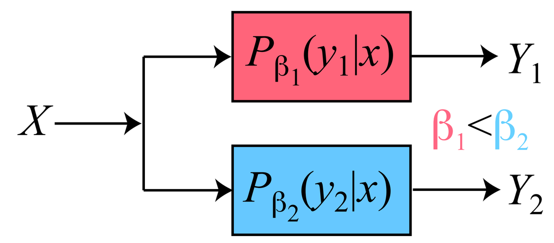
Mutual information and estimation in channels of exponential family type

Todd P. Coleman (UIUC) and **Maxim Raginsky** (Duke)

Mutual information and channel quality



A lot of recent work related to behavior of the mutual information as a function of some "channel quality" parameter. Landmark result:

- **Guo, Shamai, Verdú (2005)**: if $Y = \sqrt{\beta}X + Z$, where $Z \sim \mathcal{N}(0,1)$, then

$$\frac{d}{d\beta} I(X; \sqrt{\beta}X + Z) = \frac{1}{2}\mathbf{E}\left\{\left(X - \mathbf{E}\left\{X \middle| \sqrt{\beta}X + Z\right\}\right)^2\right\}$$

Immediately implies that $I'(\beta) \ge 0$ — mutual information is a monotonically increasing function of the channel quality parameter β.

However, the derivation of this formula relies on the fact that the channel with lower β is a degraded version of the channel with a higher β — Markov chain condition:

$$X \to Y \to Y' \qquad \Leftrightarrow \qquad Y = \sqrt{\beta}X + Z, Y' = \sqrt{\beta'}X + Z', \beta' < \beta$$

Our goal: find conditions for monotonicity of mutual information that apply to channel models for which such a Markov condition does not necessarily hold.

Channels of exponential family type

Many important noisy channel models have can be cast in the exponential family form

$$P_\beta(y|x) = \frac{e^{-\beta\rho(x-y)}}{Z(\beta)}$$

- $\beta > 0$ is the channel quality parameter
- $Z(\beta) \triangleq \sum_y e^{-\beta\rho(x-y)}$ (assumed to be independent of x) is the *partition function*

(This formalism is for DMC's; extension to continuous alphabets is analogous.)

Examples:

- Gaussian channel —
$$p_\beta(y|x) = \frac{e^{-\beta(x-y)^2}}{\sqrt{\pi/\beta}}$$
- Binary Symmetric Channel —
$$p_\beta(y|x) = \frac{e^{-\beta(x \ominus y)}}{1 + e^{-\beta}}$$
($\beta = 0$: crossover probability $= 1/2$; $\beta \to \infty$: crossover probability $\to 0$)
- Exponential Server Timing Channel (ESTC)
$$p^T_\beta(y|x, q_0) = \beta^{y_T} \exp\left\{\int_0^T -\beta\rho(q_0 + x_t - y_t)dt\right\}$$

Implications

- New results on broadcast and secrecy capacity without relying on explicit degradation assumptions.
- New results on mutual information and estimation beyond the AWGN channel and squared error criterion.

Rate-distortion perspective

Given input distribution P_X, define:

- P^β_Y: marginal distribution of the channel output
- $Q_\beta(x|y)$: backward channel, given by the posterior

$$Q_\beta(x|y) = \frac{P_X(x)e^{-\beta\rho(x-y)}}{Z(\beta|y)},$$

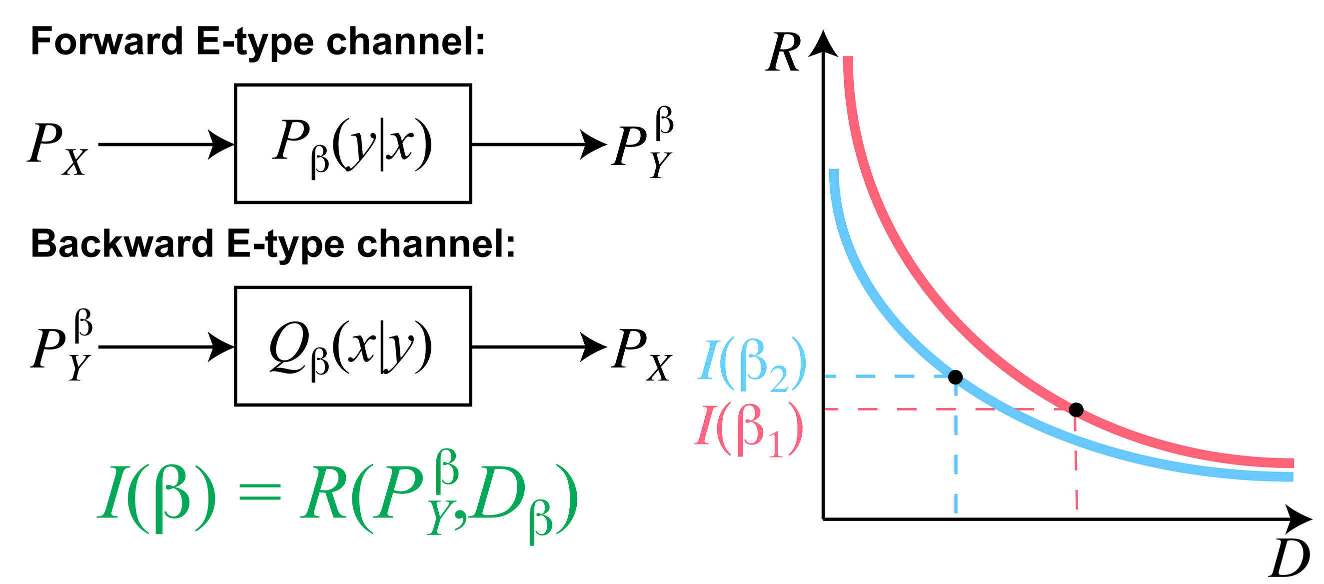
where $Z(\beta|y) \triangleq \sum_x P_X(x) e^{-\beta\rho(x-y)}$

