

# Entropy: Motivation

- **Feature selection:**
  - If we use only a few words to classify docs, what kind of words should we use?
  - $P(\text{Topic} | \text{“computer”}=1)$  vs  $p(\text{Topic} | \text{“the”}=1)$ : which is more random?
- **Text compression:**
  - Some documents (less random) can be compressed more than others (more random)
  - Can we quantify the “compressibility”?
- **In general, given a random variable  $X$  following distribution  $p(X)$ ,**
  - How do we measure the “randomness” of  $X$ ?
  - How do we design optimal coding for  $X$ ?