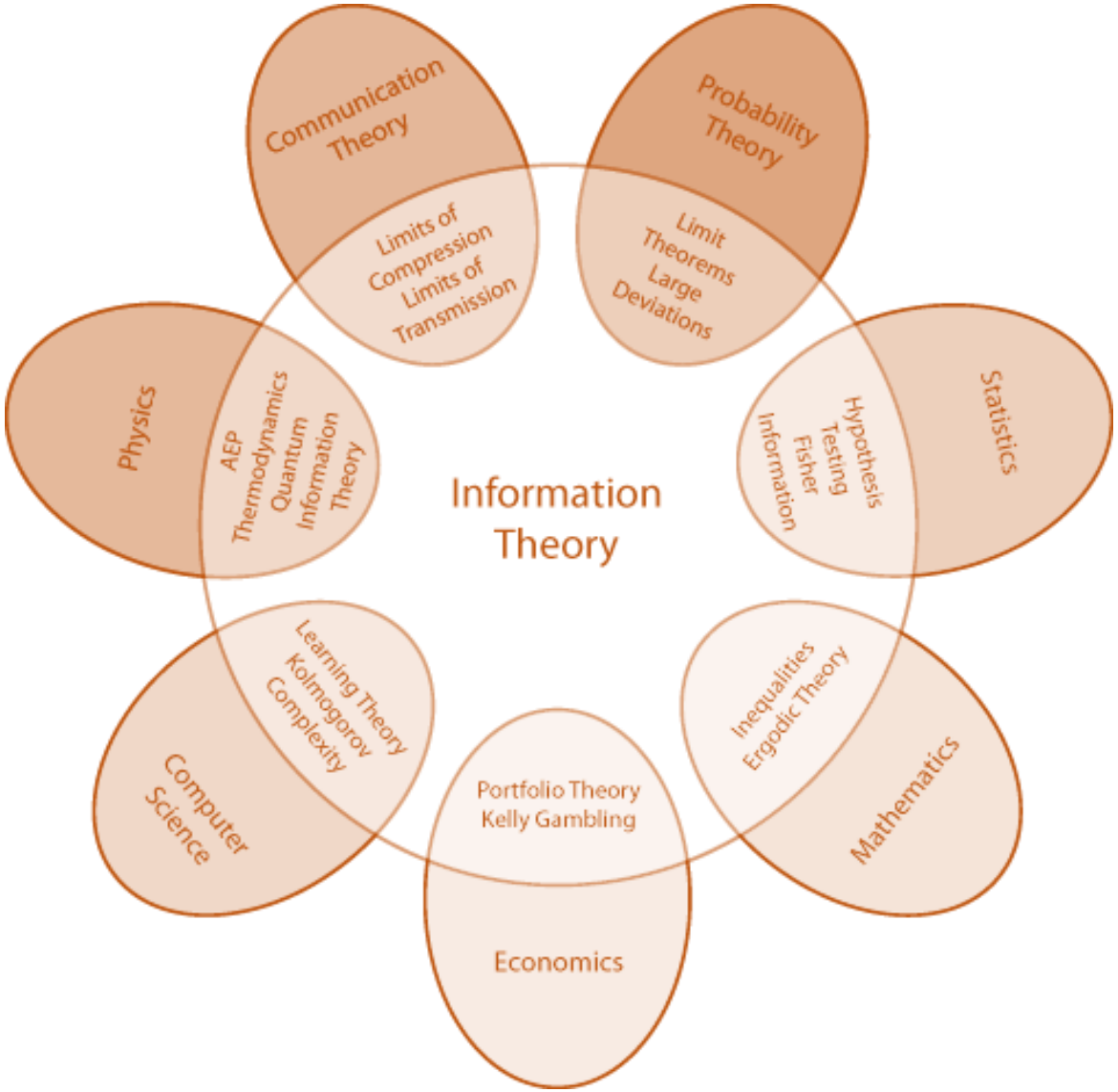
Information Theory

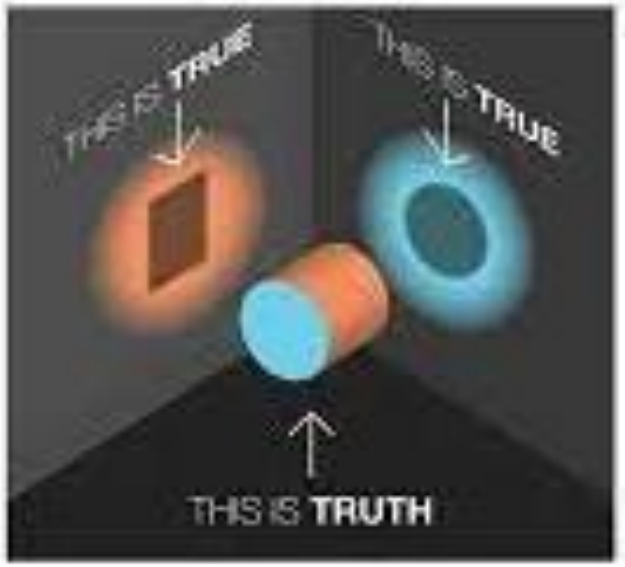
Ethan Ancell & Abigail Cummings

INFORMATION THEORY

- Information Theory is a branch