Information–Theoretic Applications of the Logarithmic Probability Comparison Bound

Rami Atar & Neri Merhav

Department of Electrical Engineering

Technion—Israel Institute of Technology

Haifa 3200004, Israel

ISIT 2015, Hong Kong, June 2015.

Background and Motivation

Key idea for lower bounds in IT and in probability theory: change of measures:

- The sphere–packing bound (Csiszár and Körner's book).
- The converse part of the source–coding error exponent (Marton '74).
- Large deviations theory – tilting.

Common recipe:

- Pass from P to Q under which the probability is high.
- The 'cost' of this passage is $D(Q\|P)$.
- Tightest bound – after optimization over Q.

This work extends this idea:

- Probability under Q – not necessarily high.
- The 'cost' of this passage is the Rényi divergence $D_\alpha(Q\|P)$.
- An extra degree of freedom for optimization – the choice of α.
- Both upper bounds and lower bounds.

The Log–Probability Comparison Bound (LPCB)

Atar, Chowdhary & Dupuis ('15): For $\alpha > 1$,

$$\frac{\ln Q(\mathcal{A})}{\alpha - 1} \leq \frac{\ln P(\mathcal{A})}{\alpha} + D_\alpha(Q \| P),$$

where

$$D_\alpha(Q \| P) = \frac{1}{\alpha(\alpha - 1)} \ln \left[\mathbf{E}_P \left\{ \left(\frac{\mathrm{d}Q}{\mathrm{d}P} \right)^\alpha \right\} \right].$$

More generally, for a given RV $Z \geq 0$:

$$\frac{\ln \mathbf{E}_Q \left\{ Z^{\alpha-1} \right\}}{\alpha - 1} \leq \frac{\ln \mathbf{E}_P \left\{ Z^\alpha \right\}}{\alpha} + D_\alpha(Q \| P),$$

Main tool of the proof: Hölder's inequality.

Objective: to demonstrate the usefulness for upper/lower bounds in IT.
In most of the examples, there are no competing bounds to the BoOK.

Bounds on Exponents

For a sequence of events $\{\mathcal{A}_n\}_{n\geq 1}$, assume that the limits

$$E_P = -\lim_{n\to\infty} \frac{\ln P_n(\mathcal{A}_n)}{n}, \quad E_Q = -\lim_{n\to\infty} \frac{\ln Q_n(\mathcal{A}_n)}{n},$$

$$\bar{D}_\alpha(P\|Q) = \lim_{n\to\infty} \frac{D_\alpha(P_n\|Q_n)}{n}, \quad \bar{D}_\alpha(Q\|P) = \lim_{n\to\infty} \frac{D_\alpha(Q_n\|P_n)}{n}$$

all exist. Then,

$$\begin{aligned} E_P &\geq \frac{\alpha-1}{\alpha} E_Q - (\alpha-1)\bar{D}_\alpha(P\|Q) \\ E_P &\leq \frac{\alpha}{\alpha-1} E_Q + \alpha \bar{D}_\alpha(Q\|P) \end{aligned}$$

Useful when easy to evaluate/bound E_Q and $\bar{D}_\alpha(\cdot\|\cdot)$.

Common practice in IT: For the upper bound, find Q with $E_Q = 0$, and then

$$E_P \leq \inf_{\alpha\geq 1} \alpha \bar{D}_\alpha(Q\|P) = D(Q\|P).$$

Here, by allowing a general Q, we also have α as an extra degree of freedom.

Auxilary Result #1: Small Perturbations

Let P and Q be 'close' in the sense that

$$P(x) = Q(x)[1 + \epsilon(x)] \qquad \forall x \in \mathcal{X}$$

Let $\quad \overline{\epsilon^2} = \sum_x Q(x)\epsilon^2(x), \qquad \epsilon_{\max} = \max_x \epsilon(x).$

Then, for every given sequence of events $\{\mathcal{A}_n\}$:

$$E_P \leq \left(\sqrt{E_Q} + \sqrt{\frac{\overline{\epsilon^2}}{2}}\right)^2 + o(\epsilon_{\max}^2).$$

Applicable to error bounds for very noisy channels.

Comment: For a parametric family $\{P_\theta,\ \theta \in \Theta\}$:

$$\lim_{\theta' \to \theta} \frac{D_\alpha(P_\theta \| P_{\theta'})}{(\theta' - \theta)^2} = \frac{J(\theta)}{2} \quad J(\theta) = \text{Fisher info.}$$

Replace $\sqrt{\overline{\epsilon^2}/2}$ above by $\sqrt{J(\theta)/2} \cdot |\theta' - \theta|$.

