

Stat Learning 1: Information Theory

Overview of the summer

- <https://github.com/JJBannister/StatisticalLearningGroup>
- Bi-weekly meetings (mostly)
- Friday 12-1
- Presentations by grad students and postdocs
- No textbook, lots of presenter freedom
- High level exploration of different topics in Statistics and Machine Learning

Lecture (David Mumford - Pattern Theory Chapter 1)

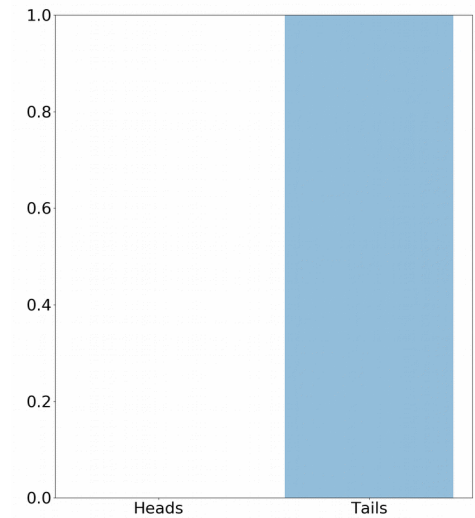
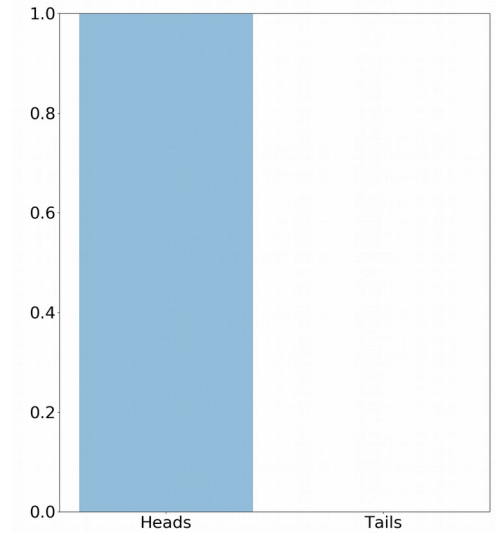
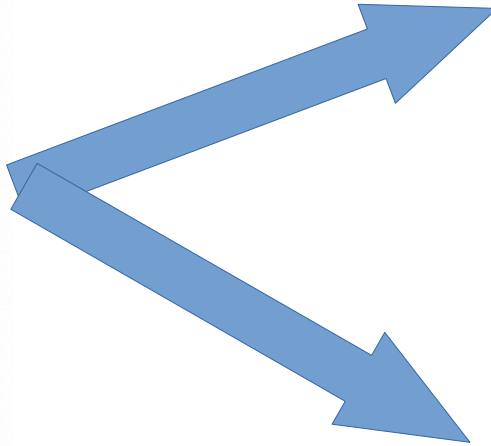
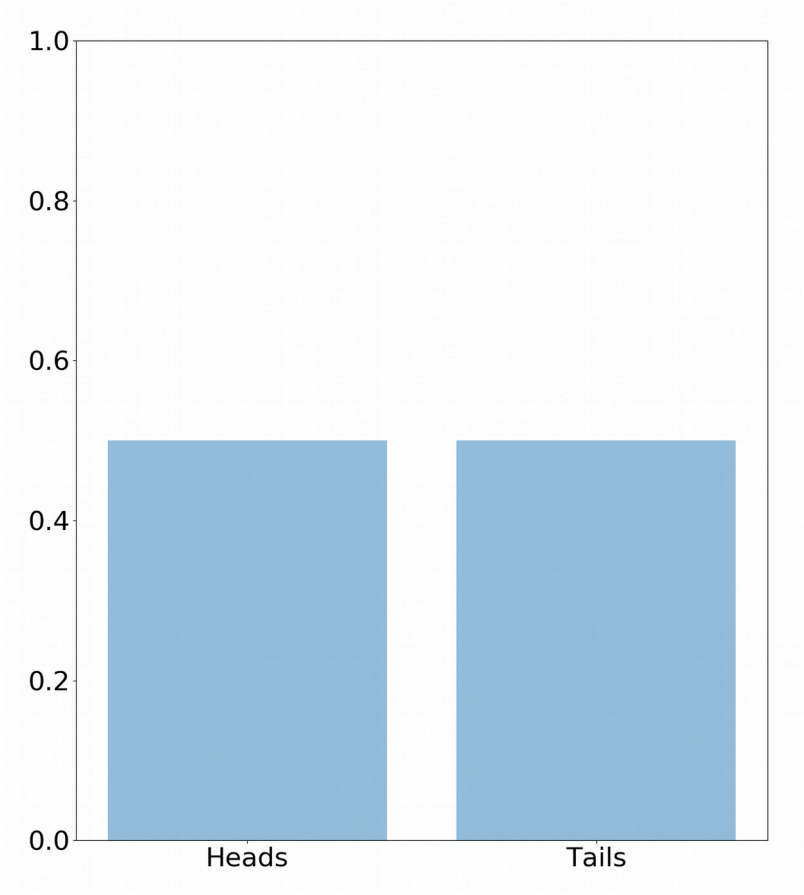
- Entropy
- Conditional Entropy
- Relative Entropy
- eg. English Text

