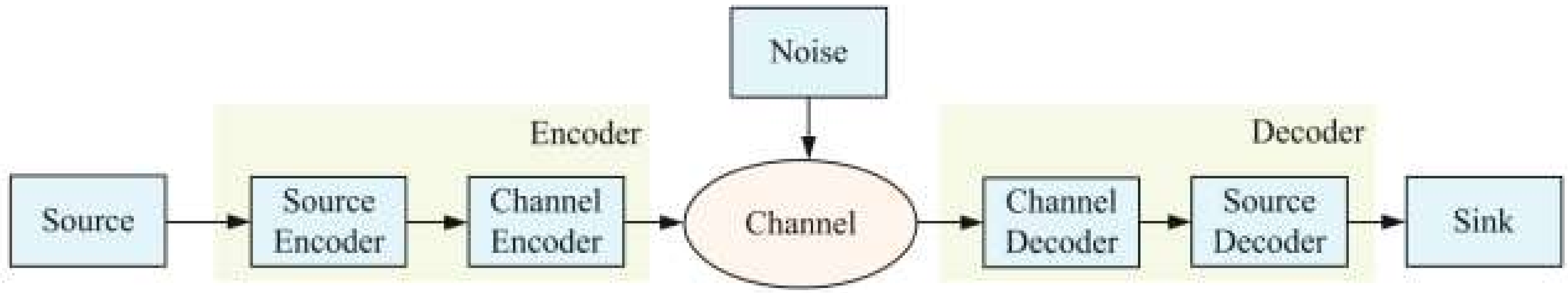
EE185524

Information Theory

Yurid Eka Nugraha
TSP DTE FTEIC ITS

How to transmit information over channel?



Shannon, *A mathematical theory of communication*, 1948

Basic insight:

- Shannon: “the rarer an event is, the more the information”
- “The information content of a message depends on its a priori probability”

(Information) Entropy?

- Simple definition: "the differential of a quantity which depends on the configuration of the system"

- In the context of information theory: "**average level of "information"**, "surprise", or "uncertainty" inherent to the variable's possible outcomes". (Shannon)

$$H(X) = -\sum_{x \in X} p(x) \log p(x) = \mathbb{E}\{-\log p(x)\}$$

(Information) Entropy?

$$H(X) = -\sum_{x \in X} p(x) \log p(x) = \mathbb{E}\{-\log p(x)\}$$

- If log base = 2, then information is expressed as **bits**

- Called information entropy because of its similarity to the entropy in statistical mechanics: "number of possible microscopic states (microstates) of a system in thermodynamic equilibrium, consistent with its macroscopic thermodynamic properties"

Review of basic probability

- $\mathbb{E}(X)$: expected value of random variable X
- Countable events:

  $\mathbb{E}(X) = \sum_{i=1}^{\infty} x_i p_i$,

  where x_i: possible outcome of random variables X with corresponding probabilities p_i
- For 6-side dice: $\mathbb{E}(X) = 3.5$

Source entropy

Requires a way of measuring:

- information content of a source
- efficiency of a code, etc.

- Maximized if all symbols **equiprobable**

Source entropy

- Fair coin toss with $p = \frac{1}{2}$:

$$H(X) = -\frac{1}{2}\log\left(\frac{1}{2}\right) - \frac{1}{2}\log\left(\frac{1}{2}\right) = 1$$

