsymbol{E}\{g(X)\} < 0$，这意味着它是一个最大值（至少是局部最大值）。因此， $\boldsymbol{E}\{f(X)g(X)\}$ 的下界为

由下式给出

$$\boldsymbol{E}\{f(X)g(X)\} \geq f\left(\frac{\boldsymbol{E}\{Xg(X)\}}{\boldsymbol{E}\{g(X)\}}\right) \cdot \boldsymbol{E}\{g(X)\}. \tag{8}$$

只要 f 关于 X 的所有分量联合凸，这一结果便可直接推广到 X 为向量的情形。特别地，它可推广到 f 和 g 分别作用于不同随机变量 X 和 Y 的情形，其联合分布为：

$$\boldsymbol{E}\{f(X)g(Y)\} \geq f\left(\frac{\boldsymbol{E}\{Xg(Y)\}}{\boldsymbol{E}\{g(Y)\}}\right) \cdot \boldsymbol{E}\{g(Y)\}. \tag{9}$$

接下来我们考虑几个例子。

例1. 令 $f(x) = -\ln x$ 且 $g(x) = x$, $x > 0$。应用不等式 (8)，

$$\boldsymbol{E}\{-X \ln X\} \geq -\boldsymbol{E}\{X\} \cdot \ln \frac{\boldsymbol{E}\{X^2\}}{\boldsymbol{E}\{X\}} = -\boldsymbol{E}\{X\} \cdot \ln(\boldsymbol{E}\{X\}) - \boldsymbol{E}\{X\} \cdot \ln\left(1 + \frac{\mathrm{Var}\{X\}}{[\boldsymbol{E}\{X\}]^2}\right). \tag{10}$$

注意到函数 $-x\ln x$ 是凹函数，而非凸函数，但在这里，我们却得到了其期望的下界（而非上界），即一个反向詹森不等式。最右侧的第一项是对 $\boldsymbol{E}\{-X \ln X\}$ 的（通常的）詹森上界，第二项则是差距项，它不仅取决于 X 的期望，还取决于其方差，这体现了围绕 $\boldsymbol{E}\{X\}$ 的波动。该不等式可直接用于求取由无记忆信源生成的序列的经验熵期望下界，这与通用信源编码 [25] 相关。经验熵的每一项都具有 $-X\ln X$ 的形式，其中 $X = N(u)/N$， $N(u)$ 表示从有限字母表为 $\mathcal{U}$ 的无记忆信源 P 随机抽取的 N 元组中，字母 u 出现的次数。显然，每个 $N(u)$ 都是一个具有 N 次试验、成功概率为 $P(u)$ 的二项随机变量。在这种情况下，$\boldsymbol{E}\{X\} = P(u)$ 且 $\mathrm{Var}\{X\} = P(u)[1 - P(u)]/N$。因此，分别用以下符号表示熵和经验熵：

$$H = -\sum_{u \in \mathcal{U}} P(u) \ln P(u) \tag{11}$$

$$\hat{H} = -\sum_{u \in \mathcal{U}} \frac{N(u)}{N} \ln\left(\frac{N(u)}{N}\right), \tag{12}$$

按照 $0 \ln 0 \triangleq 0$ 的约定，我们有：

$$\begin{aligned}
\boldsymbol{E}\{\hat{H}\} &\geq -\sum_{u\in\mathcal{U}} P(u)\ln P(u) - \sum_{u\in\mathcal{U}} P(u)\ln\left[1+\frac{P(u)[1-P(u)]/N}{P^2(u)}\right] \\
&= H - \sum_{u\in\mathcal{U}} P(u)\ln\left(1+\frac{1-P(u)}{NP(u)}\right) \\
&\geq H - \sum_{u\in\mathcal{U}} P(u)\cdot\frac{1-P(u)}{NP(u)} \\
&= H - \frac{1}{N}\sum_{u\in\mathcal{U}}[1-P(u)] \\
&= H - \frac{|\mathcal{U}|-1}{N}, \qquad (13)
\end{aligned}$$

其中，$|\mathcal{U}|$是 $\mathcal{U}$的基数。使用普通詹森不等式得到的是上界^，而非下界 $\boldsymbol{E}\{H\} \leq H$。至此，例1得证。

例2. 设 s 和 t 是两个实数，其差 $s-t$要么为负，要么大于1。现在，令 $g(x) = x^t$，且 $f(x) = x^{s-t}$。那么，

$$\begin{aligned}
\boldsymbol{E}\{X^s\} &= \boldsymbol{E}\{X^t X^{s-t}\} \\
&\geq \left(\frac{\boldsymbol{E}\{X^{t+1}\}}{\boldsymbol{E}\{X^t\}}\right)^{s-t}\cdot\boldsymbol{E}\{X^t\} \\
&= \frac{(\boldsymbol{E}\{X^{t+1}\})^{s-t}}{(\boldsymbol{E}\{X^t\})^{s-t-1}}. \qquad (14)
\end{aligned}$$

特别地，对于 $t=1$ 和 $s \notin (1,2)$，这变为

$$\boldsymbol{E}\{X^s\} \geq \frac{(\boldsymbol{E}\{X^2\})^{s-1}}{(\boldsymbol{E}\{X\})^{s-2}} = [\boldsymbol{E}\{X\}]^s\cdot\left(1+\frac{\mathrm{Var}\{X\}}{[\boldsymbol{E}\{X\}]^2}\right)^{s-1} \qquad (15)$$