of applied mathematics that deals with quantifying information, communication, and data transmission.

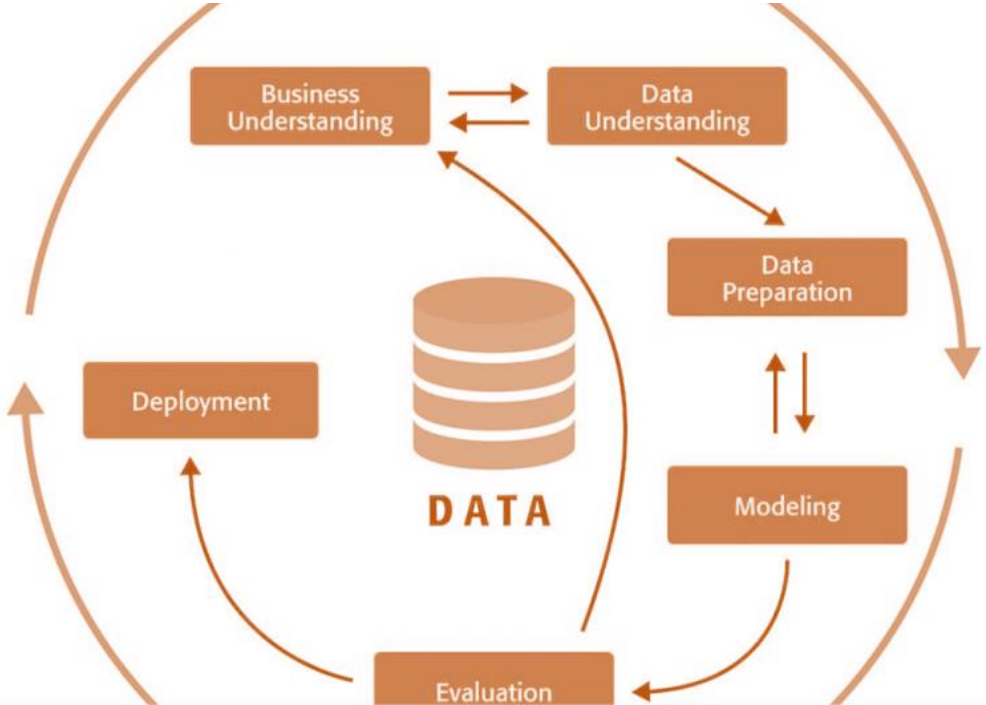


How is Information Theory Relevant?