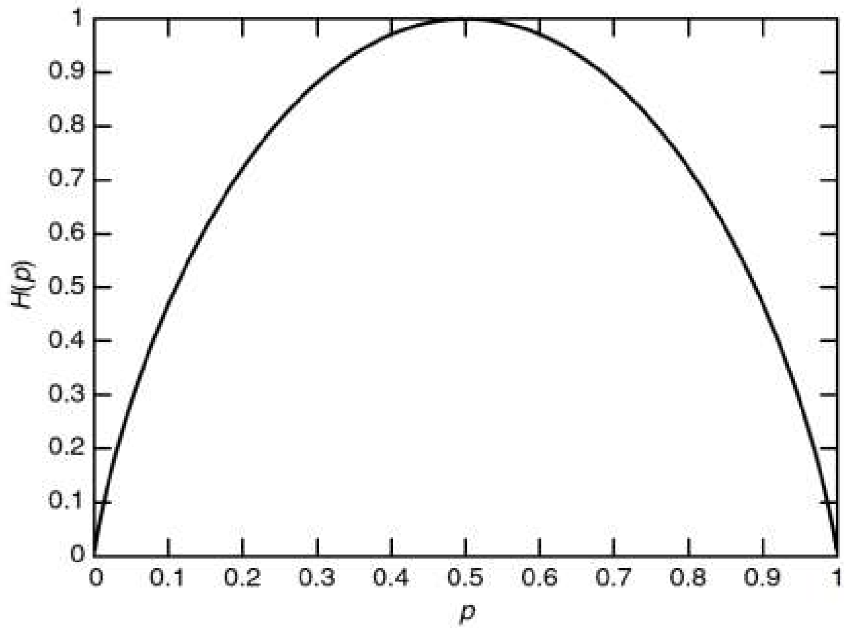
- Coin toss with head probability $p = \frac{3}{4}$:

$$H(X) = -\frac{1}{4}\log\left(\frac{1}{4}\right) - \frac{3}{4}\log\left(\frac{3}{4}\right) = 0.811$$

- Coin toss with head probability $p = \frac{4}{5}$:

$$H(X) = -\frac{1}{5}\log\left(\frac{1}{5}\right) - \frac{4}{5}\log\left(\frac{4}{5}\right) = 0.722$$

Source entropy



Conditional entropy and mutual information

- Conditional entropy $H(X|Y) := \sum_{y \in Y} p(y) H(X|Y = y)$
- $H(X|Y) < H(X)$
  - random variable X carries less information if Y is already known
- Mutual information: A drop in entropy between X and Y: $I(X;Y) \coloneqq H(X) - H(X|Y)$

Mutual information

$$
\begin{aligned}
I(X;Y) &= H(X) - H(X|Y) \\
&= -\sum_{x\in X} p(x)\log p(x) + \sum_{y\in Y, x\in X} p(x,y)\log\frac{p(x,y)}{p(y)} \\
&= \ \mathbb{E}\{\log\tfrac{p(x,y)}{p(x)p(y)}\}
\end{aligned}
$$

Also, $I(X;Y) = I(Y;X)$: mutual information **between** X and Y

Entropy rates

- For stochastic process $\{X_i\}$:

$$H(X) = \lim_{n\to\infty} \left(\frac{1}{n}\right) H(X_1, \dots, X_n)$$

when the limit exists.

- Derivative: $H'(X) = \lim_{n\to\infty} H(X_n | X_{n-1}, X_{n-2}, \dots, X_1)$
- For strongly stationary stochastic process: $H(X) = H'(X)$
- "although there are many series of results that may be produced by a random process, the one actually produced is most probably from a loosely defined set of outcomes that all have approximately the same chance of being the one actually realized"

Topological entropy

- Defined as:

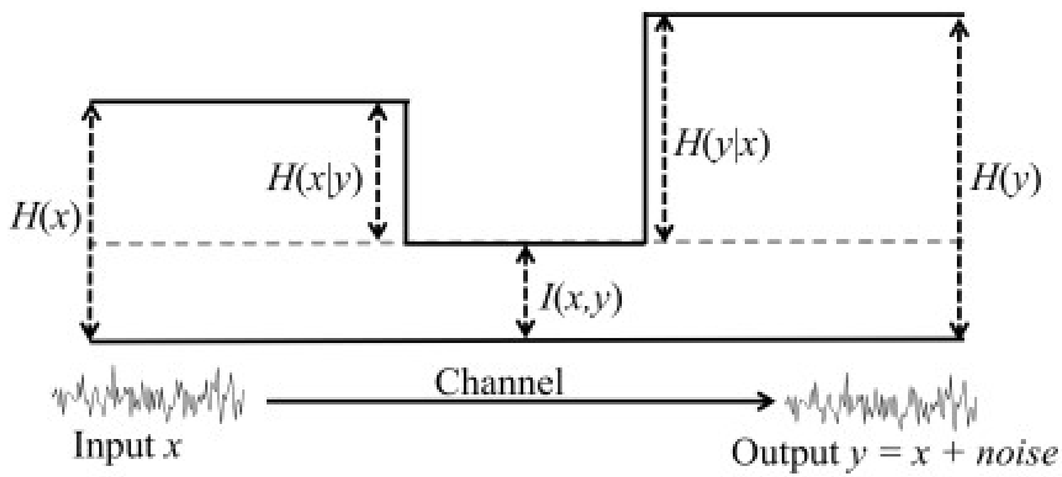
$$H_T(A) := \sum_i \max\{log_2|\lambda_i|, 0\}$$

- In information theory, **entropy rate** is used to measure the rate at which stochastic process generates information
- In feedback control theory, the rate at which a dynamical system generates information is quantified by **topological entropy**

Channel capacity

- Also called Shannon capacity
- **the tightest upper bound** on the average amount of information that can be transmitted over a communication channel

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

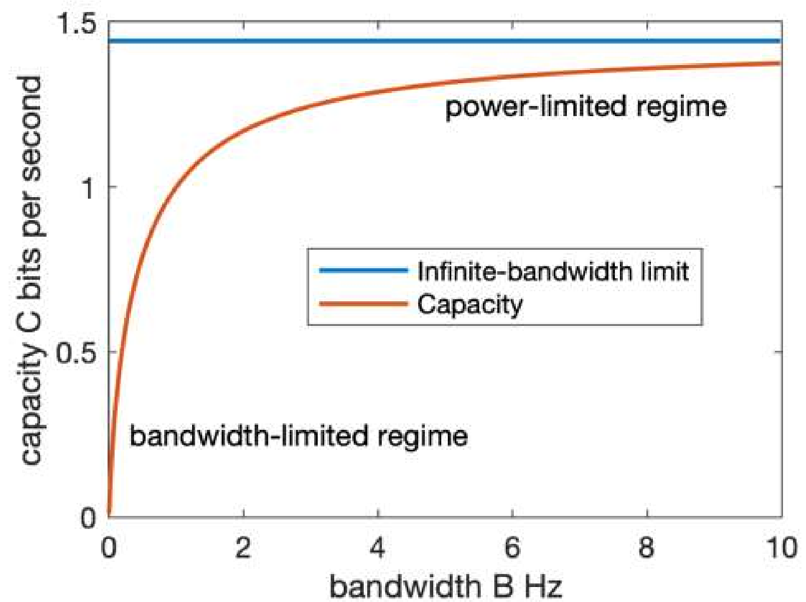


Channel capacity in wireless communications

➢ **Shannon-Hartley theorem**: For channel capacity in point-to-point scenario, the AWGN (additive white gaussian noise) channel capacity is

$$C = \underbrace{B}_{\text{bandwidth}} \log_2(1 + \underbrace{S/N}_{\text{Signal to noise ratio}})$$

capacity — bandwidth — Signal to noise ratio



Shannon's theorems

Claude Shannon

64 languages

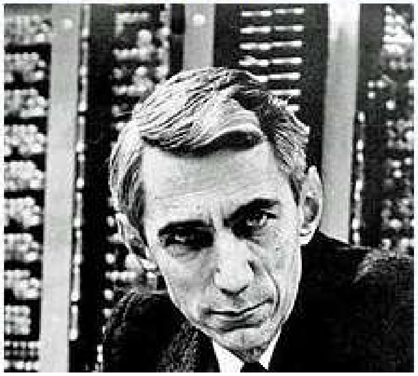
Article Talk — Read Edit View history

From Wikipedia, the free encyclopedia

Claude Elwood Shannon (April 30, 1916 – February 24, 2001) was an American mathematician, electrical engineer, and cryptographer known as a "father of information theory".[1][2]