这再次给出了一个仅取决于 X的前两个矩的界。对于 $s \in (0,1)$，函数 x^s 是凹函数，因此这是詹森不等式的一个反向版本。对于 $s \leq 0$ 和$s \geq 2$，函数 x^s 是凸函数，因此这是詹森不等式的一个改进版本：第一因子 $[\boldsymbol{E}\{X\}]^s$对应普通詹森不等式，而第二因子体现了这一改进，它取决于相对波动项 $\mathrm{Var}\{X\}/[\boldsymbol{E}\{X\}]^2$。

为了进一步具体化这个例子，考虑在分布 Q 下进行随机猜测的问题（参见例如 [26] 及其中的大量参考文献）。那么，在已知（但猜测者不知道） X 的情况下，猜中离散字母表随机变量 X 一次的概率为

$Q(x)$。在连续猜测直至首次猜中时，猜测次数 G 是参数为 $p = Q(x)$ 的几何随机变量，其均值和方差分别为 $1/p$ 和 $(1-p)/p^2$。对于 $s \in (1, 2)$,

$$\boldsymbol{E}\{G^s\} \geq \left(\frac{1}{p}\right)^s \cdot \left(1 + \frac{(1-p)/p^2}{1/p^2}\right)^{s-1} = \frac{(2-p)^{s-1}}{p^s} = \frac{[2-Q(x)]^{s-1}}{[Q(x)]^s}. \tag{16}$$

示例 2 到此完成。

例3. 设 f 为任意凸函数，且 $g(x) = e^{sx}$，其中 s 为给定实数。于是，不等式（8）变为:

$$\boldsymbol{E}\{f(X)e^{sX}\} \geq f(\psi'(s)) \cdot e^{\psi(s)} \tag{17}$$

其中

$$\psi(s) = \ln \boldsymbol{E}\{e^{sX}\} \tag{18}$$

是 X 的累积量生成函数，而 $\psi'(s)$是其导数。这给出了关于 X 的累积量生成函数及其导数的下界。普通詹森不等式是 $s = 0$ 的特殊情形，其中 $\psi(0) = 0$ 且 $\psi'(0) = \boldsymbol{E}\{X\}$。

3 单调函数与凸函数的复合

另一类类詹森不等式对应于对形如 $\boldsymbol{E}\{g[f(X)]\}$ 的表达式求下界的需求，其中 f 与之前一样是凸函数，而 g 是单调不减函数。其总体思路是对以下不等式右侧进行优化。

$$\boldsymbol{E}\{g[f(X)]\} \geq \sup_a \boldsymbol{E}\{g[f(a) + f'(a)(X-a)]\}. \tag{19}$$

在重要的特殊情况下，当 $g(x) = e^x$ 时，我们有:

$$\begin{aligned} \boldsymbol{E}\{e^{f(X)}\} &\geq \sup_a \boldsymbol{E}\{e^{f(a)+f'(a)(X-a)}\} \\ &= \sup_a e^{f(a)-af'(a)} \boldsymbol{E}\{e^{Xf'(a)}\} \\ &= \exp\left\{\sup_a\{f(a) - af'(a) + \psi[f'(a)]\}\right\}, \end{aligned} \tag{20}$$

其中，$\psi(\cdot)$ 再次表示 X 的累积量生成函数。a 的最优值 a_* 是通过令指数的导数等于零所得到方程的解，即:

$$\psi'[f'(a_*)] = a_*, \quad \text{provided that} \quad f''(a_*)\psi''[f'(a_*)] < 1, \tag{21}$$

w其中， $\psi'(\cdot)$ 和 $\psi''(\cdot)$ 分别是 $\psi(\cdot)$ 的一阶导数和二阶导数 .

例4. 考虑 $f(x) = sx^2/2$ 和 $X \sim \mathcal{N}(\mu, \sigma^2)$ 的情形，其中 $\sigma^2 < 1/s$，否则 $\boldsymbol{E}\{e^{sX^2}\} = \infty$。在这种情况下，条件 $f''(a_*)\psi''[f'(a_*)] < 1$ 等价于 $\sigma^2 < 1/s$，并且我们有 $f'(a) = sa$、$\psi(t) = \mu t + \sigma^2 t^2/2$，因此 $\psi'(t) = \mu + \sigma^2 t$，这意味着 $\psi'[f'(a)] = \mu + \sigma^2 sa$。于是，最优 a 的方程变为

$$\mu + \sigma^2 sa = a, \tag{22}$$

其解为

$$a = a_* \stackrel{\Delta}{=} \frac{\mu}{1 - \sigma^2 s}, \tag{23}$$

由此得到

$$\boldsymbol{E}\left\{e^{sX^2/2}\right\} \geq \exp\left\{sa_*^2/2 - sa_*^2 + \mu s a_* + \sigma^2 s^2 a_*^2/2\right\} = \exp\left\{\frac{\mu^2 s}{2(1 - \sigma^2 s)}\right\}. \tag{24}$$

普通詹森不等式给出

$$\boldsymbol{E}\left\{e^{sX^2/2}\right\} \geq \exp\left\{s\boldsymbol{E}\{X^2\}/2\right\} = e^{s(\mu^2 + \sigma^2)/2}, \tag{25}$$

