No input symbol should occur more frequently than $1-\frac{1}{e}$

Gowtham Kumar, Alexandros Manolakos

Electrical Engineering, Stanford University

P.I.: Prof. Thomas Cover

Abstract

Cause and Effect

Goal: Find the fundamental relationship between the input and output distributions at capacity for any discrete memoryless channel.

Question:

Is it possible to find a 2-input Discrete Memoryless Channel (DMC) $P(y|x)$ with capacity $C > 0$ for which $[0.2, 0.8]$ is a capacity achieving distribution?

Answer: No!

Question:

Now consider DMCs of arbitrary input and output alphabet sizes. What is the set of all input distributions $P(x)$ that are capacity achieving for some DMC $P(y|x)$ with strictly positive capacity?

Answer:

Those distributions $P(x)$ for which no symbol occurs more than $1 - \frac{1}{e} = 63.2\%$ of the time. Therefore

$$\max_x P^*(x) < 1 - \frac{1}{e}.$$

The Main Theorem

Theorem 2. *Consider any DMC $(\mathcal{X}, P(y|x), \mathcal{Y})$ with arbitrary but finite input and output alphabet. Let $P^*(x) := \arg\max_{P(x)} I(X;Y)$. Then*

$$\max_{x\in\mathcal{X}} P^*(x) < 1 - \frac{1}{e}.$$

Proof. (Main Idea)

Define a function $f(x)$ on the input

$$f(x) = \mathbf{1}\{x \neq 0\} = \begin{cases} 0 & \text{if } x = 0 \\ 1 & \text{if } x \neq 0 \end{cases}$$

Let $P\{f(x) = 0\} = \alpha$.

$$C = \max_{P(x)} I(X;Y)$$

$$= \max_{P(x,f(x))} [I(f(X);Y) + \bar{\alpha} I(X;Y|f(X) = 1)]$$

- $I(f(X);Y)$ denotes the capacity of a 2-Input DMC.

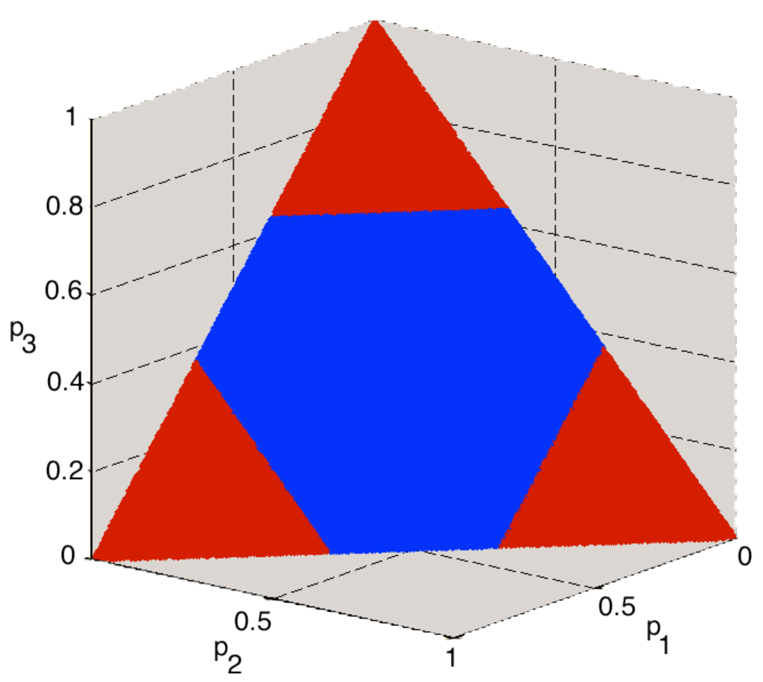


Fig. 3. Region of the Capacity Achieving Input Distributions.

What about the Output Distribution?

There is no similar restriction for the capacity achieving output distribution.

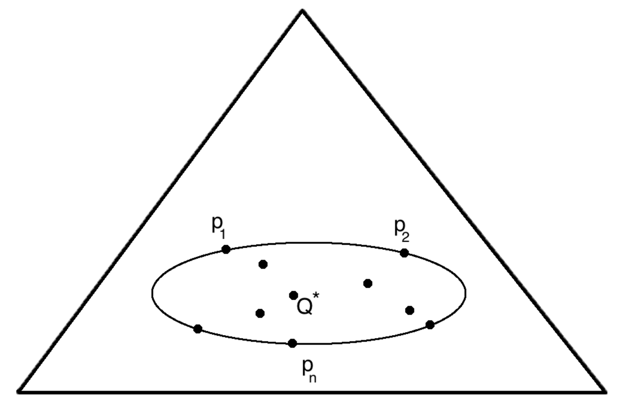
Corollary 3. *Let $Q(x)$ be a discrete probability distribution. We can always construct a DMC $(\mathcal{X}, P(y|x), \mathcal{Y})$ where $Q(x)$ is the capacity achieving output distribution.*

Review: Min-Max Duality

Consider a DMC $(\mathcal{X}, P(y|x), \mathcal{Y})$. The capacity of the channel is given by

$$C = \min_Q \max_{x\in\mathcal{X}} D(P(\cdot|x)||Q(y)).$$

The capacity C is the radius of the smallest "sphere" containing the set of all $P(\cdot|x)$. $Q^*(y)$ is the center of the "information sphere".