Thanks to the GSA Quality Money Fund, BMEG and MTC!

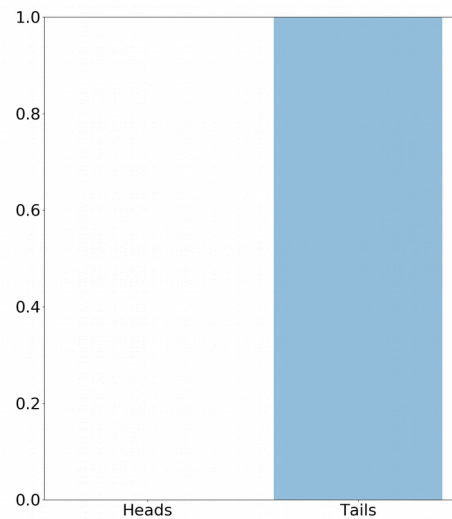
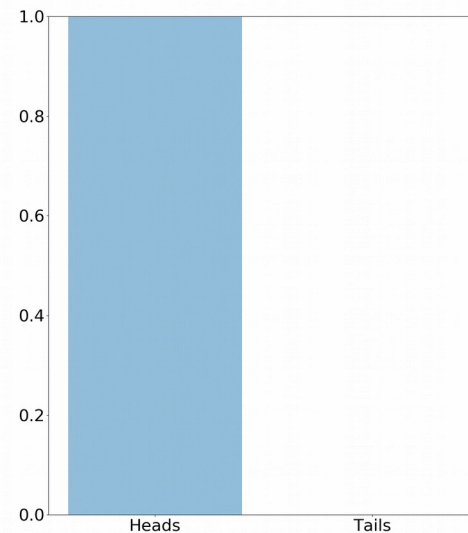
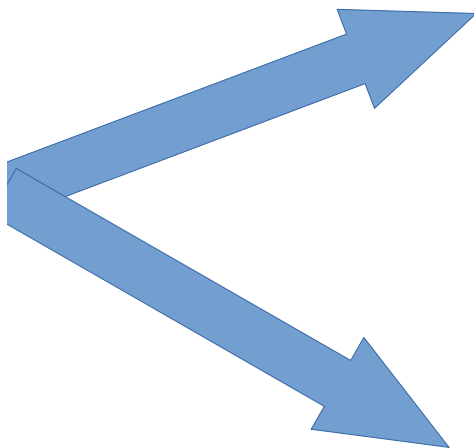
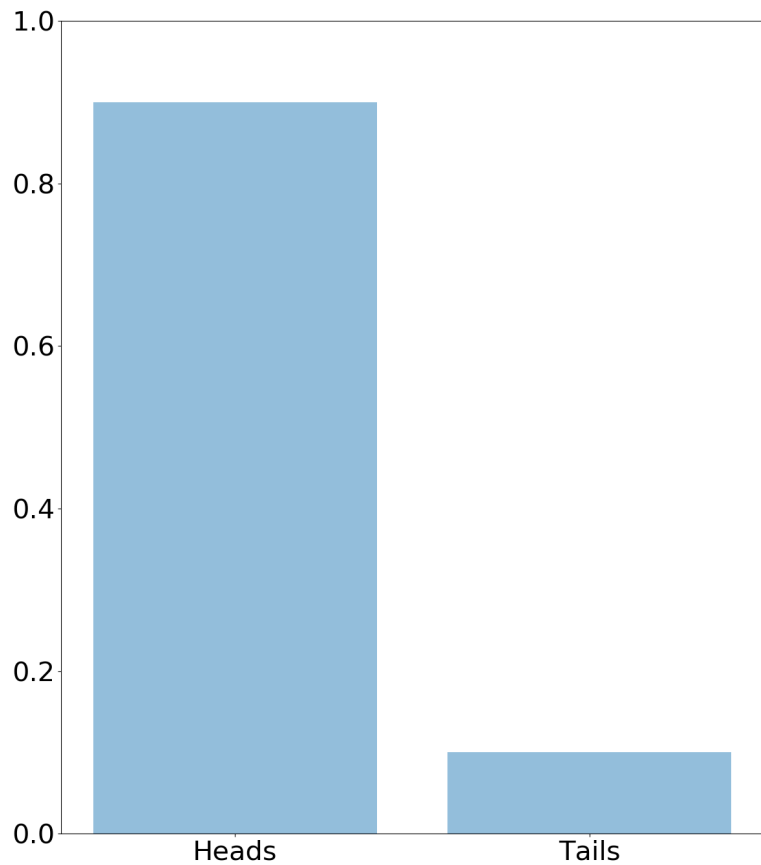
Entropy

