Reversing Jensen's Inequality for Information-Theoretic Analyses

Neri Merhav

The Viterbi Faculty of Electrical and Computer Engineering

Technion – Israel Institute of Technology

Haifa 3200003, Israel

ISIT 2022, Aalto University, Espoo, Finland, June–July 2022

Question:

How often Does Jensen's inequality work in the direction that is opposite to the one you wish?

My Answer:

Much more often than not..

Reverse Jensen's Inequalities

- ♣ Jebara & Pentland (2000).
- ♣ Budimir, Dragomir, & Pečarić (2001).
- ♣ Simić (2009).
- ♣ Dragomir (2010, 2013).
- ♣ Khan, Khan & Chu (2020).
- ♣ Wunder, Groβ, Fritschek, & Schaefer (2021).
- ♣ Ali, Budak & Zhang (2021).
- ♣ Budak, Ali & Tarhanaci (2021).

In most of these works, the bounds depend merely on global properties of the given convex/concave function.

Reverse Jensen's Inequalities

♣ Jebara & Pentland (2000).

♣ Budimir, Dragomir, & Pečarić (2001).

♣ Simić (2009).

♣ Dragomir (2010, 2013).

♣ Khan, Khan & Chu (2020).

♡ Wunder, Groβ, Fritschek, & Schaefer (2021).

♣ Ali, Budak & Zhang (2021).

♣ Budak, Ali & Tarhanaci (2021).

The blue work is an exception in that sense.

Starting Point

Lemma [WGFS21]: Let $f : \mathbb{R}^+ \to \mathbb{R}$ be concave with $f(x) \geq f(0)$ $\forall x \geq 0$. Let $X \geq 0$, $\mathbf{E}\{X\} = \mu$. Then,

$$\mathbf{E}\{f(X)\} \geq \sup_{a>0} \left[\frac{\mu}{a} \cdot f(a) + \left(1 - \frac{\mu}{a}\right) \cdot f(0) - \frac{f(a) - f(0)}{a} \cdot \underbrace{\mathbf{E}\left\{X \cdot \mathcal{I}[X > a]\right\}}_{q(a)} \right].$$

In [WGFS21], $q(a)$ is upper bounded by combining the Markov and Hölder inequalities:

$$q(a) \leq \inf_{p>1} \left\{ \left(\mathbf{E}\{X^p\}\right)^{1/p} \cdot \left(\frac{\mu}{a}\right)^{1-1/p} \right\}.$$

The first factor is weak for large p and the second factor is weak for small p.

