Physics of the Shannon Limits

Neri Merhav

Department of Electrical Engineering

Technion – Israel Institute of Technology

Technion City, Haifa, Israel

ITW 2009, Taormina, Sicily, October 11–16, 2009

Outline

- Laws of physics:

  boundaries between the possible and impossible in Nature.

- Coding theorems $+$ converses in IT:

  boundaries between the possible and impossible in Communications.

In this talk we:

- Briefly review basic background in physics.
- Discuss some physical interpretations of the Shannon limits.
- Introduce the broader physical picture – Jarzynski's equality.
- Propose an informational version of Jarzynski's equality.

Background in Statistical Physics

Consider a system with $n >> 1$ particles which can lie in various microstates, $\{\boldsymbol{x} = (x_1, \ldots, x_n)\}$, e.g., a combination of locations, momenta, angular momenta, spins, ...

For every $\boldsymbol{x}$, $\exists$ energy $\mathcal{E}(\boldsymbol{x})$ – Hamiltonian.

Example: For $x_i = (\boldsymbol{p}_i, \boldsymbol{r}_i)$,

$$\mathcal{E}(\boldsymbol{x}) = \sum_{i=1}^{n} \left(\frac{\|\boldsymbol{p}_i\|^2}{2m} + mgh_i \right).$$

Background (Cont'd)

In thermal equilibrium, $\boldsymbol{x} \sim$ Boltzmann–Gibbs distribution:

$$P(\boldsymbol{x}) = \frac{e^{-\beta \mathcal{E}(\boldsymbol{x})}}{Z(\beta)}$$

where $\beta = \frac{1}{kT}$, k – Boltzmann's constant, T – temperature, and

$$Z(\beta) = \sum_{\boldsymbol{x}} e^{-\beta \mathcal{E}(\boldsymbol{x})}, \quad \text{a normalization factor} = \text{partition function}$$

$\phi(\beta) = \ln Z(\beta) \quad \Rightarrow$ many physical quantities:

free energy: $F = -\frac{\phi}{\beta}$;

mean internal energy: $E = -\frac{d\phi}{d\beta}$;

entropy: $S = \phi - \beta \frac{d\phi}{d\beta}$.

Background (Cont'd)

Easy to see that

$$F = E - TS.$$

Physical meaning:

$\Delta F = F_1 - F_0 =$ the minimum work it takes to transfer the system between two equilibrium points, '0' and '1', for fixed T.

Minimum – achieved by a reversible process – so slow that the system is always almost in equilibrium.

The Information Inequality

Essentially all fundamental limits of IT are based on the information inequality in some form (DPT, Fano's inequality, "conditioning reduces entropy," ...).

For any two distributions, P and Q, over an alphabet $\mathcal{X}$:

$$D(P\|Q) \triangleq \sum_x P(x) \log \frac{P(x)}{Q(x)} \geq 0.$$

In physics, it is known as the Gibbs inequality.

The Gibbs Inequality

Let $\mathcal{E}_0(x)$ and $\mathcal{E}_1(x)$ be two Hamiltonians of a system. For a given β, let

$$P_i(x) = \frac{e^{-\beta\mathcal{E}_i(x)}}{Z_i}, \qquad Z_i = \sum_x e^{-\beta\mathcal{E}_i(x)}, \quad i = 0, 1.$$

Then,

$$\begin{aligned} 0 &\leq D(P_0 \| P_1) = \boldsymbol{E}_0 \left\{ \ln \frac{e^{-\beta\mathcal{E}_0(X)}/Z_0}{e^{-\beta\mathcal{E}_1(X)}/Z_1} \right\} \\ &= \ln Z_1 - \ln Z_0 + \beta \cdot \boldsymbol{E}_0\{\mathcal{E}_1(X) - \mathcal{E}_0(X)\} \end{aligned}$$

or

$$\begin{aligned} \boldsymbol{E}_0\{\mathcal{E}_1(X) - \mathcal{E}_0(X)\} &\geq kT \ln Z_0 - kT \ln Z_1 \\ &= F_1 - F_0 \end{aligned}$$

Interpretation of $\boldsymbol{E}_0\{\mathcal{E}_1(X) - \mathcal{E}_0(X)\} \geq \Delta F$

- A system with Hamiltonian $\mathcal{E}_0(x)$ – in equilibrium $\forall\, t < 0$.
  Free energy $= -kT \ln Z_0$.
- At $t = 0$, the Hamiltonian jumps, by $W = \mathcal{E}_1(x) - \mathcal{E}_0(x)$: from $\mathcal{E}_0(x)$ to $\mathcal{E}_1(x)$ – by abruptly applying a force. Energy injected:
  $\boldsymbol{E}_0\{W\} = \boldsymbol{E}_0\{\mathcal{E}_1(X) - \mathcal{E}_0(X)\}$.
- New system, with Hamiltonian $\mathcal{E}_1$, equilibrates.
  Free energy $= -kT \ln Z_1$.

Gibbs inequality: $\boldsymbol{E}_0\{W\} \geq \Delta F$.

$$\boldsymbol{E}_0\{W\} - \Delta F = kT \cdot D(P_0 \| P_1)$$

is the dissipated energy $=$ entropy production (system $+$ environment) due to irreversibility of the abruptly applied force.