- Physics: **Disorder** (Boltzman 1877)
 - Statistical mechanics (entropy+energy), thermodynamics, quantum computing, black holes!
- Communication Systems: **Information Capacity** (Shannon 1948)
 - Channel capacity, coding, compression...
- Statistics: **Uncertainty** (Kullback 1959)
 - Model training, model comparison, hypothesis testing, experimental design...

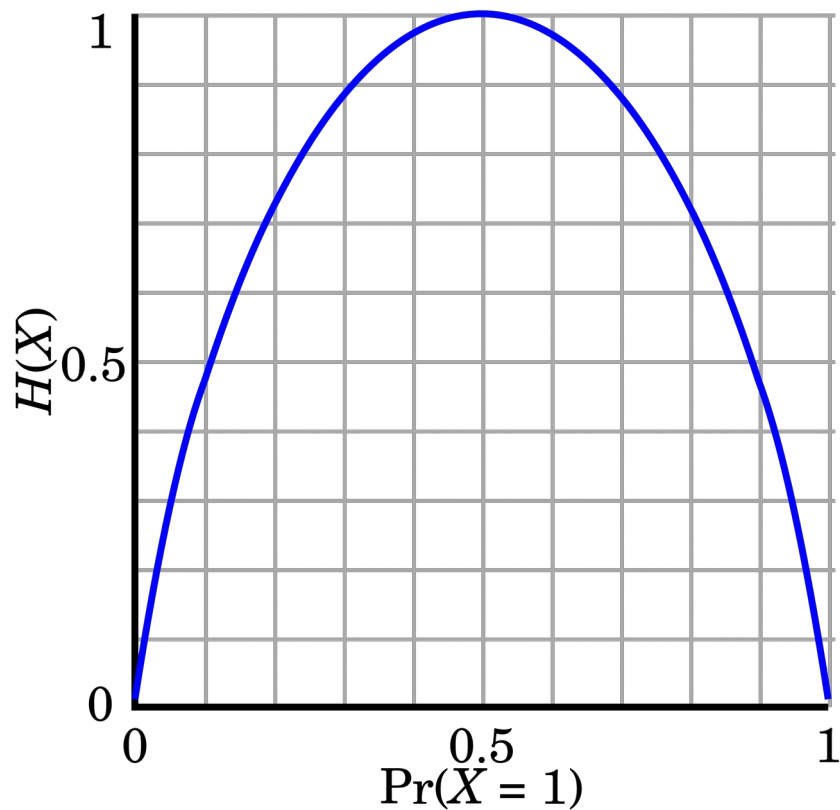
The Fair Coin



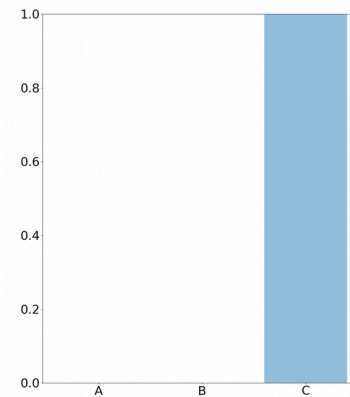
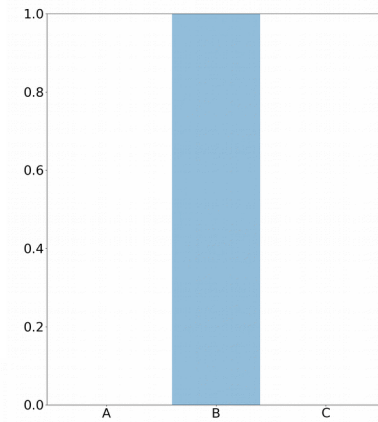
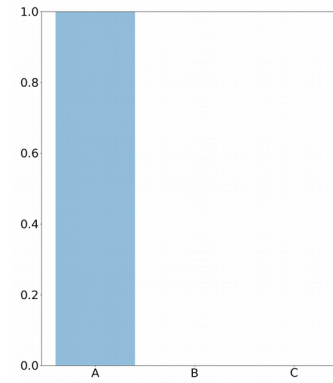
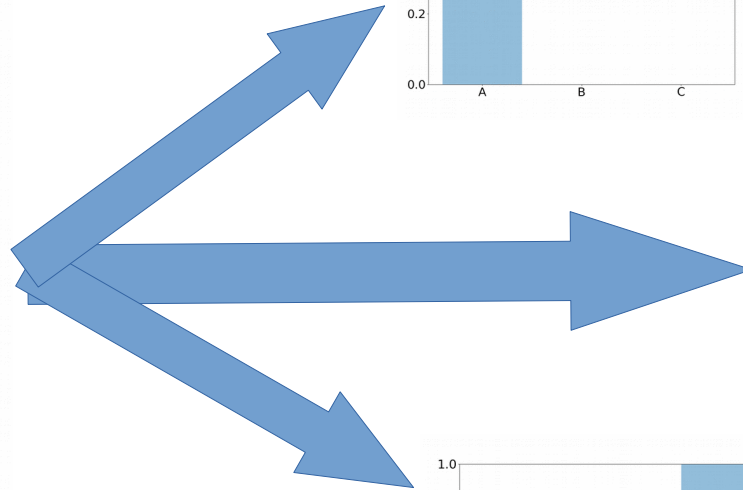
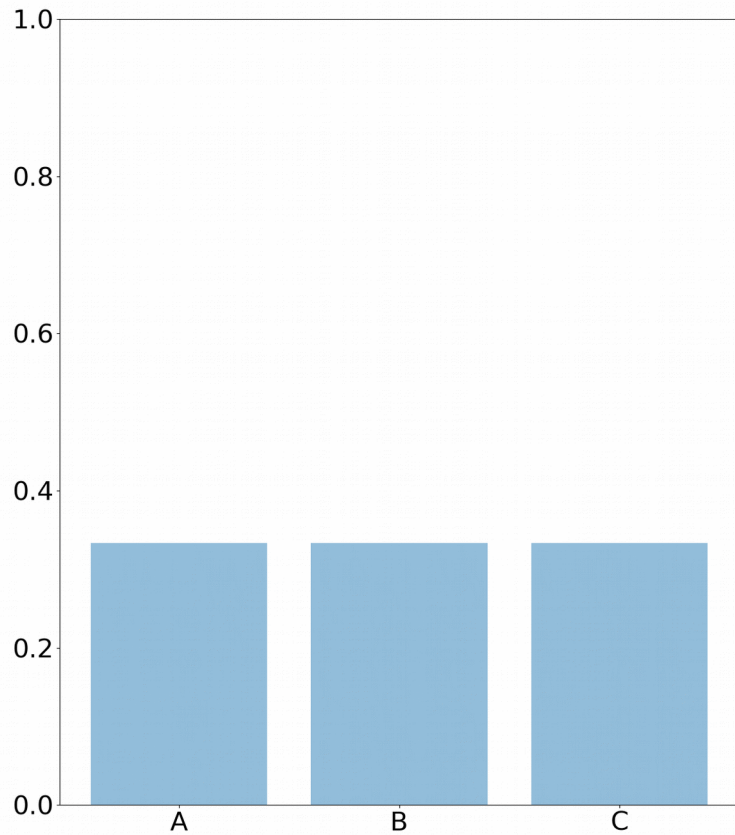
The Loaded Coin