这无法捕捉 $s = 1/\sigma^2$ 处的奇点。精确计算得到

$$\boldsymbol{E}\left\{e^{sX^2/2}\right\} = \frac{1}{\sqrt{1 - \sigma^2 s}} \cdot \exp\left\{\frac{\mu^2 s}{2(1 - \sigma^2 s)}\right\}, \tag{26}$$

也就是说，类詹森界 (24) 给出了正确的指数项（以及 $s = 1/\sigma^2$ 处的奇点），与精确值的差异仅在于指数前因子。

4 凸函数与单调凸复合函数的乘积

另一类类詹森不等式对应于对两个函数乘积的期望给出下界，其中一个是凸函数，另一个是由正单调非递减函数与凸函数复合而成的函数，即，

$$\boldsymbol{E}\{h[f(X)]g(X)\} \geq \sup_{a,b} \boldsymbol{E}\{h[f(a) + f'(a)(X - a)] \cdot [g(b) + g'(b)(X - b)]\}, \tag{27}$$

其中 f 和 g 是凸函数， h 是单调非递减且非负的。对于 $h(x) = e^x$的情况，我们得到一个依赖于 X 的累积量生成函数及其导数的界：

$$\begin{aligned}
\boldsymbol{E}\{e^{f(X)}g(X)\} &\geq \boldsymbol{E}\left\{e^{f(a)+f'(a)(X-a)}[g(b)+g'(b)(X-b)]\right\} &(28)\\
&= e^{f(a)-af'(a)}\boldsymbol{E}\left\{e^{Xf'(a)}[g(b)-bg'(b)+g'(b)X]\right\} &(29)\\
&= \exp\{f(a)-af'(a)+\psi[f'(a)]\}\{g(b)+g'(b)(\psi'[f'(a)]-b)\}. &(30)
\end{aligned}$$

在固定 a 的同时，对 b 求最大值，得到 $b_* = \psi'[f'(a)]$, ，因此我们得到：

$$\boldsymbol{E}\{e^{f(X)}g(X)\} \geq \sup_a \exp\{f(a)-af'(a)+\psi[f'(a)]\}\cdot g(\psi'[f'(a)]). \tag{31}$$

示例5. 考虑 $f(x) = -\ln x$ 且 $g(x) = x\ln x$的情形，我们可以得到一个反向Jensen-like 不等式，即凹函数 $\ln X$的期望的下界：

$$\begin{aligned}
\boldsymbol{E}\{\ln X\} &= \boldsymbol{E}\left\{e^{-\ln X}\cdot X\ln X\right\} &(32)\\
&\geq \sup_{a\geq 0}\exp\{-\ln a+1+\psi(-1/a)\}\cdot\psi'(-1/a)\ln\psi'(-1/a) &(33)\\
&= \sup_{\alpha\geq 0}\exp\{\ln\alpha+1+\psi(-\alpha)\}\psi'(-\alpha)\ln\psi'(-\alpha) &(34)\\
&= e\cdot\sup_{\alpha\geq 0}\alpha e^{\psi(-\alpha)}\psi'(-\alpha)\ln\psi'(-\alpha) &(35)\\
&= e\cdot\sup_{\alpha\geq 0}\alpha\boldsymbol{E}\{Xe^{-\alpha X}\}\ln\frac{\boldsymbol{E}\{Xe^{-\alpha X}\}}{\boldsymbol{E}\{e^{-\alpha X}\}}. &(36)
\end{aligned}$$

定义矩母函数 $\phi(s) = \boldsymbol{E}\{e^{sX}\} = e^{\psi(s)}$，我们有：

$$\begin{aligned}
\boldsymbol{E}\{\ln X\} &\geq e\cdot\sup_{\alpha\geq 0}\alpha\phi'(-\alpha)\ln\psi'(-\alpha) &(37)\\
&= e\cdot\sup_{\alpha\geq 0}\alpha\phi(-\alpha)\psi'(-\alpha)\ln\psi'(-\alpha) &(38)\\
&= e\cdot\sup_{\alpha\geq 0}\alpha\phi'(-\alpha)\ln\frac{\phi'(-\alpha)}{\phi(-\alpha)}. &(39)
\end{aligned}$$

我们得到了一个用矩母函数及其导数（或等价地，用累积量生成函数及其导数）表示的下界，这在 X 是独立同分布随机变量之和的情况下很有用。

因此，我们现在进一步具体考察这一例子，研究 $X = 1+\sum_{i=1}^k Y_i^2$的情形，其中 $Y_i \sim \mathcal{N}(0,\sigma^2)$、$i = 1$、.....、$k$为相互独立的随机变量。考察 $\boldsymbol{E}\left\{\ln\left(1+\sum_{i=1}^k Y_i^2\right)\right\}$这类表达式有双重动机。首先，它有助于界定单输入多输出（SIMO）信道的遍历容量，

其中 $\{Y_i\}$ 表示随机信道传输系数（例如参见 [24], [27], [28] 及其中的参考文献）。其次，它与界定多元柯西分布对应的联合微分熵有关。这里，(Y_1、.....、Y_k)并非如上所定义的高斯随机变量，但其多元柯西分布可表示为零均值独立同分布高斯随机变量的连续混合，其中混合遍及所有可能的方差——详见 [24, Example6] 。在这种情况下，

$$\begin{aligned}\phi(s) &= \boldsymbol{E}\left\{\exp\left(s\left[1+\sum_{i=1}^{k}Y_i^2\right]\right)\right\} &(40)\\ &= e^s\left(\boldsymbol{E}\{e^{sY^2}\}\right)^k &(41)\\ &= \frac{e^s}{(1-2s\sigma^2)^{k/2}}, \qquad s<\frac{1}{2\sigma^2}. &(42)\end{aligned}$$