Auxiliary Result #2: Iterated Use of the LPCB

Sometimes it proves convenient to pass from P to Q via a third measure S:

$$\begin{aligned} E_P &\geq \frac{\alpha-1}{\alpha} E_S - (\alpha-1)\bar{D}_\alpha(P\|S) \\ E_S &\geq \frac{\beta-1}{\beta} E_Q - (\beta-1)\bar{D}_\beta(S\|Q) \end{aligned}$$

Thus,

$$E_P \geq \frac{(\alpha-1)(\beta-1)}{\alpha\beta} E_Q - \frac{(\alpha-1)(\beta-1)}{\alpha} \bar{D}_\beta(S\|Q) - (\alpha-1)\bar{D}_\alpha(P\|S).$$

where $\alpha, \beta > 1$, Q and S are subject to optimization.

- Straightforward extension to any number of steps.
- Similar idea works for the upper bound on E_P.

Application to Channel Coding

Setup:

- Real channel: $P(\boldsymbol{y}|\boldsymbol{x}) = \prod_i p(y_i|x_i)$.
- Reference channel: $Q(\boldsymbol{y}|\boldsymbol{x}) = \prod_i q(y_i|x_i)$.
- Codebook: $\mathcal{C}_n = \{\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{M-1}\} \subseteq \mathcal{X}^n$, $M = e^{nR}$.
- A message m – picked uniformly at random among M messages.
- m is mapped to $\boldsymbol{x}_m \in \mathcal{C}_n$ and transmitted.
- ML decoding w.r.t. P.
- The error event $\mathcal{E}_n = \{\hat{m} \neq m\}$.

Example # 1: Channel with Interference

Channel P:

$$Y_t = X_t + g_t(X^n, Y^{t-1}) + W_t \qquad W_t \sim \mathcal{N}(0, \sigma^2) \text{ i.i.d.}$$

Channel Q:

$$Y_t = X_t + W_t \qquad W_t \text{ same}$$

Theorem: Assume that $|g_t| \le \Gamma_t$ with $\sum_t \Gamma_t^2 \le n\Gamma^2$. Let $E_U(R,Q)$ be any upper bound on the error exponent of Q (e.g., SP or SL bound). Then

$$E(R,P) \le \left(\sqrt{E_U(R,Q)} + \frac{\Gamma}{\sqrt{2}\sigma} \right)^2 .$$

Comments:

- Similar to the small–perturbation result, but here there is no limitation.
- The bound does not vanish above capacity: alleviated by SL at $(C,0)$.
- For $R = 0$, $E_U(0,Q)$ can be taken to be the zero–rate expurgated bound.
- For $R > R_{\text{crit}}$, $E_U(R,Q) = E_{\text{sp}}(R)$.
- In between, take the SL bound.

Example #1 (Cont'd): Very Noisy Channel

Considering the input constraint $\sum_t X_t^2 \leq nS$, the exact reliability function $E(Q,R)$ is known for the very noisy channel

$$E(R,Q) = \begin{cases} C_Q/2 - R & R < C_Q/4 \\ (\sqrt{C_Q} - \sqrt{R})^2 & C_Q/4 \leq R < C_Q \\ 0 & R < C_Q \end{cases}$$

where $C_Q = S/2\sigma^2 \ll 1$. Accodingly,

$$E(R,P) \leq \begin{cases} \left[\sqrt{\frac{C_Q}{2} - R} + \frac{\Gamma}{\sqrt{2}\sigma}\right]^2 & R < C_Q/4 \\ (\sqrt{C} - \sqrt{R})^2 & C_Q/4 \leq R < C_Q \\ \frac{\Gamma^2(C-R)}{2\sigma^2(C-C_Q)} & C_Q < R < C \\ 0 & R > C \end{cases}$$

where $C = (\sqrt{S} + \Gamma)^2/2\sigma^2$.

At least in the range $[C_Q/4, C_Q)$, the bound is tight in the sense that it is achieved when $g_t \propto x_t$ – coherent sum.

Example #1 (Cont'd): A Lower Bound

The following lower bound is achieved by random coding and ML decoding that ignores the interference (a-fortiori, by ML decoding):

$$E(R,P) \geq \begin{cases} \left(\sqrt{E(R,Q)} - \frac{\Gamma}{\sqrt{2}\sigma}\right)^2 & E(R,Q) \geq \frac{\Gamma^2}{2\sigma^2} \\ 0 & \text{elsewhere} \end{cases}$$

where $E(R,Q)$ is the random coding error exponent of Q.

The bound is attained by an anti–coherent interference $g_t \propto -x_t$.