The Loaded Coin



The 3-sided Die



What is the entropy of english text?

- **English text** {a,b,c,d,...}
- Morse Code {*, -, ' '}
- DNA {G,A,T,C}
- Anything digital! {1,0}

Example

How much information did our model learn?

- If we can express it we can optimize for it!
- What is the expected difference in information between the “correct” model, and the naive model?

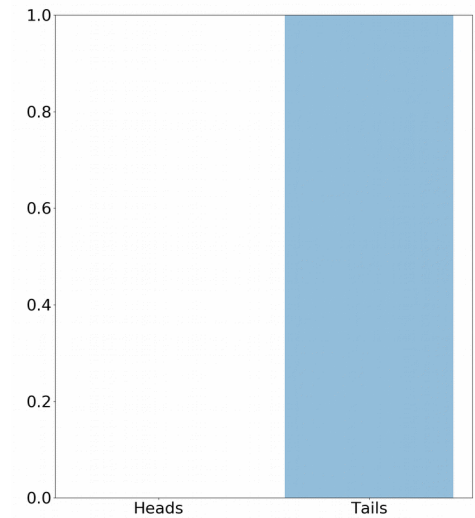
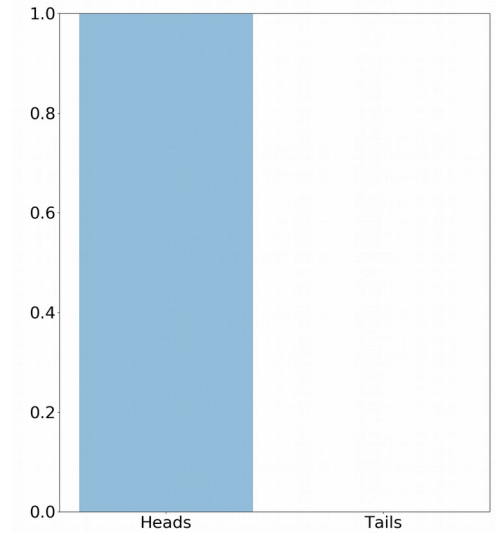
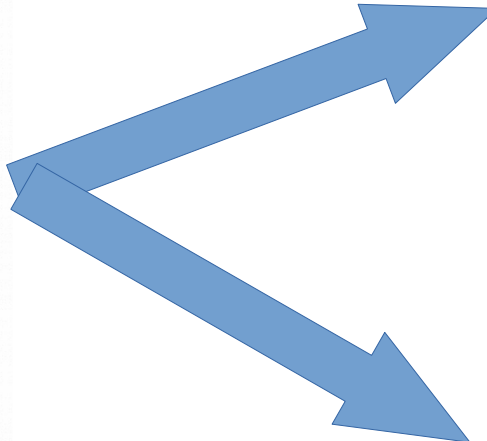
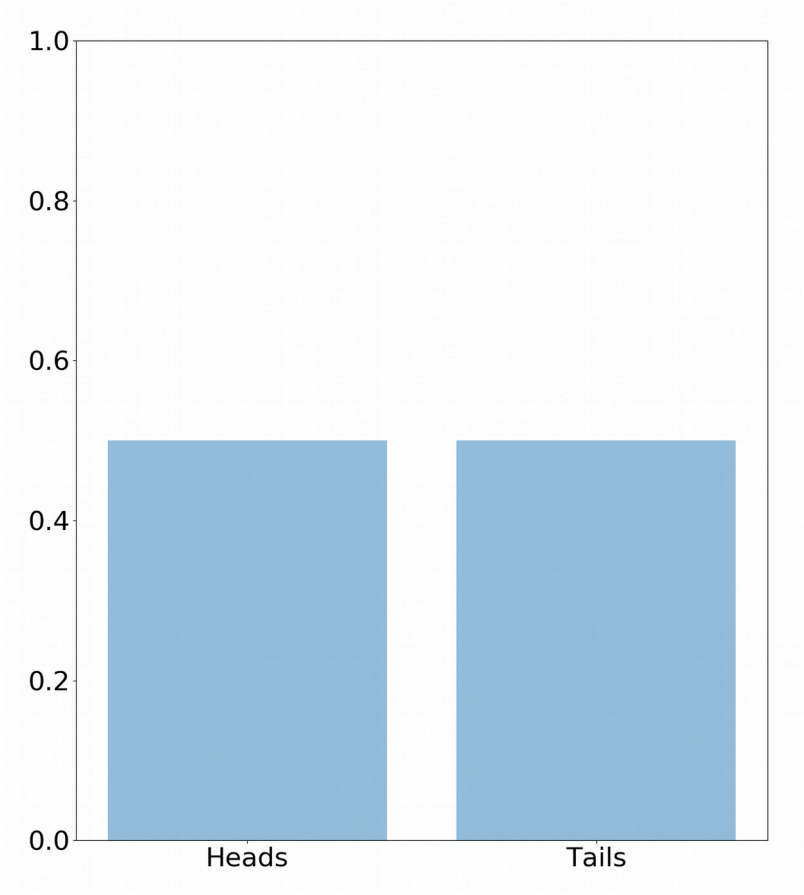
KL Divergence

- Not symmetric, not a distance metric.