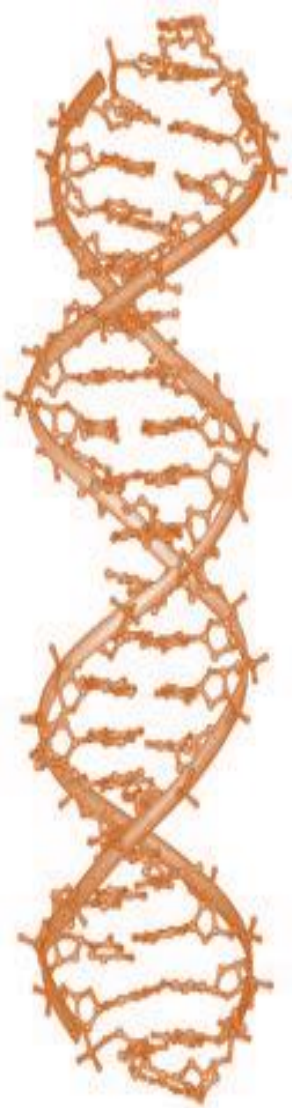
Data Science and Technology

- Machine learning
- AI
- Data Mining



Communication Technology

- Internet
- Mobile networks
- Broadcasting



Biological Sciences and Psychology/Human Perception

- Genetics analyzing
- Neuroscience
- Cognitive development
- Ecology

Fundamental Concepts of Information Theory

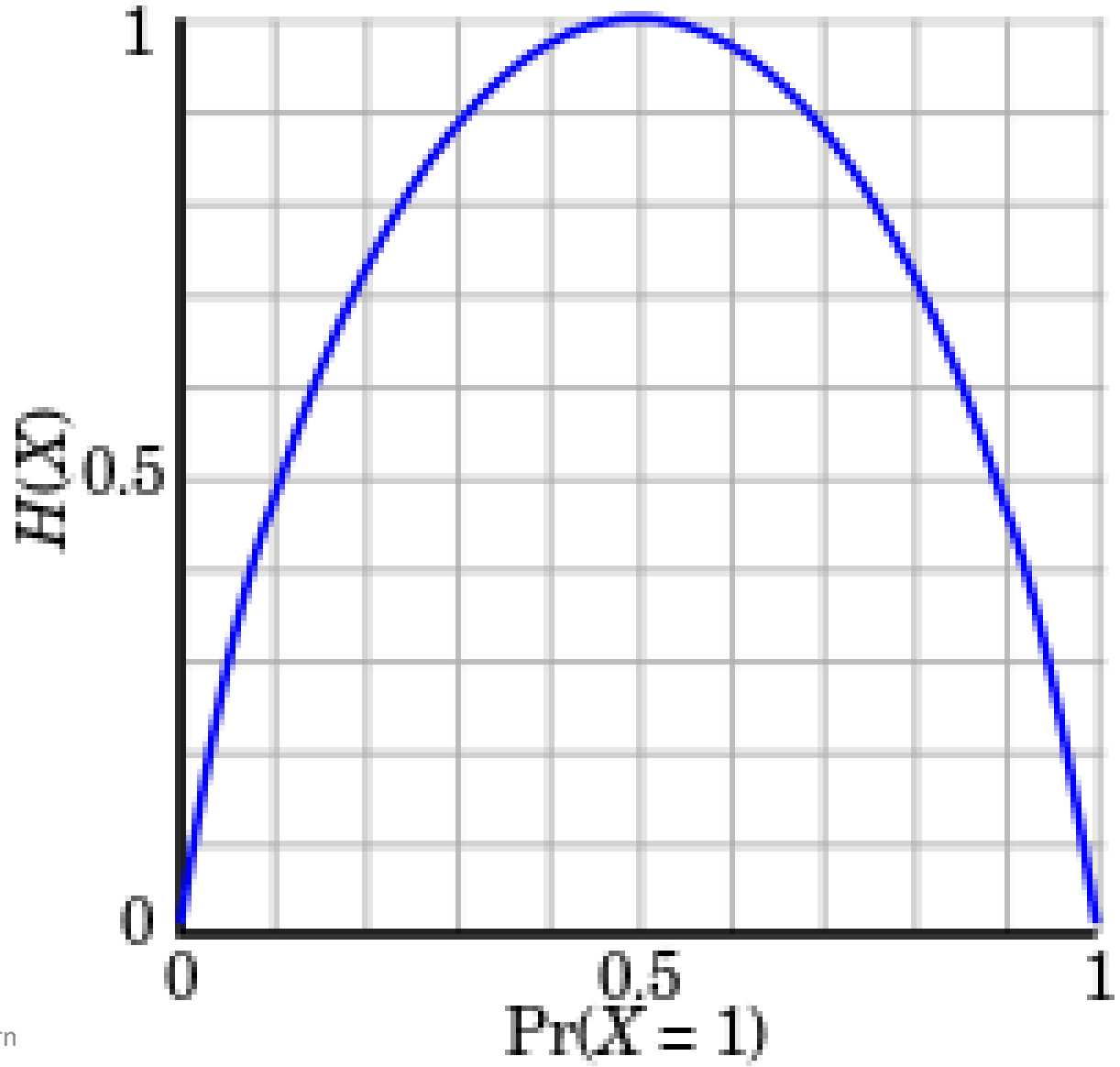
Entropy

Differential Entropy

For continuous variable with pdf *f:*

$$H(X) = \int f(x) \log \frac{1}{f(x)} dx$$

H(X) is the expectation of log(1/f(x)) which is called the *loss.*

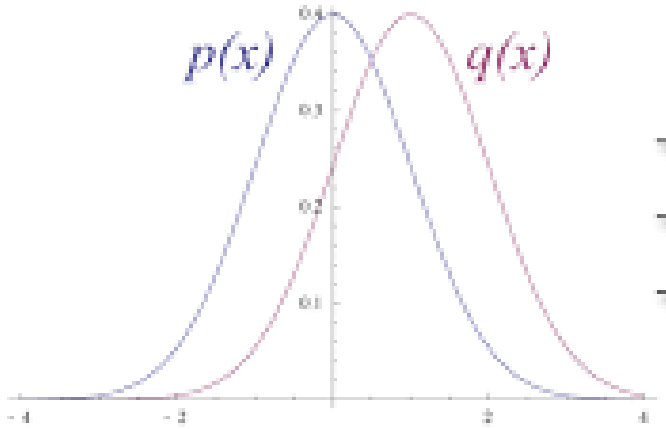


Kullback-Leibler Divergence

