Some Families of Jensen-Like Inequalities with Application to Information Theory

Neri Merhav

The Viterbi Faculty of Electrical & Computer Engineering
Technion—Israel Institute of Technology
Haifa, Israel

ITA 2024, February 18-23, 2024

Jensen's Inequality

- Extremely useful fundamental and important in many scientific fields, including IT.
- Includes as special cases many famous inequalities in general:
    - the Shwartz-Cauchy inequality, which in turn supports uncertainty relations, including the CRB)
    - the Lyapunov inequality
    - the Hölder inequality
    - the harmonic-geometric-arithmetic means inequalities.
- In information theory, it is stands at the basis of:
    - The information inequality, $D(P\|Q) \geq 0$
    - The data processing inequality and Fano's inequality
    - Conditioning reduces entropy
    - Derivation of single-letter expressions
    - Maximum entropy under moment constraints

Refinemens, Improvements, Variations and Extensions

- Refinements: Xiao & Lu (2020); Deng *et al.* (2021); Wu *et al.* (2022); Sayyari *et al.* (2023).
- Improvements: Seuret *et al.* (2012); Walker (2014); Liao & Berg (2019).
- Variations: Jaafari *et al.* (2020); Matković *et al.* (2007); Bakula *et al.* (2008).
- Extensions: Simić (2021).

There have also been research efforts to derive "reversed" Jensen inequalities:

- Mixtures of exponential families: Jebara & Pentland (2000).
- Global bounds: Budimir *et al.* (2001); Simić (2009); Dragomir (2013).
- Functions of self-adjoint operators: Dragomir (2010).
- Bounds via Green functions: Khan *et al.* (2020).
- Bounds via Chebychev and Chernoff bounds: Wunder *et al.* (2021); Merhav (2022).
- Other: Ali *et al.* (2022); Budak & Ali (2020).

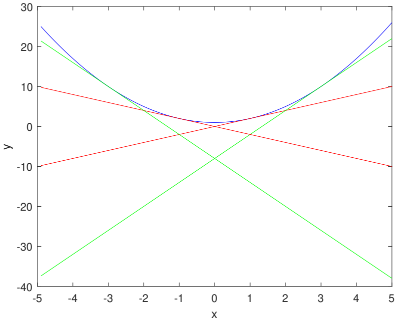
Our Approach to Jensen-Type Inequalities

In the proof of Jensen's inequality,

$$\mathbf{E}\{f(X)\} \geq \sup_a \mathbf{E}\{f(a) + f'(a)(X - a)\} = f(\mathbf{E}\{X\}) \text{ attained by } a^* = \mathbf{E}\{X\}$$

But what if f is just part of a more complicated expression, e.g., $\mathbf{E}\{f(X)g(X)\}$, $\mathbf{E}\{g[f(X)]\}$, $\mathbf{E}\{h[f(X)] \cdot g(X)\}$, etc.?

The optimal value of a is generally different.



Just A Few Examples of Jensen-Like Inequalities

$$\mathbf{E}\{-X\ln X\} \geq -\mathbf{E}\{X\}\cdot\ln(\mathbf{E}\{X\}) - \mathbf{E}\{X\}\cdot\ln\left(1+\frac{\mathsf{Var}\{X\}}{\mathbf{E}^2\{X\}}\right)$$

$$\mathbf{E}\{X^s\} \geq \mathbf{E}^s\{X\}\cdot\left(1+\frac{\mathsf{Var}\{X\}}{\mathbf{E}^2\{X\}}\right)^{s-1} \qquad s\notin(1,2)$$

$$\mathbf{E}\{\ln^2(1+X)\} \leq \ln(1+\mathbf{E}\{X\})\cdot\ln\left(1+\frac{\mathbf{E}\{X\}\ln(1+\mathbf{E}\{X^2\}/\mathbf{E}\{X\})}{\ln(1+\mathbf{E}\{X\})}\right).$$

- Bounds in terms of: (i) first two moments, and (ii) MGF and its derivative.
- In many cases, easy to optimize in closed-form.
- Reverse Jensen inequalities.
- Bounds for functions that are neither convex nor concave.
- Extend easily to multivariate convex functions.
- Applicable to many information-theoretic analyses.

Bounds on $\mathbf{E}\{$convex function $\cdot$ non-negative function$\}$

Let f be convex and g be non-negative:

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

The tightest (maximum) bound is obtained for

$$a = a_* \triangleq \frac{\mathbf{E}\{Xg(Y)\}}{\mathbf{E}\{g(Y)\}},$$

which yields

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

$\mathbf{E}\{$convex function $\cdot$ non-negative function$\}$ (Cont'd)

Example 1 Let $f(x) = -\ln x$ and $g(x) = x$, $x > 0$:

$$\begin{aligned}\mathbf{E}\{-X\ln X\} &\geq -\mathbf{E}\{X\}\cdot\ln\frac{\mathbf{E}\{X^2\}}{\mathbf{E}\{X\}} \\ &= -\mathbf{E}\{X\}\cdot\ln(\mathbf{E}\{X\}) - \mathbf{E}\{X\}\cdot\ln\left(1+\frac{\mathsf{Var}\{X\}}{[\mathbf{E}\{X\}]^2}\right).\end{aligned}$$