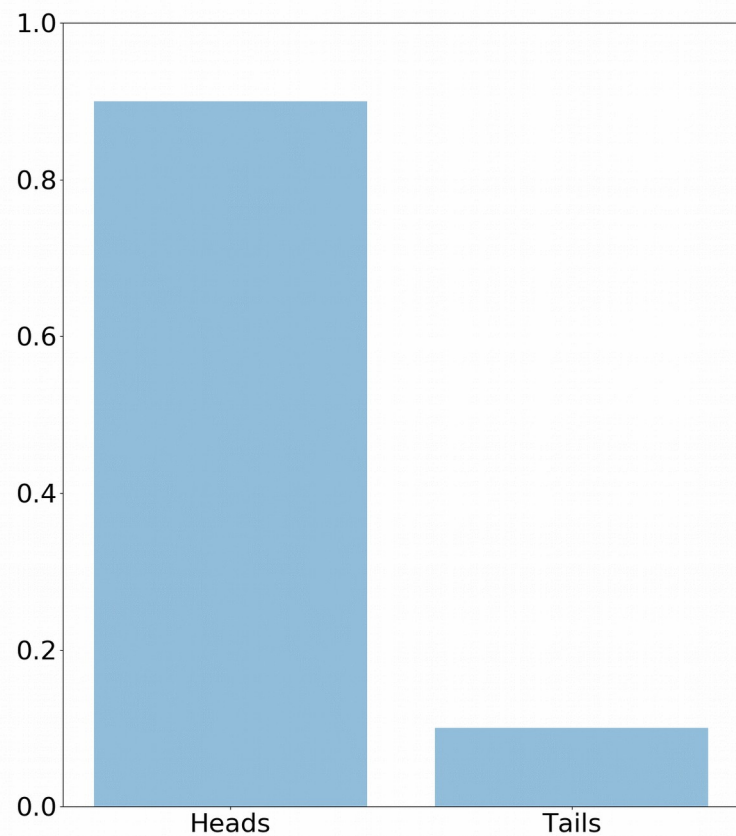
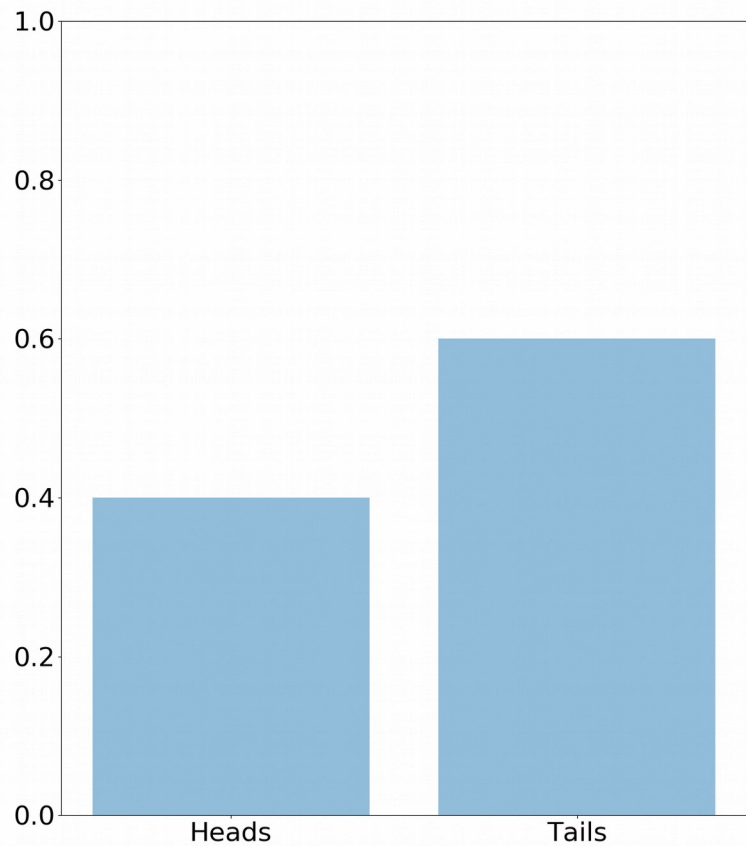
KL Divergence

- Model Fitting (MLE, VI, VAE, ICA, InfoGAN)
- Model comparisson (AIC, WAIC, GIC, TIC, Bayes Factor)
 - Goodness of fit + complexity penalty
 - Entropy + energy