因此，

$$\psi(s) = s - \frac{k}{2}\ln(1-2s\sigma^2), \tag{43}$$

and

$$\psi'(s) = 1 + \frac{k\sigma^2}{1-2s\sigma^2}. \tag{44}$$

由此可得

$$\boldsymbol{E}\left\{\ln\left(1+\sum_{i=1}^{k}Y_i^2\right)\right\} \geq e\cdot\sup_{\alpha\geq 0}\left\{\frac{\alpha e^{-\alpha}}{(1+2\alpha\sigma^2)^{k/2}}\left(1+\frac{k\sigma^2}{1+2\alpha\sigma^2}\right)\ln\left(1+\frac{k\sigma^2}{1+2\alpha\sigma^2}\right)\right\}. \tag{45}$$

詹森上界、$\ln(1+k\sigma^2)$和下界(45) 针对 $\sigma^2=1$ 和$k=1$、2、.. .、100显示于图1 中。如图所示，这些界相当接近。有趣的是，选择 $\alpha = 1/(k\sigma^2)$ 所得到的结果与最优 α 的结果非常接近。

该例的另一个实例是圆对称复高斯信道，其信噪比（SNR） Z是一个随机变量（例如，由衰落引起），且发射机和接收机均知晓该变量。其容量为 $C = \boldsymbol{E}\{\ln(1+gZ)\}$，其中 g 是某个确定性增益因子，而期望是相对于 Z的随机性而言的。为简单起见，假设 Z 服从指数分布，即，

$$p(z) = \begin{cases}\theta e^{-\theta z} & z\geq 0\\ 0 & z<0\end{cases} \tag{46}$$

其中参数 $\theta>0$ 已给定。在这种情况下，

$$\phi(-\alpha) = \frac{\theta e^{-\alpha}}{\theta+g\alpha} \tag{47}$$

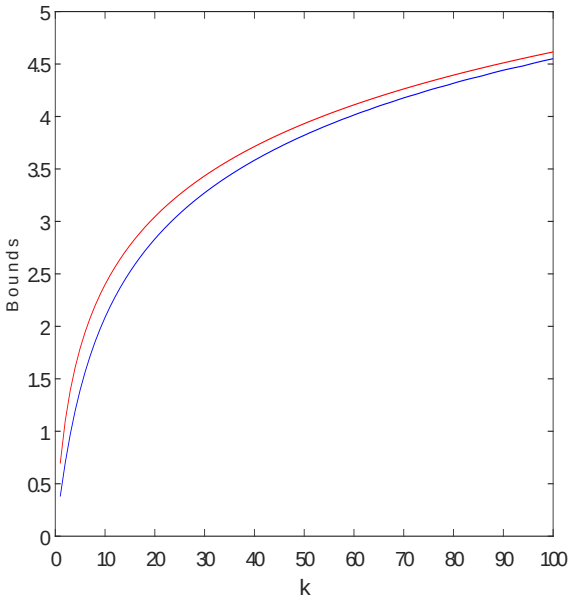

图1：对于$\sigma^2 = 1$和 $k = 1,\ 2,\ \ldots\ldots,\ 100,\ \ Y_i \sim \mathcal{N}(0, \sigma^2)$ 为独立同分布时，$\boldsymbol{E}\left\{\ln\left(1 + \sum_{i=1}^{k} Y_i^2\right)\right\}$ 的上界和下界。红色曲线是应用普通詹森不等式得到的上界 $\ln(1 + k\sigma^2)$。蓝色曲线是公式（45）的下界，其中对 α 的搜索以 0.001 的分辨率进行。

and

$$\psi(-\alpha) = \ln\theta - \ln(\theta + g\alpha) - \alpha, \tag{48}$$

因此，

$$\boldsymbol{E}\{\ln(1 + gZ)\} \geq e\theta \cdot \sup_{\alpha \geq 0} \frac{\alpha e^{-\alpha}}{\theta + g\alpha} \cdot \left(1 + \frac{g}{g + \theta\alpha}\right) \ln\left(1 + \frac{g}{g + \theta\alpha}\right). \tag{49}$$

在图2中，我们将此下界绘制为 θ 关于 $g = 5$ 的函数，并将其与詹森上界 $\ln(1 + g/\theta)$（红色曲线）以及第4.1节例 1] 中的 [24, 下界进行比较。可以看出，此处提出的下界要紧得多，尤其是在 θ 较小时。

例 6. 这类类詹森不等式的另一个例子，可用于求得 $\boldsymbol{E}\{X^t\}$ 的下界，其中 t 为任意实数。给定 t，令 $s \geq 0$ 大于或等于