Rate-distortion perspective: view P^β_Y as the source, $\rho(x-y)$ as the distortion function, and

$$D_\beta \triangleq \mathbf{E}_\beta\{\rho(X-Y)\}$$

Then $Q_\beta(x|y)$ satisfies variational conditions to attain rate-distortion function $R(P^\beta_Y, D_\beta)$:

$$\boxed{I(\beta) = I(X;Y) = R(P^\beta_Y, D_\beta)}$$



Area theorems for an *Arbitrary* P_X

- Area Theorem for Posterior Information Gain

$$D\left(P^\beta_{X|Y=y} \| P_X\right) = \left[\int_\beta^\infty \frac{1}{\bar\beta^2} D\left(P^{\bar\beta}_{X|Y=y} \| P_X\right) d\bar\beta\right] - \beta\mathbb{E}_\beta\{\rho(X-Y)|Y=y\} + \beta E_0(y)$$

- Area Theorem for Mutual Information

$$I(P_X;\beta) = \frac{1}{Z(\beta)}\int_\beta^\infty Z(\bar\beta)\mathbb{E}_{\bar\beta}\left\{\rho(X-Y) D\left(P^{\bar\beta}_{X|Y=y} \| P_X\right)\right\} d\bar\beta$$

Towards monotonicity $(\beta_1 > \beta_2)$

1. The following condition can be proven many ways: $\sum_{x,y} P^{\beta_2}_Y(y) Q_{\beta_1}(x|y)\rho(x-y) \le D_{\beta_2}$. This implies $I(\beta_2) = R(P^{\beta_2}_Y, D_{\beta_2}) \le I(P^{\beta_2}_Y, Q_{\beta_1})$.
2. We would like to show that $I(P^{\beta_2}_Y, Q_{\beta_1}) \le I(P^{\beta_1}_Y, Q_{\beta_1}) = I(\beta_1)$. This holds if and only if

$$\mathbf{E}_{\beta_2}\{D(Q_{\beta_1}(\cdot|Y) \| P_X)\} \le \mathbf{E}_{\beta_1}\{D(Q_{\beta_1}(\cdot|Y) \| P_X)\}$$

— thus, we have related monotonicity of mutual information to monotonicity of average information gain due to posterior estimates at β_1 and β_2

Summary

- **Status quo** In a network where quality of communication links may differ depending on location, need to characterize the impact of channel quality on operational characteristics (probability of error, end-to-end distortion, etc.)
  - To quantify the impact, one often needs to assume that the channel family is ordered by degradation
  - Need to check appropriate conditions on a case-by-case basis
- **New insights** Many important channel models have an exponential family structure.
  - Can exploit connections between information theory and statistics. E.g., maximum entropy characterization of exponential families (Kullback, Csiszár); Shannon lower bounds on rate-distortion functions
  - Instead of degradation, exploit monotonicity of posterior information gain (Mitter-Newton, Yuan-Clarke)
- **Achievement** Analysis of dependence of mutual information on channel quality reduces to a rate-constrained estimation problem with distortion (loss) function $\rho(x-y)$
  - **How it works:** structure of E-type channels leads to a dual characterization of mutual information as the minimum rate needed to describe the channel output Y via channel input X under a given constraint on $\mathbf{E}\{\rho(X-Y)\}$
  - **Limitations and assumptions:** for a general E-type channel, can prove monotonicity of mutual information only in the high-SNR (high-β) regime
- **Impact**
  - New results on broadcast and secrecy capacity without relying on explicit degradation assumptions.
  - New results on mutual information and estimation beyond the AWGN channel and squared error criterion.
  - **Next-phase goals** Explore connections between information theory and statistical estimation over E-type channels to obtain new performance results in the network setting.