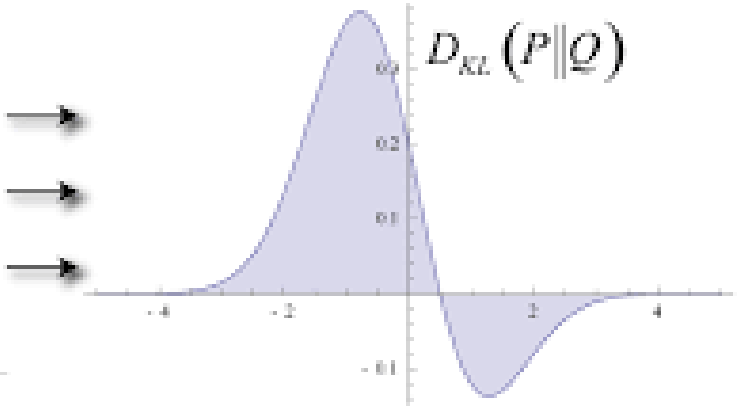
Directed Divergence

Kullback-Leibler risk of *g* related to *f:*

$CE(P||Q) = \int P(x) \log \frac{P(x)}{Q(x)} dx$



Original Gaussian PDF's

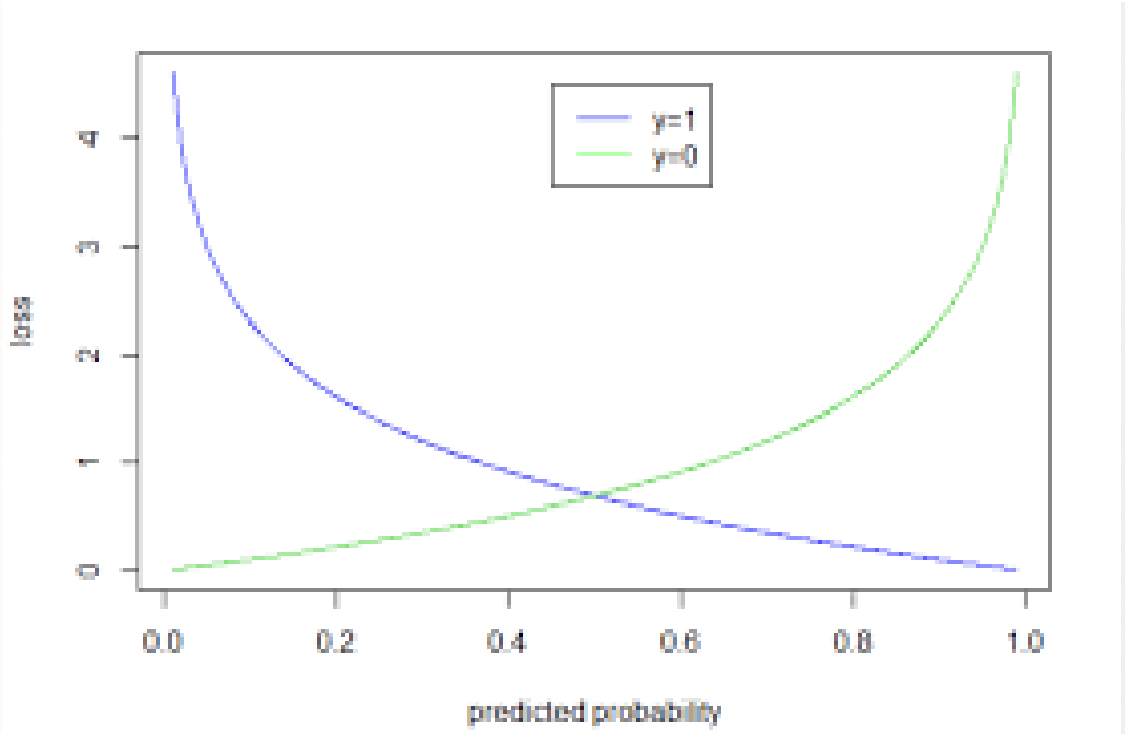


KL Area to be Integrated

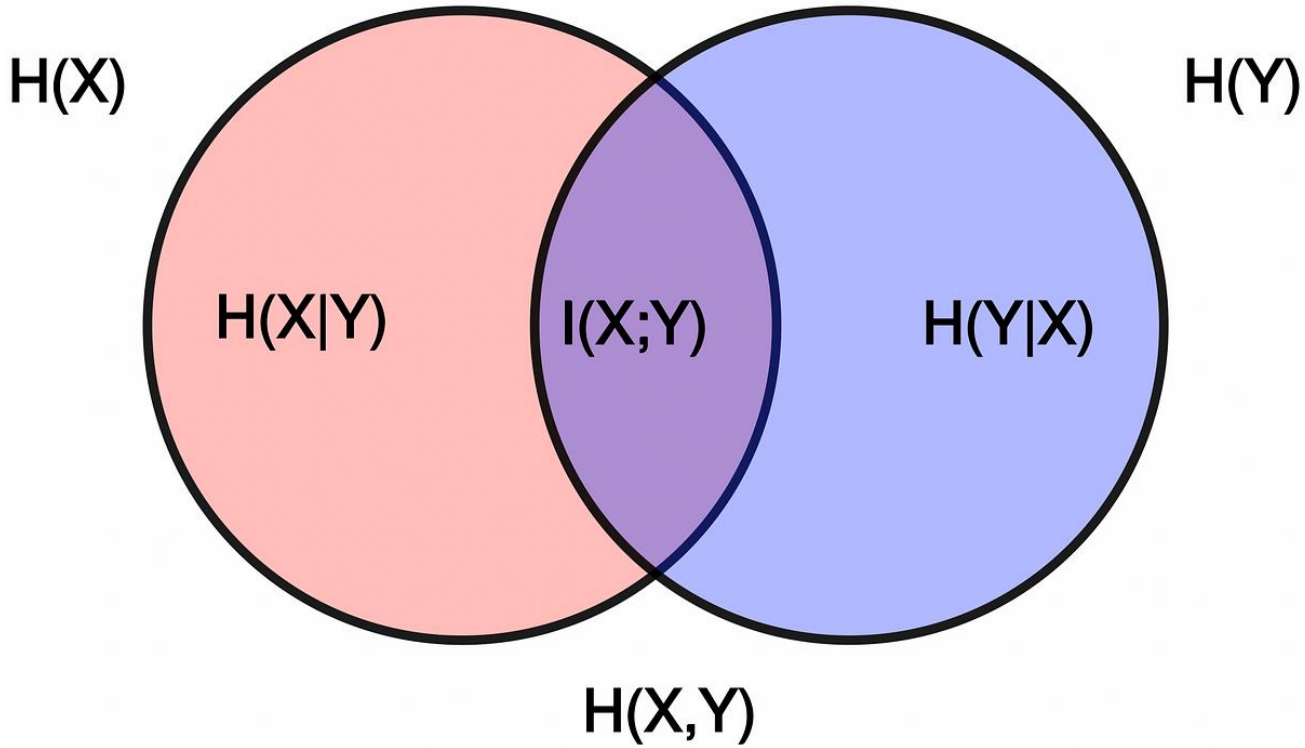
Cross Entropy

- Cross Entropy

$$CE(g|f) = \int f(x) \log \frac{1}{g(x)} dx$$



Mutual Information



Applications of Information Theory

Parameter Estimation

- Information Theory offers methods that are relevant to parameter estimation including:
  - Maximum Likelihood Estimator (MLE)
  - Bayesian Estimation

Data Analysis

- Offers tools and concepts that are invaluable to data analysis
  - Entropy and Information Gain
  - Mutual Information
  - Kullback-Leibler Divergence
  - Coding theory and Compression

Variational Inference

- Alternative to MCMC
- This is a technique used in Bayesian statistics and machine learning and is meant to approximate complex probability distributions

Data Compression

- Primary goal: represent information in a more efficient manner, using as few bits as possible to store or transmit data
  - Entropy and Redundancy
  - Lossless vs Lossy Compression
  - Arithmetic Coding
    - Representing messages by a number within a specified range


Thank you!