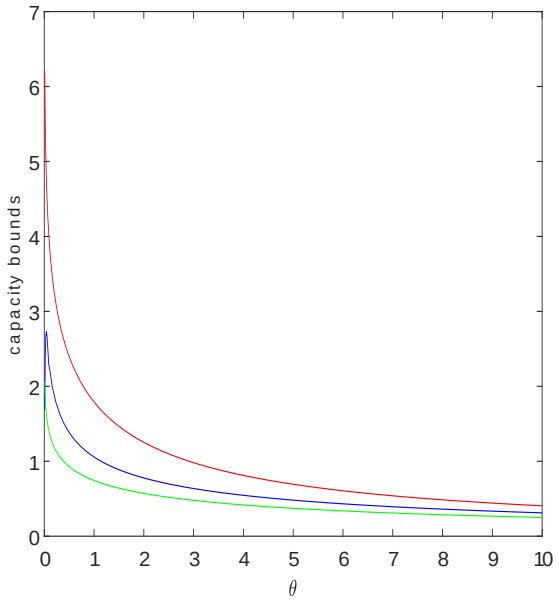

图 2：对于 $g=5$，当 Z 服从参数为 θ 的指数分布时，$\boldsymbol{E}\{\ln(1+gZ)\}$ 的上界和下界随 θ 的变化。红色曲线是上界 $\ln(1+g/\theta)$，由应用普通詹森不等式得到。蓝色曲线是式 (49) 的下界，其中对 α 的搜索以 0.001 为分辨率进行。绿色曲线是[24, Example 1]的下界。

$1-t$ 或小于 $-t$，考虑 $f(x)=x^{t+s}$, $g(x)=-s\ln x$ 且 $h(x)=e^x$ 的情况。然后，

$$\begin{aligned}
\boldsymbol{E}\{X^t\} &= \boldsymbol{E}\{e^{-s\ln X}X^{t+s}\} && (50)\\
&\geq \boldsymbol{E}\left\{\exp\left[s\left(-\ln a-\frac{1}{a}(X-a)\right)\right]\cdot\left[b^{t+s}+(t+s)b^{t+s-1}(X-b)\right]\right\} && (51)\\
&= e^{s[1-\ln a]}\phi\left(-\frac{s}{a}\right)\left[b^{t+s}+(t+s)b^{t+s-1}\left(\psi'\left(-\frac{s}{a}\right)-b\right)\right]. && (52)
\end{aligned}$$

取 $b=\psi'(-s/a)$，并将优化变量 a 改为 $\alpha=1/a$，得到

$$\boldsymbol{E}\{X^t\}\geq\sup_{\alpha\geq 0}(\alpha e)^s\phi(-\alpha s)[\psi'(-\alpha s)]^{t+s}. \tag{53}$$

更具体地，若 $X=\sum_{i=1}^n Y_i$，其中 $\{Y_i\}$ 是参数为 p 的独立同分布伯努利分布变量，则 $\phi(s)=(pe^s+q)^n$，其中 $q=1-p$。于是得到

$$\boldsymbol{E}\{X^t\}\geq\sup_{\alpha\geq 0}(\alpha e)^s(pe^{-\alpha s}+q)^n\cdot\left(\frac{npe^{-\alpha s}}{pe^{-\alpha s}+q}\right)^{t+s}. \tag{54}$$

选择 $\alpha=1/(np)$，我们得到

$$\boldsymbol{E}\{X^t\}\geq(np)^t\cdot\frac{e^s(pe^{-s/(np)}+q)^n e^{-s(t+s)/(np)}}{(pe^{-s/(np)}+q)^{t+s}}. \tag{55}$$

第一个因子是 $(\boldsymbol{E}X)^t$。第二个因子随着 n 增大而趋于 1，因为 $pe^{-s/np}+q \approx p(1-s/(np))+q=1-s/n$，因此 $(pe^{-s/np}+q)^n \approx (1-s/n)^n \approx e^{-s}$。对于 $t \geq 1$ 和 $t \leq 0$，函数 $f(x)=x^t$ 是凸函数，因此 $(\boldsymbol{E}X)^t$ 是通常的詹森下界。在这种情况下，如果乘法因子，管用。

$$\frac{e^s(pe^{-s/(np)}+q)^n e^{-s(t+s)/(np)}}{(pe^{-s/(np)}+q)^{t+s}},$$

大于 1。若 $0<t<1$，函数 $f(x)=x^t$ 是凹函数，则$(\boldsymbol{E}X)^t$ 是上界。当然，参数 s 也可以进行优化。 $t=0.5$ 的一些数值结果显示于图 3 中。可以看出，上界和下界相当接近。

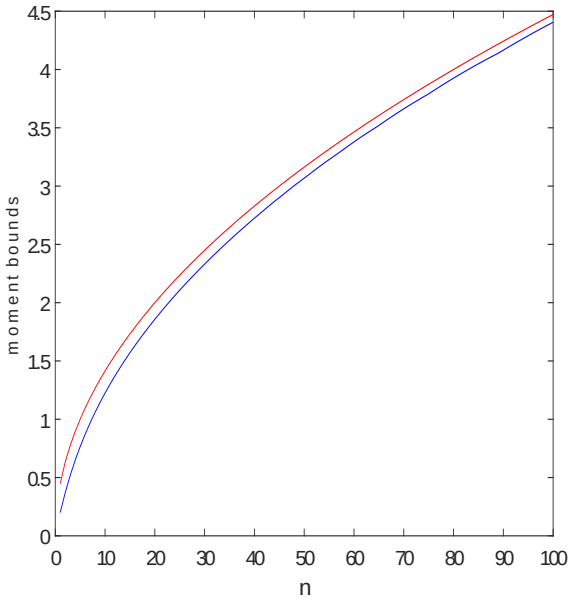

图 3： $\boldsymbol{E}\{\sqrt{\sum_{t=1}^n Y_t}\}$ 关于 n的上界和下界，其中 $\{Y_t\}$ 为独立同分布的伯努利分布(0.2)。红色曲线为詹森上界$\sqrt{np}$，蓝色曲线为所提出的下界，其中 α 在范围 $[0,10]$ 内优化， s 在范围 $[0.5,10]$ 内优化，两者的分辨率均为 0.01。