Implication on robust decoding:

$E(R,P,d)$ – random coding error exponent for decoding metric d.
$\mathcal{P}$ – class of all Gaussian channels with $|g_t|$ bounded as above.

$$\sup_d \inf_{P\in\mathcal{P}} E(R,P,d) \quad \leq \quad E(R,Q)$$

$$\sup_d \inf_{P\in\mathcal{P}} E(R,P,d) \quad \geq \quad \left(\sqrt{E(R,Q)} - \frac{\Gamma}{\sqrt{2}\sigma}\right)^2$$

Example #1 (Cont'd): Non–Gaussian Noise

Channel P:

$$Y_t = X_t + g_t(X^n, Y^{t-1}) + W_t \qquad W_t \text{ i.i.d. non–Gaussian}$$

Channel Q:

$$Y_t = X_t + \tilde{W}_t \qquad \tilde{W}_t \sim \mathcal{N}(0, \sigma^2) \text{ i.i.d.}$$

The Rényi divergence between P and Q may be difficult to handle.
A natural approach is to pass via a third channel "in between", S:

$$Y_t = X_t + g_t(X^n, Y^{t-1}) + \tilde{W}_t \qquad \tilde{W}_t \sim \mathcal{N}(0, \sigma^2) \text{ i.i.d.}$$

and to iterate the LPCB as before using $D_\alpha(P\|S)$ and $D_\beta(S\|Q)$.

- $D_\alpha(P\|S)$ – bounded in terms of $D_\alpha(f_W\|f_{\tilde{W}})$.
- $D_\beta(S\|Q)$ has already been derived (bounded) before.

More details – in the paper.

Example # 2: Gaussian Channel with Fading

Channel P:

$Y_t = (1+\theta_t)X_t + W_t \quad W_t \sim \mathcal{N}(0,\sigma^2)$ i.i.d., θ_t - Gaussian with spectrum $\Sigma(\omega)$

Channel Q:

$$Y_t = X_t + W_t \qquad W_t \sim \mathcal{N}(0,\sigma^2) \text{ i.i.d.}$$

Theorem: Let $|X_t| \leq A$ for all t and let α be small enough that $c(\alpha) = \alpha(\alpha-1)A^2/2\sigma^2$ obeys $2c(\alpha)\sup_\omega \Sigma(\omega) < 1$. Then,

$$\begin{aligned} E(P,R) &\leq \frac{\alpha}{\alpha-1}E(R,Q) - \frac{1}{4\pi(\alpha-1)}\int_0^{2\pi} \ln[1-2c(\alpha)\Sigma(\omega)]\mathrm{d}\omega \\ E(P,R) &\geq \frac{\alpha-1}{\alpha}E(R,Q) + \frac{1}{4\pi\alpha}\int_0^{2\pi} \ln[1-2c(\alpha)\Sigma(\omega)]\mathrm{d}\omega \end{aligned}$$

Comments:

- For certain forms of $\Sigma(\omega)$, the optimization of α can be made explicit.
- Similar bounds for continuous–time Gaussian channels.

Example #3: Rate–Distortion Coding

Consider the source P

$$Y_t = X_t + Z_t, \quad X_t \sim \mathcal{N}(0, \sigma^2), \; Z_t \text{ arbitrary independent process.}$$

The source is compressed at rate R. Find a lower bound on

$$\text{Pr}\left\{ \sum_t (\hat{Y}_t - Y_t)^2 \geq nD \right\}.$$

Under P: $(\boldsymbol{Y}, \boldsymbol{Z}) \sim f_Z(\boldsymbol{z}) g(\boldsymbol{y} - \boldsymbol{z})$, i.e., $Y_t = X_t + Z_t$.

Under Q: $(\boldsymbol{Y}, \boldsymbol{Z}) \sim f_Z(\boldsymbol{z}) g(\boldsymbol{y})$, i.e., $Y_t = X_t$ (hence Z_t is irrelevant).

Example #3 (Cont'd)

We know (from Marton '74) that

$$\liminf_{n\to\infty} \frac{\ln Q\left\{\sum_t (\hat{Y}_t - Y_t)^2 \geq nD\right\}}{n} \geq -\Phi[R - R_{\mathsf{G}}(D)],$$

where

$$R_{\mathsf{G}}(D) = \frac{1}{2}\ln\frac{\sigma^2}{D}; \qquad \Phi(u) = \frac{e^u - 1}{2} - u.$$

Now, assuming that $|Z_t| \leq A$ for all t:

$$\liminf_{n\to\infty} \frac{\ln P\left\{\sum_t (\hat{Y}_t - Y_t)^2 \geq nD\right\}}{n} \geq -\left(\sqrt{\Phi[R - R_{\mathsf{G}}(D)]} + \frac{A}{\sqrt{2}\sigma}\right)^2.$$

Wrapping Up

- We discussed a framework of change–of–measure bounds.
- An extension of a tool already used in IT and large deviations theory.
- We demonstrated the usefulness in various examples.
- Many more examples – in the paper.
- We are not aware of competing bounds in the literature.
- Applicable also to exponential moments in general.