Applicable to $\mathbf{E}\{$empirical entropy$\}$ of a sequence drawn by a memoryless source:
Let $X = N(u)/N$, with $N(u)$ =number of occurrences of a letter u in a randomly drawn N-tuple from a DMS P.
$N(u) =$ binomial RV with N trials and probability of success, $P(u)$:

$$\mathbf{E}\{X\} = P(u); \quad \mathsf{Var}\{X\} = \frac{P(u)[1-P(u)]}{N}.$$

$$H \geq \mathbf{E}\{\hat{H}\} \geq H - \frac{|\mathcal{U}|-1}{N}.$$

$\mathbf{E}\{$convex function $\cdot$ non-negative function$\}$ (Cont'd)

Example 2 Let $s, t \in \mathbb{R}$ with $s-t$ either negative or larger than unity. Let $g(x) = x^t$, and $f(x) = x^{s-t}$. Then,

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

In particular, for $t = 1$ and $s \notin (1, 2)$, this becomes

$$\mathbf{E}\{X^s\} \geq \frac{(\mathbf{E}\{X^2\})^{s-1}}{(\mathbf{E}\{X\})^{s-2}} = [\mathbf{E}\{X\}]^s \cdot \left(1 + \frac{\text{Var}\{X\}}{[\mathbf{E}\{X\}]^2}\right)^{s-1} .$$

For $s \in (0, 1)$, x^s is concave, and so, this is a reversed version of Jensen inequality. For $s \leq 0$ and $s \geq 2$, x^s is convex, so this is an improved Jensen inequality: While $[\mathbf{E}\{X\}]^s$ corresponds to the ordinary Jensen inequality, the second factor expresses the improvement, which depends on $\text{Var}\{X\}/[\mathbf{E}\{X\}]^2$.

$\mathbf{E}\{$convex function $\cdot$ monotone[convex function]$\}$

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

where f and g are convex; h is monotonically non-decreasing and non-negative.
For $h(x) = e^x$, we get a bound that depends on the CGF of X and its derivative:

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

with $\psi(s) = \ln \mathbf{E}\{e^{sX}\}$.
Maximizing w.r.t. b for a given a gives:

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

$\mathbf{E}\{$convex function $\cdot$ monotone[convex function]$\}$ (Cont'd)

Example 3 For $f(x) = -\ln x$ and $g(x) = x \ln x$, we obtain a reversed Jensen-like inequality:

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

Defining $\phi(s) = \mathbf{E}\{e^{sX}\} = e^{\psi(s)}$, we have:

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

$\mathbf{E}\{$convex function $\cdot$ monotone convex function$\}$ (Cont'd)

Particularizing further, let $X = 1 + \sum_{i=1}^{k} Y_i^2$, with $Y_i \sim \mathcal{N}(0, \sigma^2)$, $i = 1, \ldots, k$, being independent RVs, with application to bounding the ergodic capacity of the SIMO channel. Here,

$$\begin{aligned} & \mathbf{E}\left\{\ln\left(1 + \sum_{i=1}^{k} Y_i^2\right)\right\} \\ \geq \quad & e \cdot \sup_{\alpha \geq 0}\left\{\frac{\alpha e^{-\alpha}}{(1 + 2\alpha\sigma^2)^{k/2}}\left(1 + \frac{k\sigma^2}{1 + 2\alpha\sigma^2}\right)\ln\left(1 + \frac{k\sigma^2}{1 + 2\alpha\sigma^2}\right)\right\}. \end{aligned}$$

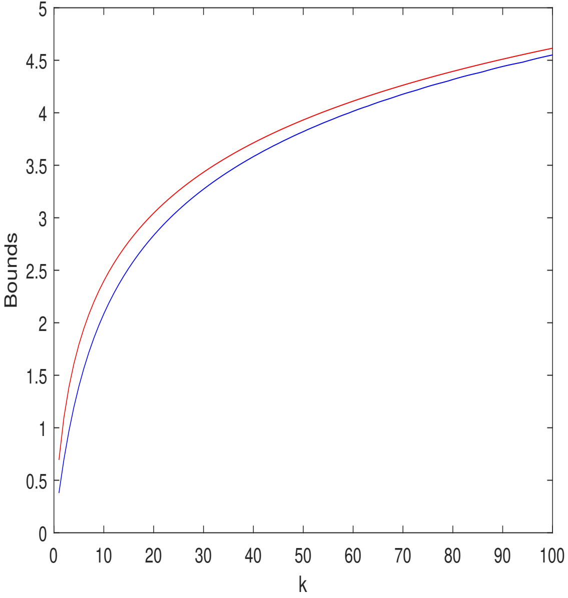


Figure: Bounds on $\mathbf{E}\left\{\ln\left(1+\sum_{i=1}^{k} Y_i^2\right)\right\}$, where $Y_i \sim \mathcal{N}(0,1)$ and $k = 1, 2, \ldots, 100$. Red curve = upper bound, $\ln(1+k\sigma^2)$ - the ordinary Jensen inequality. Blue curve = lower bound.