此例的另一项应用与估计理论有关。设 $\theta \in \mathrm{IR}$，且令$Y_1,\ldots,Y_n$ 为独立同分布随机变量，其均值为 θ，方差为 σ^2。考虑估计误差的 t 阶矩 $\boldsymbol{E}_\theta\left|\frac{1}{n}\sum_{i=1}^n Y_i-\theta\right|^t$。定义 $X=\left(\frac{1}{n}\sum_{i=1}^n Y_i-\theta\right)^2$，则有

$$\phi(s)=\frac{1}{\sqrt{1-2s\sigma^2/n}}; \quad \psi(s)=-\frac{1}{2}\ln\left(1-\frac{2s\sigma^2}{n}\right). \tag{56}$$

因此，

$$\phi(-\alpha s) = \frac{1}{\sqrt{1+2\alpha s\sigma^2/n}}; \quad \psi'(-\alpha s) = \frac{\sigma^2/n}{1+2\alpha s\sigma^2/n}. \tag{57}$$

$$\begin{aligned} \boldsymbol{E}_\theta \left| \frac{1}{n}\sum_{i=1}^n Y_i - \theta \right|^t &= \boldsymbol{E}_\theta X^{t/2} && (58) \\ &\geq \frac{(\alpha e)^s}{\sqrt{1+2\alpha s\sigma^2/n}} \left(\frac{\sigma^2/n}{1+2\alpha s\sigma^2/n} \right)^{t/2+s} \\ &= \left(\frac{\sigma^2}{n}\right)^{t/2+s} \cdot \frac{(\alpha e)^s}{(1+2\alpha s\sigma^2/n)^{(t+1)/2+s}}. && (59) \end{aligned}$$

其中 ei 无论是 $s \geq 1-t/2$ 还是 $s \leq -t/2$。对于 $\alpha = \zeta n/\sigma^2$ （$\zeta > 0$ 为常数），我们有 ：

$$\boldsymbol{E}_\theta \left| \frac{1}{n}\sum_{i=1}^n Y_i - \theta \right|^t \geq \frac{\sigma^t}{n^{t/2}} \cdot \sup_{\zeta>0,\ s>1-t/2} \frac{(\zeta e)^s}{(1+2\zeta s)^{(t+1)/2+s}} \tag{60}$$

其中对于 $t \in [0,2]$，第一个因子 $\sigma^t/n^{t/2}$ 是詹森上界。第二个因子，

$$\mu_t = \sup_{\zeta>0,\ s>1-t/2} \frac{(\zeta e)^s}{(1+2\zeta s)^{(t+1)/2+s}}, \tag{61}$$

是詹森上界与所提出的下界之间的差距。在图4中，我们展示了这个因子。结果 $\mu_2 = 1$ 是意料之中的，因为对于 $t=2$ 和 $s=0$，计算 是显然精确的。注意，对于给定的 s，对 ζ的最大化可以通过令 $\ln[(\zeta e)^s/(1+2\zeta s)^{(t+1)/2+s}]$ 关于 ζ 的偏导数等于零，以闭式求出。最优的 ζ 结果等于 $1/(t+1)$（与 s 无关），因此，

$$\mu_t = \sup_{s>1-t/2} \left(\frac{t+1}{t+2s+1} \right)^{(t+1)/2} \cdot \left(\frac{e}{t+2s+1} \right)^s. \tag{62}$$

最后需要指出的是，这一类类詹森界也为形式为 $\boldsymbol{E}\{f(X)/g(X)\}$ 的下界计算开辟了道路，其中 f 是正凸函数， g 是正凹函数。利用恒等式 $1/s = \int_0^\infty e^{-st}\mathrm{d}t$，我们得到:

$$\begin{aligned} \boldsymbol{E}\left\{\frac{f(X)}{g(X)}\right\} &= \boldsymbol{E}\left\{ f(X) \cdot \int_0^\infty e^{-tg(X)}\mathrm{d}t \right\} && (63) \\ &= \int_0^\infty \boldsymbol{E}\left\{ e^{-tg(X)} f(X) \right\} \mathrm{d}t && (64) \end{aligned}$$