As a 21-year-old master's degree student at the Massachusetts Institute of Technology (MIT), he wrote his thesis demonstrating that electrical applications of Boolean algebra could construct any logical numerical relationship.[3] Shannon contributed to the field of cryptanalysis for national defense of the United States during World War II, including his fundamental work on codebreaking and secure telecommunications.

Claude Shannon



Biography [edit]

Shannon's theorems

- Source samples X_n are encoded into a digital representation at rate R (bits/sample)
- Decoder produces sample estimates $\hat{X}_n$

Shannon's theorems

- **Shannon's channel coding theorems**: probability of error could be made nearly zero for $R < C$
- **Shannon's source coding theorems**: average number of bits required to represent result of a random event is given by its entropy $R > H$
  - When $R < H$, distortion always happens (never if $R > H$)

Distortion measures

- "Distance" $d(x, \hat{x})$: quantitative measure between two variables x and $\hat{x}$
  - **Hamming distance:** $d(x, \hat{x}) = 1$ if $x \neq \hat{x}$, zero if $x = \hat{x}$.
  - **Squared error:** $d(x, \hat{x}) = (x - \hat{x})^2$

- Distance between sequences $x_1^n \coloneqq \{x_1, \dots, x_n\}$ and $\hat{x}_1^n$:

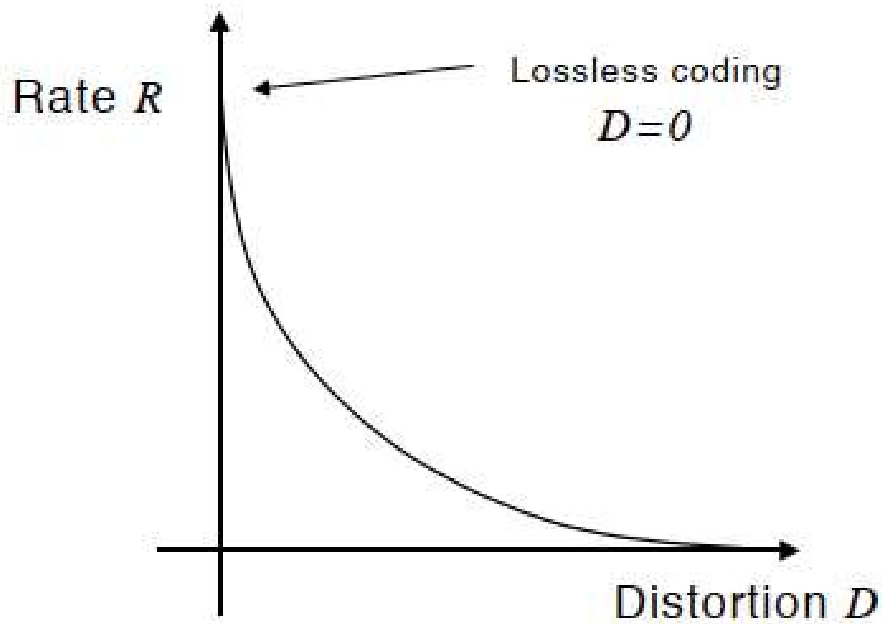
$$d(x_1^n, \hat{x}_1^n) \coloneqq \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{x}_i)^2$$

- Average distortion between two random sequences:

$$D \coloneqq \mathbb{E}[d(X_1^n, \hat{X}_1^n)]$$

Rate distortion theory

- Lossy compression: lower the bit rate R by allowing some acceptable distortion D of the signal