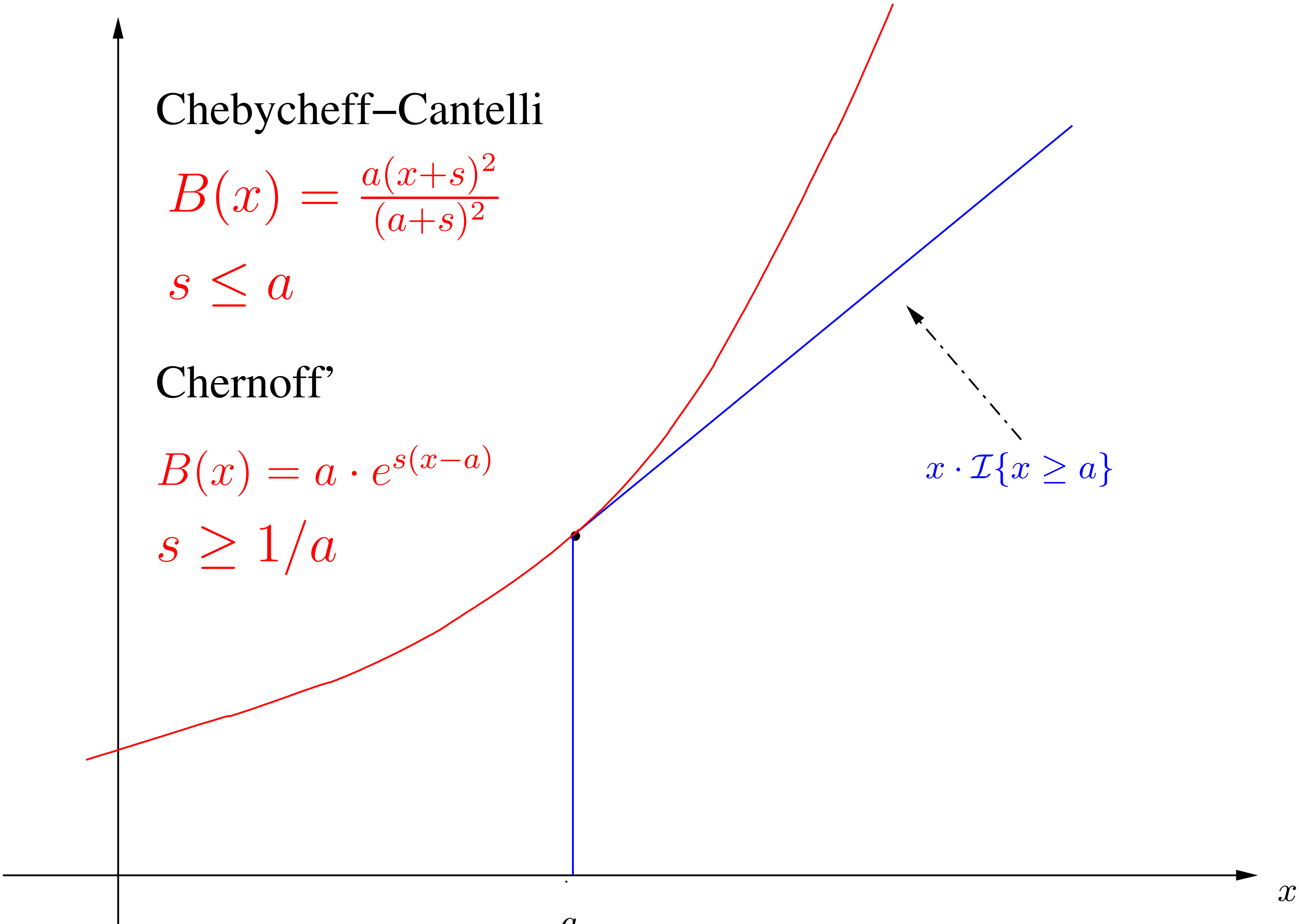
Our Approaches

♠ Chernoff: $\mathbf{E}\{X \cdot \mathcal{I}[X > a]\} \leq \mathbf{E}\{Xe^{s(X-a)}\} = e^{-sa}\frac{\partial}{\partial s}\mathbf{E}\{e^{sX}\}$.

♠ Chernoff': $\mathbf{E}\{X \cdot \mathcal{I}[X > a]\} \leq \mathbf{E}\{ae^{s(X-a)}\} = ae^{-sa}\mathbf{E}\{e^{sX}\}$, $s \geq 1/a$.

♠ Chebychev-Cantelli: $\mathbf{E}\{X \cdot \mathcal{I}[X > a]\} \leq \frac{\mathbf{E}\{a(X+s)^2\}}{(a+s)^2}$, $s \leq a$.

Our Approach (Cont'd)



When X = Sum of I.I.D. RV's

For concave f, the Chernoff approach yields:

$$\mathbf{E}\left\{f\left(\sum_{i=1}^{n} Y_i\right)\right\} \geq \frac{nf(a)}{a}\left[\mu_Y - \inf_{s\geq 0}\left\{e^{-sa}[\Phi_Y(s)]^{n} \cdot \frac{\mathrm{d}\ln\Phi_Y(s)}{\mathrm{d}s}\right\}\right]$$

and we can take $a = n(\mu_Y + \epsilon)$.

The Chebycheff-Cantelli approach gives:

$$\mathbf{E}\left\{f\left(\sum_{i=1}^{n} Y_i\right)\right\} \geq \frac{nf(a)}{a}\left[\mu_Y - \frac{\sigma_Y^2}{\sigma_Y^2 + n\epsilon^2}\right].$$

Example - Gaussian Channel with Random SNR

Consider a complex Gaussian channel whose SNR Z is a RV.

The capacity is given by $C = \mathbf{E}\{\ln(1 + gZ)\}$, where g is a gain factor.