我们可以对被积函数运用与之前相同的思路，并可自由优化界参数，使其可能依赖于 t。

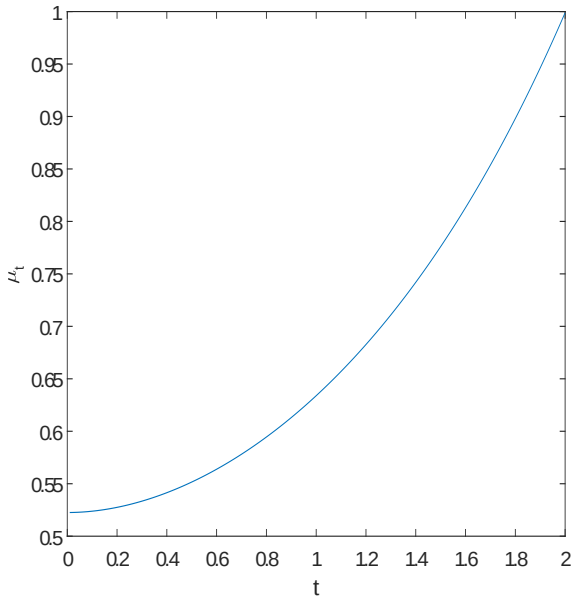

图 4：间隙因子 μ_t 随 t 变化的函数关系。参数 s 在范围 $[1-t/2, 10]$ 内进行优化，分辨率为 0.001。

5 两个正凸函数的乘积

本文介绍的最后一类詹森型界与两个非负凸函数的乘积有关。设 f 和 g 都是关于 $x \geq 0$ 的非负凸函数。那么，

$$
\begin{aligned}
\boldsymbol{E}\{f(X)g(X)\} &\geq \boldsymbol{E}\{[f(a)+f'(a)(X-a)]\cdot g(X)\} && (65)\\
&= [f(a)-af'(a)]\boldsymbol{E}\{g(X)\}+f'(a)\boldsymbol{E}\{Xg(X))\} && (66)\\
&\geq [f(a)-af'(a)]\boldsymbol{E}\{[g(b)+g'(b)(X-b)]\}+ \\
&\quad f'(a)\boldsymbol{E}\{X[g(c)+g'(c)(X-c)]\} \qquad f(a)\geq af'(a)\geq 0 && (67)\\
&= [f(a)-af'(a)]\cdot[g(b)-bg'(b)+g'(b)\boldsymbol{E}\{X\}]+ \\
&\quad f'(a)[(g(c)-cg'(c))\boldsymbol{E}\{X\}+g'(c)\boldsymbol{E}\{X^2\}\}]. && (68)
\end{aligned}
$$

最优的 b 和 c 分别为 $b^*=\boldsymbol{E}\{X\}$ 和 $c^*=\boldsymbol{E}\{X^2\}/\boldsymbol{E}\{X\}$。因此，

$$
\boldsymbol{E}\{f(X)g(X)\}\geq [f(a)-af'(a)]\cdot g(\boldsymbol{E}\{X\})+f'(a)\boldsymbol{E}\{X\}\cdot g\left(\frac{\boldsymbol{E}\{X^2\}}{\boldsymbol{E}\{X\}}\right). \qquad (69)
$$

Let

$$a^* = \frac{\boldsymbol{E}\{X\} \cdot g(\boldsymbol{E}\{X^2\}/\boldsymbol{E}\{X\})}{g(\boldsymbol{E}\{X\})} \tag{70}$$

以及 assu 使得 $f(a^*) \geq a^* f'(a^*) \geq 0$。那么， a^* 是 a的最优值，由此得到 ds

$$\boldsymbol{E}\{f(X)g(X)\} \geq f\left(\frac{\boldsymbol{E}\{X\} \cdot g(\boldsymbol{E}\{X^2\}/\boldsymbol{E}\{X\})}{g(\boldsymbol{E}\{X\})}\right) \cdot g(\boldsymbol{E}\{X\}). \tag{71}$$

更一般地，当 X 和 Y 是具有联合分布的两个随机变量时，上述推导很容易推广为

$$\boldsymbol{E}\{f(X)g(Y)\} \geq f\left(\frac{\boldsymbol{E}\{X\} \cdot g(\boldsymbol{E}\{XY\}/\boldsymbol{E}\{X\})}{g(\boldsymbol{E}\{Y\})}\right) \cdot g(\boldsymbol{E}\{Y\}). \tag{72}$$

如果 f 和 g 都是凹函数，而不是凸函数，那么不等式方向相反。

例 7。再次考虑具有随机 SNR 的 AWGN 容量示例，$c(Z) = \ln(1 + gZ)$，并假设我们希望界定 $c(Z)$的方差，以评估其波动（例如，为了界定中断概率）。那么，显然，

$$\mathrm{Var}\{c(Z)\} = \boldsymbol{E}\{c^2(Z)\} - [\boldsymbol{E}\{c(Z)\}]^2 = \boldsymbol{E}\{\ln^2(1 + gZ)\} - [\boldsymbol{E}\{\ln(1 + gZ)\}]^2. \tag{73}$$

为了给 $\mathrm{Var}\{c(Z)\}$提供上界，我们可以为 $\boldsymbol{E}\{\ln^2(1 + gZ)\}$ 推导一个上界，并为 $\boldsymbol{E}\{\ln(1 + gZ)\}$推导一个下界。对于后者，例 5 中已经提出了一个下界。对于前者，我们可以使用当前不等式，取 $f(z) = g(z) = \ln(1 + gz)$，很容易证明其满足要求。于是得到如下上界，它仅取决于 Z的前两个矩：

$$\boldsymbol{E}\{\ln^2(1 + gZ)\} \leq \ln(1 + g\boldsymbol{E}\{Z\}) \cdot \ln\left(1 + \frac{g\boldsymbol{E}\{Z\}\ln(1 + g\boldsymbol{E}\{Z^2\}/\boldsymbol{E}\{Z\})}{\ln(1 + g\boldsymbol{E}\{Z\})}\right). \tag{74}$$