Causality, feedback, and directed information

- based on Massey, Causality, feedback, and directed information, *IEEE Int. Symp. on Inform. Theory and Its Appl.*, 1990
- Discrete channel is **without feedback** if

  $p(x_n|x^{n-1}, y^{n-1}) = p(x^n|x^{n-1}), \forall x^n, y^{n-1}$

Directed information

- Given a pair of random sequences X^n and Y^n, directed information is defined as

$$I(X^n \to Y^n) \coloneqq \sum_{t=1}^{n} I(X^t; Y^t \,|Y^{t-1})$$
$$= H(Y^n) - H(Y^n || X^n)$$
$$= H(Y^n) - \sum_{t=1}^{n} H(Y^t | Y^{t-1}, X^t)$$

Causally conditional entropy

Directed information

- Compared to mutual information, directed information has the **causally conditional entropy** in place of the conditional entropy

$$I(X^n;Y^n) = H(Y^n) - H(Y^n|X^n)$$

- Unlike mutual information, directed information is, in general, non symmetric

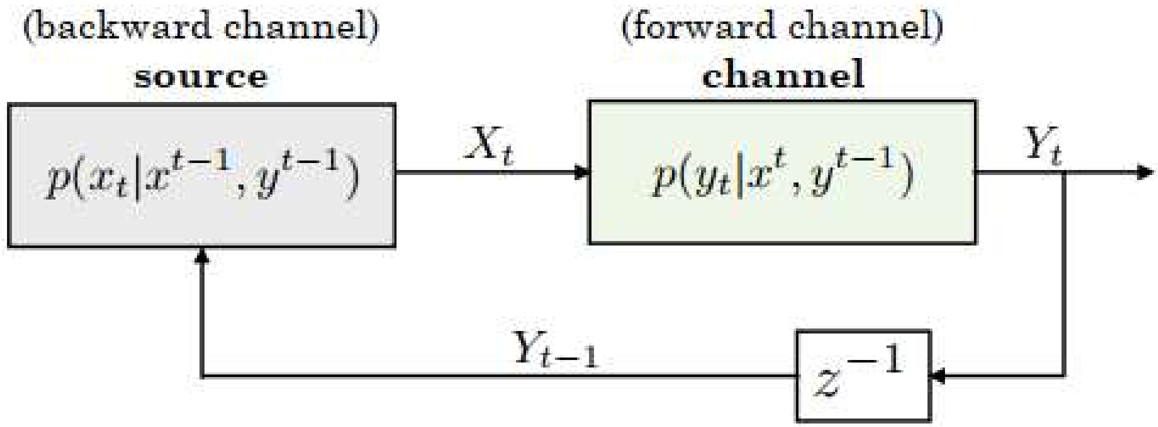
$$I(X^n \to Y^n) \neq I(Y^n \to X^n)$$

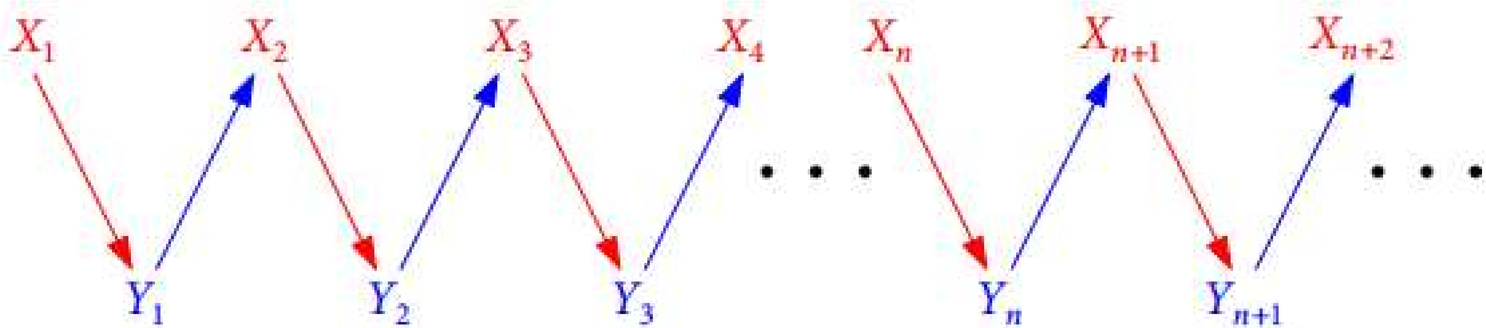
Properties of directed information

- Conservation law: $I(X^n; Y^n) = I(X^n \to Y^n) + I(Y^{n-1} \to X^n)$
$$= I(X^{n-1} \to Y^n) + I(Y^n \to X^n)$$

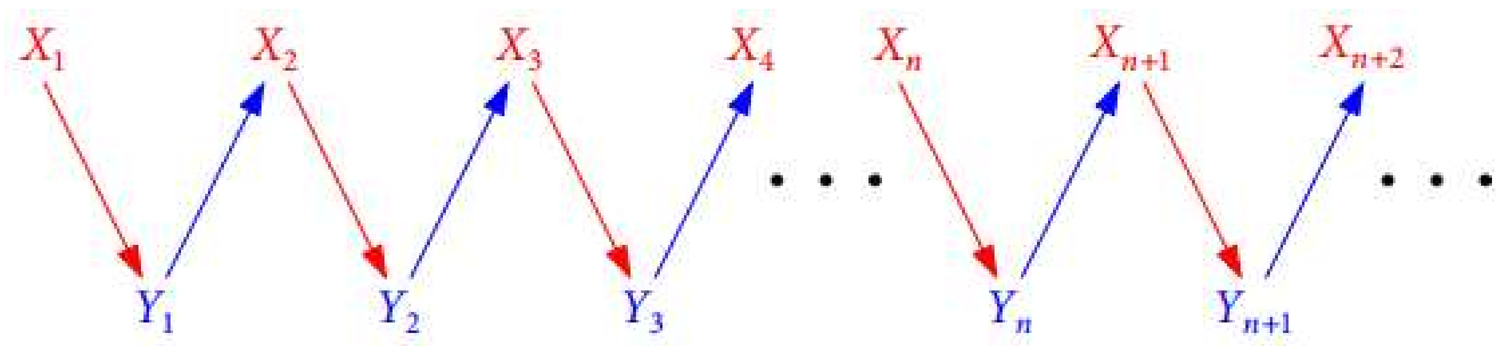
- No feedback case: $I(X^n; Y^n) = I(X^n \to Y^n)$

Causal influence





Causal influence



- Forward link exists **if and only if** $I(X^n \to Y^n) > 0$
- Backward link **if and only if** $I(Y^n \to X^n) > 0$

Topological entropy

- Topological entropy of an **LTI system** with open-loop matrix A (as in $x[k+1] = Ax$) is defined as

$$H_T(A) \coloneqq \sum_i \max\{\log_2|\lambda_i|, 0\},$$ where λ_i are eigenvalues of A

- In information theory, entropy rate is used to measure **the rate at which a stochastic process generates information**
- In feedback control theory, **the rate at which a dynamical system generates information** is quantified by topological entropy

Mahler measure

$$M(A) := M\,\det(zI - A) = \prod_i \max\{|\lambda_i|, 1\} = 2^{H_T(A)}$$

- No reference to any controller or feedback communication – **an intrinsic property of the dynamical system**

Data rate theorem

- Addresses how much information needs to be communicated between the quantizer and controller **for stabilizing a discrete LTI system** $C > H_T(A)$

Conclusion

- Information theory deals with general problems of reliable transmission of data and capacity of channels required for that purpose

- Parallel between information theory and control of unstable system?