6.897
L11.5
Mar. 10, 2005

Information-theory basics

- $H(x)$ = entropy of x
  = # bits to represent x as a sample from its distribution
  $= \sum_{x_0} \Pr\{x = x_0\} \lg \frac{1}{\Pr\{x = x_0\}}$
- $H(x|y)$ = entropy of x given y
  = # bits to represent x if you know y's value
  $= E_{y_0}[H(x|y=y_0)]$ → propagate into Pr's
- $I(x:y)$ = shared information between x & y
  $= H(x) + H(y) - H((x,y))$
- $I(x:y|z) = E_{z_0}[I(x:y|z=z_0)]$ → propagate into H's

Proof sketch of round-elimination lemma:

- call Alice's first message $m = m(x_1, \ldots, x_k)$
- $a = |m| \geq H(m) = \sum_{i=1}^{k} I(x_i : m | x_1, \ldots, x_{i-1})$
  ↑ theorem from information theory
- if i is distributed uniformly,
  then $E_i[I(x_i : m | x_1, \ldots, x_{i-1})]$ = average term in sum $\leq \frac{H(m)}{k} \leq \frac{a}{k}$.
- intuition: Bob knows $x_1, \ldots, x_{i-1}$ & receives m
  $\Rightarrow$ learns $I(x_i : m)$ about x_i
- build protocol for $f(x)$ as follows:
  - fix $x_1, \ldots, x_{i-1}$ & i randomly in advance (indep. of x)
  - set $x_i = x$
  - run $f^{(k)}$ protocol, starting at second message,
    assuming first message $m = m(x_1, \ldots, x_{i-1}, \tilde{x}_i, \ldots, \tilde{x}_k)$
    for $\tilde{x}_i, \ldots, \tilde{x}_k$ chosen uniformly by Bob (who doesn't know x_i)
  - guess $I(x_i : m)$ correctly with probability $\approx a/k$
  - claim that, with probability $\sqrt{a/k}$,
    $\exists x_{i+1}, \ldots, x_k$ such that $m(x_1, \ldots, x_{i-1}, \tilde{x}_i, \ldots, \tilde{x}_k) = m(x_1, \ldots, x_k)$
    distributed roughly the same (so error prob. δ is preserved)

"Average Encoding Theorem"