有趣的是，函数 $\ln^2(1 + gx)$ 既非凸函数，也非凹函数，但我们的方法给出了一个上界，只要能够计算出 Z 的前两个矩，这个上界就相当容易计算。

参考文献

[1] Cover, T. M.; Thomas, J. A.《信息论要素》，第2版；John Wiley & Sons，新泽西州霍博肯，2006年。

[2] Xiao, L.；Lu, G.《带有信息论应用的詹森不等式新改进》，《Open Mathematics》，第18卷，第1748–1759页，2020年。[3] Deng, Y.；Ullah, H.；Khan, M. A.；Iqbal, S.；Wu, S.《通过主化结果改进詹森不等式及其在信息论中的应用》，《Hindawi Journal of Mathematics》，第2021卷，文章编号1951799，2021年。[4] Wu, S.；Khan, M. A.；Saeed, T.；Sayed, Z. M. M. M.《与任意正有限序列相关的詹森不等式改进形式》，《Mathematics》，第10卷，4817，2022年。[5]Sayyari, Y.；Barsam, H,；Sattarzadeh, A. R.《利用一致凸函数对詹森不等式的新改进及其应用》，《Applicable Analysis》，DOI：10.1080/00036811.2023.2171873[6] Jaafari, E.；Asgari, M. S.；Hosseini, M. S.；Moosavi, B.《关于詹森不等式及其变体》，《AIMS Mathematics》，第5卷，第2期，第1177–1185页，2020年。[7] Matkovi´c, A.；Pe˘cari´c, J.《多变量凸函数的詹森不等式一种变体》，《Journal of Mathematical Inequalities》，第1卷，第1期，第45–51页，2007年。[8] Bakula, M. K.；Matkovi´c, A.；Pe˘cari´c, J.《关于非递减增量函数的詹森不等式一种变体》，《J. Korean Math. Soc.》，第45卷，第3期，第821–834页，2008年。[9] Seuret, A.；Gouaisbaut, F.《利用Wirtinger不等式缩小詹森不等式的差距》，提交至《Automatica》的预印本，2012年7月12日。在线获取地址：https://www.semanticscholar.org/paper/Reducing-the-gap-of-Jensen’s-inequality

-通过使用- Seur

[10] Walker, S. G.；《詹森不等式的一个下界》，《SIAM 数学分析杂志》，第 46 卷，第 5 期，第 3151–3157 页，2014 年。[11] Liao, J. G.；Berg, A.《詹森不等式的强化》，《美国统计学家》，第 73 卷，第 3 期，第 278–281 页，2019 年。[12] Simi´c, S.；Almohsen, B.《詹森不等式的一些推广》，《当代数学》，第 2 卷，第 1 期，2021年。DOI：https://doi.org/10.37256/cm.212021686

[13] Jebara, T.; Pentland,A.；“关于反转詹森不等式”，第13届神经信息处理系统国际会议（NIPS 2000）论文集，第213–219页，丹佛，科罗拉多州，美国，2000年1月。[14] Budimir,I.;Dragomir,S.S.; Pe˘cari´c,J. “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”，纯粹与应用数学不等式杂志，2001年，2，1–14。[15] Simi´c, S.“关于詹森不等式的一个上界”，纯粹与应用数学不等式杂志，第10卷，第2期，文章60，2009年。[16] Simi´c, S.“关于詹森不等式的一个新逆命题”，数学研究所出版物，2009年，85(99)，107–110。[17] Dragomir, S.S.“希尔伯特空间中自伴算子函数的詹森不等式的一些逆向形式”，不等式与应用杂志，2010年，文章ID496821，1–15。[18] Dragomir,S.S. “詹森不等式的一些逆向形式及其应用”，澳大利亚数学学会公报，2013年，87，177–194。[19] Khan, S.;Khan,M.A.; Chu,Y-M.“通过格林函数得到的詹森不等式新逆命题及其应用”，西班牙皇家精确、物理与自然科学学院杂志：A辑，数学，2020年7月，114(3)[20] Khan, S.;Khan,M.A.; Chu, Y-M. “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”，应用科学数学方法，2020年，43，2577–2587。[21]Wunder, G.; Groβ, B.; Fritschek,R.; Schaefer, R.F. “一个反向詹森不等式结果及其在互信息估计中的应用”，2021IEEE信息论研讨会（ITW 2021）论文集，金泽，2021年10月，日本。另可在线获取：https://arxiv.org/pdf/2111.06676.pdf[22] Ali,M.A.;BudakH.;ZhangZ. “量子可微凸函数的量子辛普森不等式和量子牛顿不等式的新扩展”，应用科学数学方法，2021年，https://doi.org/10.1002/mma.7889

[23] Budak H.; Ali, M. A.; Tarhanaci M. “关于协调凸函数的一些新的量子 Hermite-Hadamard 型不等式”，《优化理论与应用杂志》，186(3)，2020年，899–910。[24] N. Merhav，“用于信息论分析的詹森不等式反向形式”，《Information》，2022年，13(1)，39；https://doi.org/10.3390/info13010039 - 2022年1月13日。[25] Krichevsky, R.E.; Trofimov, V.K. “通用编码的性能”，《IEEE 信息论汇刊》，1981年，27(2)，199-207。[26] Merhav, N.; Cohen, A. “具有异步分散式暴力攻击应用的通用随机猜测”，《IEEE 信息论汇刊》，2020年，66(1)，114–129。[27] Dong, A.; Zhang, H.; Wu, D.; Yuan, D. “无线通信性能评估中指数随机变量之和的对数期望”，《2015年 IEEE 第82届车辆技术会议（VTC2015-Fall）论文集》，美国马萨诸塞州波士顿，2015年9月。[28] Tse, D.; Viswanath, P.《无线通信基础》，剑桥大学出版社，2005年。