EE 231A: Information Theory

Rick Wesel
Boelter 6412A
(310) 256-2150
wesel@ee.ucla.edu

1

EE 231E: Information Theory
Lecture 1

A. Introduction
B. Entropy
C. Relative Entropy
D. Mutual Information

2

Relative entropy

- A way to compare how close two distributions are.
- The "penalty" for compressing using the wrong distribution.
- Specifically, if $X \sim p(x)$ but we represent it in a way that would be efficient if $X \sim q(x)$, our representation will require

$$\underbrace{H(p)}_{\text{entropy}} + \underbrace{D(p \| q)}_{\substack{\text{relative entropy} \\ \text{between p and q}}} \quad \text{bits}$$

3

Relative entropy definition

$$D(p \| q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$$

$$D(p \| q) = \underbrace{-\sum_x p(x) \log q(x)}_{} - \left(\underbrace{-\sum_x p(x) \log p(x)}_{} \right)$$

bits required by assuming *q(x)* when *X~p(x)*

bits required by correct assumption *X~p(x)*

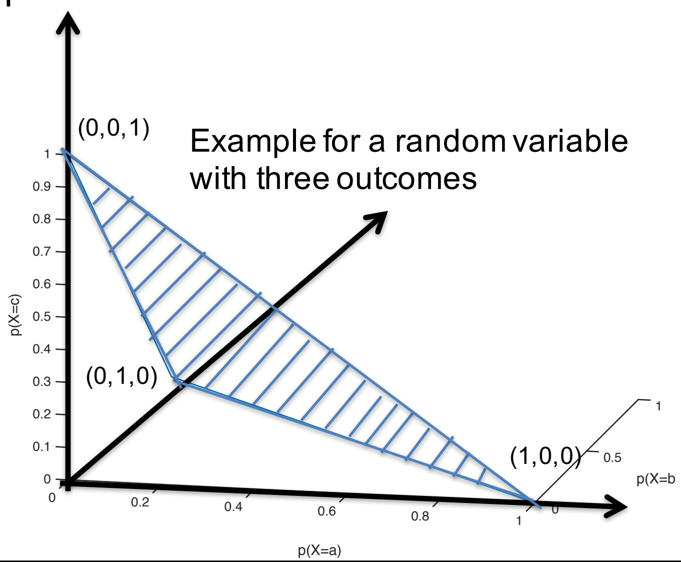
- For two distributions $p(x)$ and $q(x)$

$$D(p \| q) = E_X\left[\log p(X)\right] - E_X\left[\log q(X)\right] = E_X\left[\log \frac{p(X)}{q(X)}\right]$$

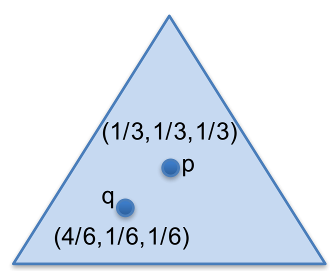
4

A probability simplex

- A probability simplex is the set of all possible pmf's with a particular number of outcomes.



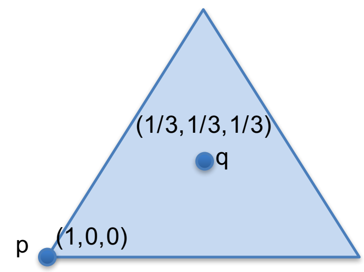
Example D(p||q) Computation



$$D(p\,\|\,q)=\sum_{x}p(x)\log\frac{p(x)}{q(x)}$$

$$=\frac{1}{3}\log\frac{1/3}{4/6}+\frac{1}{3}\log\frac{1/3}{1/6}+\frac{1}{3}\log\frac{1/3}{1/6}=\frac{1}{3}$$

D(p||q) at a corner point



$$D(p\,\|\,q) = 1\log\frac{1}{1/3} + \underbrace{0\log\frac{0}{1/3}}_{0} + \underbrace{0\log\frac{0}{1/3}}_{0} = \log 3$$

$$D(q\,\|\,p) = \frac{1}{3}\log\frac{1/3}{1} + \underbrace{\frac{1}{3}\log\frac{1/3}{0}}_{\infty} + \underbrace{\frac{1}{3}\log\frac{1/3}{0}}_{\infty} = \infty$$

7

Conditional relative entropy

$$D(p(y\,|\,x)\,\|\,q(y\,|\,x)) = E_X E_{Y|X=x}\left[\log\frac{p(y\,|\,x)}{q(y\,|\,x)}\right]$$

$$= \sum_x p(x) \sum_y p(y\,|\,x)\log\frac{p(y\,|\,x)}{q(y\,|\,x)}$$

$$= \sum_{x,y} p(x,y)\log\frac{p(y\,|\,x)}{q(y\,|\,x)}$$

$$= E_{p(x,y)}[-\log q(Y\,|\,X) - -\log p(Y\,|\,X)]$$

8

Chain rule for relative entropy

$$D(p(x,y)\|q(x,y)) = D(p(x)\|q(x)) + D(p(y|x)\|q(y|x))$$

$$\begin{aligned}
&D(p(x,y)\|q(x,y)) \\
&= E_{p(x,y)}[-\log q(X,Y)] - E_{p(x,y)}[-\log p(X,Y)] \\
&= E_{p(x,y)}[-\log q(X)q(Y|X)] - E_{p(x,y)}[-\log p(X)p(Y|X)] \\
&= E[-\log q(X)] + E[-\log q(Y|X)] - E[-\log p(X)] - E[-\log p(Y|X)] \\
&= D(p(x)\|q(x)) + D(p(y|x)\|q(y|x))
\end{aligned}$$

9

EE 231E: Information Theory Lecture 1

A. Introduction
B. Entropy
C. Relative Entropy
D. Mutual Information
E. Axiomatic Development of Entropy

10