Example – Data Compression and the Ising Model

Let $\boldsymbol{X} \in \{-1,+1\}^n \sim$ Markov chain $P_0(\boldsymbol{x}) = \prod_i P_0(x_i|x_{i-1})$ with

$$P_0(x|x') = \frac{\exp(Jx \cdot x')}{Z_0}, \qquad x, x' \in \{-1,+1\}$$

Code designer thinks that $\boldsymbol{X} \sim$ Markov with parameters:

$$P_1(x|x') = \frac{\exp(Jx \cdot x' + Kx)}{Z_1(x')}.$$

$D(P_0\|P_1) =$ loss in compression due to mismatch. Easy to see that

$$\mathcal{E}_0(\boldsymbol{x}) = -J \cdot \sum_i x_i x_{i-1}; \quad \mathcal{E}_1(\boldsymbol{x}) = -J \cdot \sum_i x_i x_{i-1} - B \cdot \sum_i x_i$$

where

$$B = K + \frac{1}{2} \ln \frac{\cosh(J-K)}{\cosh(J+K)}.$$

Thus, $W = -B \cdot \sum_i x_i$ means an abrupt application of the magnetic field B.

Physics of the Data Processing Theorem (DPT)

DPT – supports virtually all Shannon limits: For $X \to U \to V$:

$$I(X;U) - I(X;V) = \boldsymbol{E}\{D(P_{X|U,V}(\cdot|U,V)\|P_{X|V}(\cdot|V))\} \geq 0.$$

Let $\beta = 1$. Given (u, v), let

$$\mathcal{E}_0(x) = -\ln P(x|u,v) = -\ln P(x|u); \quad \mathcal{E}_1(x) = -\ln P(x|v).$$

$$Z_0 = \sum_x e^{-1\cdot[-\ln P(x|u,v)]} = \sum_x P(x|u,v) = 1$$

and similarly, $Z_1 = 1$. Thus, $F_0 = F_1 = 0$, and so, $\Delta F = 0$.
After averaging over P_{UV}:

$$\begin{aligned} \boldsymbol{E}_0\{W(X)\} &= \boldsymbol{E}\{-\ln P(X|V) + \ln P(X|U)\} \\ &= H(X|V) - H(X|U) = I(X;U) - I(X;V). \end{aligned}$$

$$\boldsymbol{E}_0\{W\} = I(X;U) - I(X;V) \geq 0 = \Delta F.$$

Discussion

The relation

$$\boldsymbol{E}_0\{W\} - \Delta F = kT \cdot D(P_0 \| P_1) \geq 0$$

is known (Jarzynski ‘97, Crooks ‘99, ..., Kawai *et. al.* ‘07), but with different physical interpretations, which require some limitations.

Present interpretation – holds generally; Applied in particular to the DPT.

In our case:

- Maximum irreversibility: $\boldsymbol{E}_0\{W\}$ – fully dissipated: $\Delta F = 0$.
- All dissipation – in the system, none in the environment:

$$\boldsymbol{E}_0\{W\} = T\Delta S = 1 \cdot [H(X|V) - H(X|U)].$$

- Rate loss due to gap between mutual informations: irreversible process $\Longleftrightarrow$ irreversible info: $I(X;U) > I(X;V) \longrightarrow U$ cannot be retrieved from V.

Relation to Jarzynski's Equality

Let

$$\mathcal{E}_\lambda(x) = \mathcal{E}_0(x) + \lambda[\mathcal{E}_1(x) - \mathcal{E}_0(x)]$$

interpolate $\mathcal{E}_0$ and $\mathcal{E}_1$. λ – a generalized force.
Jarzynski's equality (1997): $\forall$ protocol $\{\lambda_t\}$ with $\lambda_t = 0 \; \forall \; t \leq 0$ and $\lambda_t = 1$ $\forall \; t \geq \tau$ ($\tau \geq 0$), the injected energy

$$W = \int_0^\tau \mathrm{d}\lambda_t [\mathcal{E}_1(x_t) - \mathcal{E}_0(x_t)]$$

satisfies

$$\boldsymbol{E}\left\{e^{-\beta W}\right\} = e^{-\beta \Delta F}.$$

Jensen: $\boldsymbol{E}\{e^{-\beta W}\} \geq e^{-\beta \boldsymbol{E}\{W\}}$ so, $\boldsymbol{E}\{W\} \geq \Delta F$ more generally.
Equality – for a reversible process – $W =$ deterministic.

Informational Jarzynski Equality

Taking

$$\mathcal{E}_0(x) = -\ln P_0(x), \quad \mathcal{E}_1(x) = -\ln P_1(x), \quad \beta = 1$$

and defining a “protocol” $0 \equiv \lambda_0 \to \lambda_1 \to \ldots \to \lambda_n \equiv 1$, and

$$W = \sum_{i=0}^{n-1} (\lambda_{i+1} - \lambda_i) \ln \frac{P_0(X_i)}{P_1(X_i)}, \qquad X_i \sim P_{\lambda_i} \propto P_0^{1-\lambda_i} P_1^{\lambda_i},$$

one can show:

$$\boldsymbol{E}\{e^{-W}\} = 1 = e^{-\Delta F}.$$

Jensen: generalized information inequality:

$$\int_0^1 \mathrm{d}\lambda_t \cdot \boldsymbol{E}_{\lambda_t} \left\{ \ln \frac{P_0(X)}{P_1(X)} \right\} \geq 0.$$

Summary

- Suboptimum commun. system $\Longleftrightarrow$ irreversible process.
- Info rate loss $\Longleftrightarrow$ dissipated energy $\rightarrow$ entropy $\uparrow$
- Fundamental limits of IT $\Longleftrightarrow$ second law.
- Possible implications of Jarzynski's equality in IT.