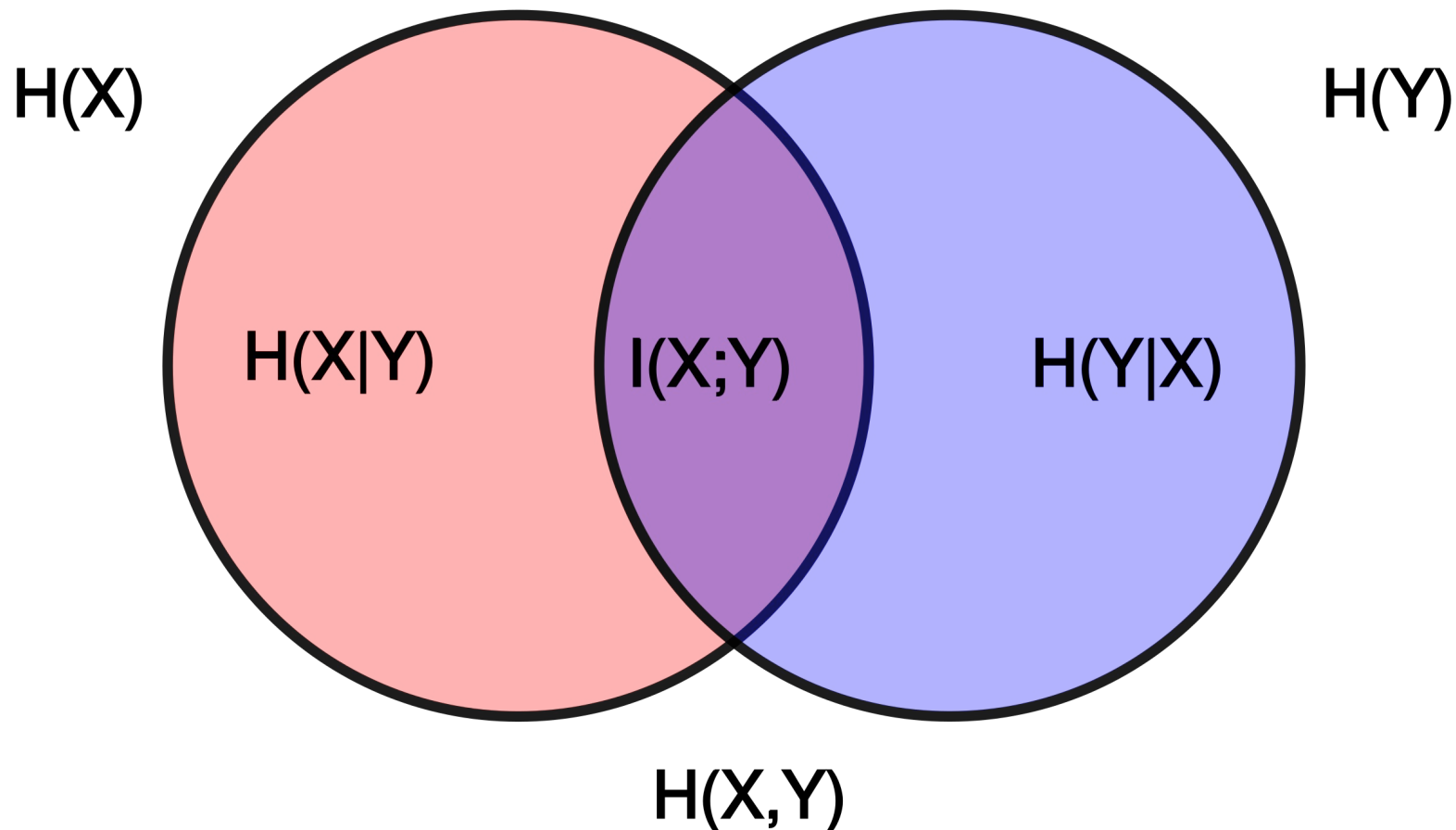
The Fair Coin



Incremental Learning



Mutual Information



What about Continuous Variables?

- Differential Entropy is not exactly entropy...
 - The true entropy of a continuous function is infinity
 - Differential entropy must be relative to a reference measure and so:
 - Units matter (mm vs m)
 - Can be negative