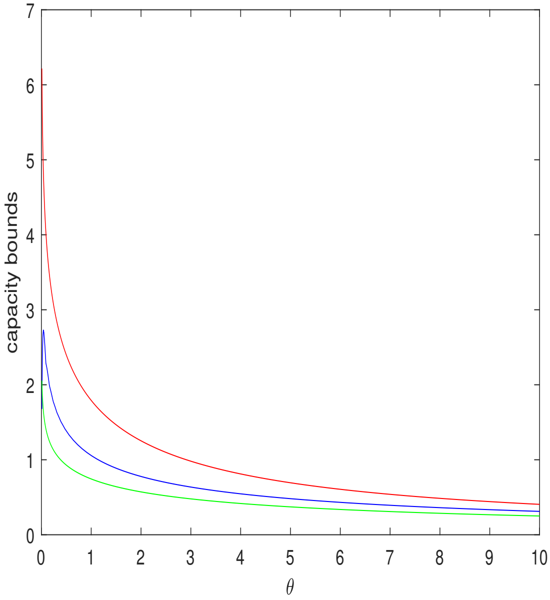


Figure: Upper and lower bounds on $\mathbf{E}\{\ln(1+gZ)\}$, where $Z \sim \theta e^{-\theta z} u(z)$ and $g = 5$. Upper bound $= \ln(1+5/\theta)$ (ordinary Jensen), lower bound = reverse Jensen-like inequality, lower bound = reverse Jensen inequality (Merhav 2022).

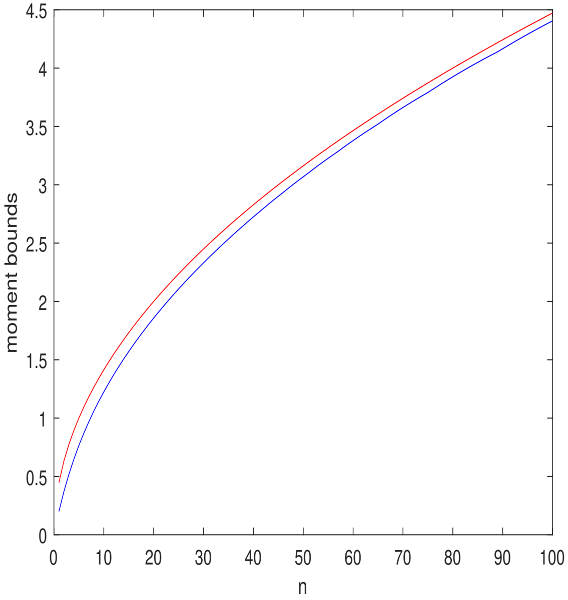


Figure: Upper and lower bounds on $\mathbf{E}\{\sqrt{\sum_{i=1}^{n} Y_i}\}$, where $Y_i \sim$ Bernoulli(0.2). Upper bound $= \sqrt{0.2n}$ (ordinary Jensen), lower bound $=$ reverse Jensen-like inequality.

$\mathbf{E}\{$Product of Two Non-negative Convex Functions$\}$

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

After optimizing a, b and c, we get:

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

More generally,

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

Example 4. The capacity of the AWGN with a random SNR, $c(Z) = \ln(1 + gZ)$. We wish to bound $\mathrm{Var}\{c(Z)\}$.

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

To upper bound $\mathrm{Var}\{c(Z)\}$, we may derive an upper bound to $\mathbf{E}\{\ln^2(1 + gZ)\}$ and a lower bound to $\mathbf{E}\{\ln(1 + gZ)\}$.
For the latter we can use earlier results.
For the former, let $f(z) = g(z) = \ln(1 + gz)$.

$$\mathbf{E}\{\ln^2(1 + gZ)\} \le \ln(1 + g\mathbf{E}\{Z\}) \cdot \ln\left(1 + \frac{g\mathbf{E}\{Z\}\ln(1 + g\mathbf{E}\{Z^2\}/\mathbf{E}\{Z\})}{\ln(1 + g\mathbf{E}\{Z\})}\right).$$

The function $\ln^2(1 + gx)$ is neither convex nor concave, yet our approach offers an upper bound, which is fairly easy to calculate provided that one can compute the first two moments of Z.

Summary and Conclusion

- Often, the optimal value of the parameter(s) can be found in closed form.
- Two types of bounds: (i) bounds that depend on the first two moments, and (ii) bounds that depend on CGF and its derivative.
- Most of our derivations extend to multivariate functions.
- Allowing flexibility to obtain bounds on functions that are not necessarily convex and reverse Jensen inequalities.
- We demonstrate the utility in examples of information-theoretic relevance.
- The bounds become tighter as X becomes concentrated around its mean.