Z is distributed exponentially, i.e.,

$$p_Z(z) = \theta e^{-\theta z}, \quad z \geq 0.$$

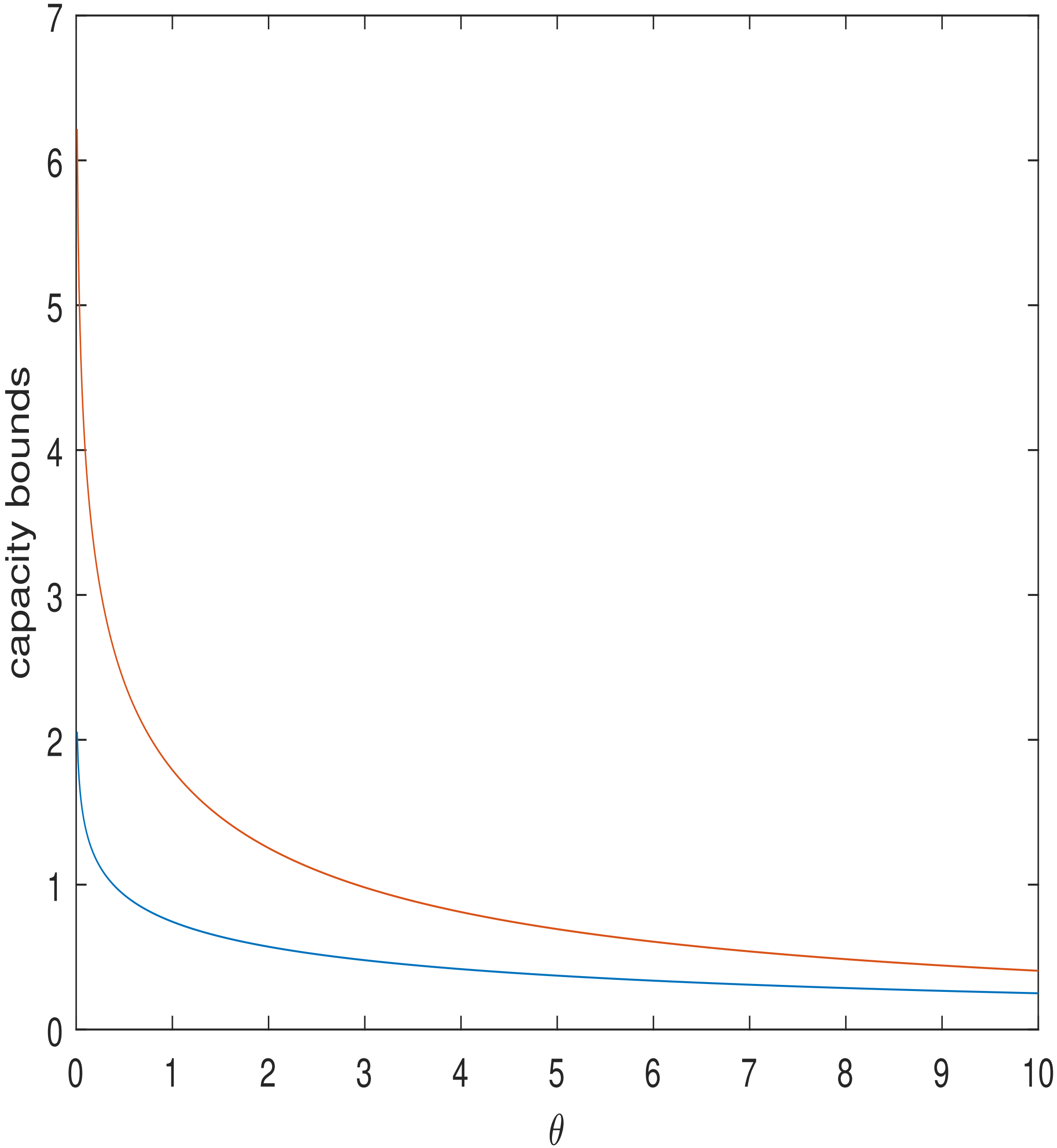
Here,

$$q(a) = \theta \cdot \int_a^\infty z e^{-\theta z} \mathsf{d}z = \left(a + \frac{1}{\theta}\right) \cdot e^{-\theta a}$$

and

$$C \geq \sup_{s \geq 1} \left[\frac{1 - (s+1)e^{-s}}{s}\right] \cdot \ln\left(1 + \frac{gs}{\theta}\right).$$

Example (Cont'd)



More Examples in the Paper

- ♢ Guesswork.
- ♢ Moments of parameter estimation error.
- ♢ Universal source coding.
- ♢ Ergodic capacity of the Rayleight SIMO channel.
- ♢ Differential entropy the generalized multivariate Cauchy distribution.
- ♢ ... and more.

Discussion

- The maximization over a is not eally necessary.
- Assumption $f(x) \geq 0$ can be replaced by $\exists\, \Delta > 0 \;\; f(x) + \Delta x \geq f(0)$.
- Convex functions - by flipping the signs.
- Easy to extend to functions that are neither convex nor concave.
- Extension to multivariate convex/concave functions.

Summary of Main Contributions

- ♣ Significant improvement relative to [WGFS21].
- ♣ Relaxing some assumptions on f.
- ♣ A more natural analogous result for convex functions.
- ♣ Extension to bivariate (and multivariate) functions.
- ♣ Providing examples of information-theoretic relevance.