Converse

Theorem 3. *Let $P(x)$ be a discrete input probability distribution over a discrete memoryless channel $(\mathcal{X}, P(y|x), \mathcal{Y})$. If $\exists\, S \subset \mathcal{X} : \sum_{x\in S} P(x) \in (\frac{1}{e}, 1 - \frac{1}{e})$ then there exists a channel for which $P(x)$ is a capacity achieving input distribution.*

Proof. (Main Idea)

- Let $p = \sum_{x\in S} P(x)$. Then, $p \in (\frac{1}{e}, 1 - \frac{1}{e})$.
- We can always construct a 2-Input DMC Channel for which the distribution $[p, 1-p]$ is capacity achieving **[?]**.
- We use this channel to construct a $|\mathcal{X}|-$Input DMC that has the $P(x)$ as a capacity achieving distribution.

Extreme Channel

There is no DMC channel with positive capacity that has the distribution $[\frac{1}{\epsilon}, 1 - \frac{1}{\epsilon}]$ as a capacity achieving distribution. However, the channel in Fig. 4 has a capacity achieving distribution that approaches this bound:

$$P(x) = \begin{cases} \frac{1}{e} + \epsilon(\delta) & \text{if } x = 0 \\ 1 - \frac{1}{e} - \epsilon(\delta) & \text{if } x = 1 \end{cases}$$

where $\epsilon(\delta) \to 0, C(\epsilon(\delta)) \to 0$ as $\delta \to 0$.

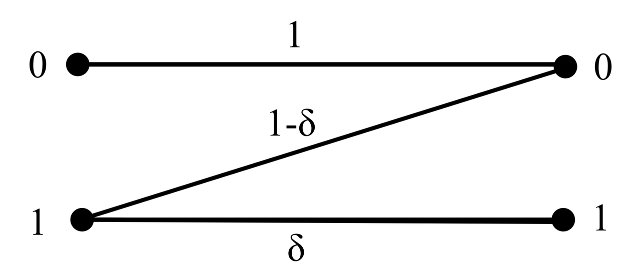
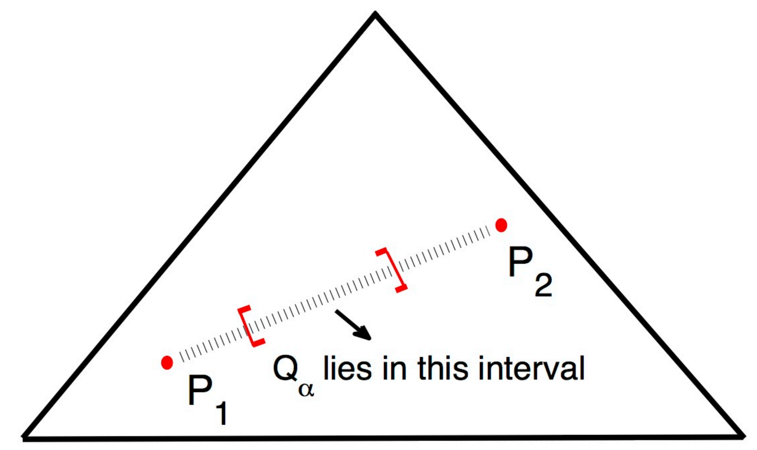


Fig. 4. Extreme Channel.

Binary input DMC

Theorem 1. **(Primal)** *Let $P_1(y)$ and $P_2(y)$ be any two probability distributions on $\mathcal{Y}$. Let $\alpha \in [0,1]$ be chosen such that $D(P_1||Q_\alpha) = D(P_2||Q_\alpha)$ where $Q_\alpha(y) = \alpha P_1(y) + (1-\alpha)P_2(y)$. Then $\alpha \in (\frac{1}{e}, 1 - \frac{1}{e})$, where e is the base of the natural logarithm.*



Corollary 1. **(Dual)** *Let $(\mathcal{X}, P(y|x), \mathcal{Y})$ be a DMC with $P_1(y)$ and $P_2(y)$ as the rows of the channel matrix. Let the capacity achieving distribution be $P^*(x) = [\alpha, 1-\alpha]$. Then α satisfies*

$$\frac{1}{e} < \alpha < 1 - \frac{1}{e}.$$

Information Geometric "Mid-point"

Corollary 2. *Let $P_1(y)$ and $P_2(y)$ be any two probability distributions on $\mathcal{Y}$. Let $P_{\frac{1}{e}}(y) = \frac{1}{e}P_1(y) + (1 - \frac{1}{e})P_2(y)$. Then,*

$$D(P_1||Q_{\frac{1}{e}}) - D(P_2||Q_{\frac{1}{e}}) \geq 0.$$

The farthest that $Q_\alpha(x)$ can be from $P_1(x)$ and still always be closer to $P_1(x)$ than to $P_2(x)$ is for $\alpha = \frac{1}{e}$.

Acknowledgement

The authors wish to thank their advisor Tom Cover for insightful discussions. Also, they would like to acknowledge very helpful conversations with Kartik Venkat, Bobbie Chern and Yeow-Khiang Chia.

References

[1] Kumar, G. and Manolakos A.,"No input symbol should occur more frequently than $1 - \frac{1}{e}$", To be submitted in ISIT.

[2] Silverman, R., "On binary channels and their cascades," Information Theory, IRE Transactions on , vol.1, no.3, pp.19-27, December 1955

[3] T.M. Cover and J.A. Thomas, "Elements of Information Theory". New York: John Wiley, 1991