Summary

of the talk ”Entropy, relative entropy and mutual information”

Sophia Koch, 23.04.2015

1 Entropy H

X discrete random variable, $\mathcal{X}$ alphabet, $p(x) = Pr(X = x),\ x \in \mathcal{X}$, probability mass function, $0 \log 0 := 0$

$$H(X) := -\sum_{x\in\mathcal{X}} p(x) \log_2 p(x) =: H(p)$$
$$H(X,Y) := -\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x,y) \log p(x,y)$$
$$H(Y|X) := -\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x,y) \log p(y|x)$$

- $H(X) \geq 0$
- $H(X) \leq \log |\mathcal{X}|$, equality iff X uniform distribution
- $H(X|Y) \leq H(X)$, equality iff $X,\ Y$ independent
- $H(X_1, \ldots, X_n) \leq \sum_{i=1}^n H(X_i)$, equality iff $\{X_i,\ i = 1, \ldots, n\}$ independent
- $H(p)$ concave

2 Relative entropy D

X discrete random variable with probability mass functions $p(x),\ q(x),\ 0 \log \frac{0}{0} := 0$, $0 \log \frac{0}{q} := 0$, $p \log \frac{p}{0} := \infty$

$$D(p\|q) := \sum_{x\in\mathcal{X}} p(x) \log \frac{p(x)}{q(x)}$$
$$D(p(Y|X)\|q(Y|X)) := \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x,y) \log \frac{p(y|x)}{q(y|x)}$$

- $D(p\|q) \geq 0$, equality iff $p(x) = q(x)$ for all $x \in \mathcal{X}$
- convex: $D(\lambda p_1 + (1-\lambda)p_2 \| \lambda q_1 + (1-\lambda)q_2) \leq \lambda D(p_1\|q_1) + (1-\lambda) D(p_2\|q_2)$

3 Mutual information I

X, Y discrete random variables, $p(x,y)$ joint probability mass function, $p(x)$, $p(y)$ marginal probability mass functions

$$I(X;Y) := \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}$$

$$I(X;Y|Z) := \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}\sum_{z\in\mathcal{Z}} p(x,y,z)\log\frac{p(x,y|z)}{p(x|z)p(y|z)}$$

- $I(X;Y) \geq 0$, equality iff X, Y independent
- $I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = I(Y;X)$
  $= H(X) + H(Y) - H(X,Y)$
  $I(X;X) = H(X)$
- concave for $p(x)$ variable, $p(y|x)$ fixed; convex for $p(y|x)$ variable, $p(x)$ fixed

4 Chain rule

Entropy:	$H(X_1,\ldots,X_n)$	$=$	$\sum_{i=1}^{n} H(X_i\|X_{i-1},\ldots,X_1)$
Relative Entropy:	$D(p(x,y)\\|q(x,y))$	$=$	$D(p(x)\\|q(x)) + D(p(y\|x)\\|q(y\|x))$
Mutual Information:	$I(X_1,\ldots,X_n;Y)$	$=$	$\sum_{i=1}^{n} I(X_i;Y\|X_{i-1},\ldots,X_1)$

5 Data processing inequality

X, Y, Z random variables, Markov chain $X\to Y\to Z$ ($\Leftrightarrow$ conditional distribution of Z only depends on Y; Z conditionally independent of X)

$$I(X;Y) \geq I(X;Z), \text{ equality iff } I(X;Y|Z) = 0$$

6 Fano's inequality

$\hat{X}$ estimation of X knowing Y s.t. $X\to Y\to\hat{X}$, $P_e := Pr(X\neq\hat{X})$

$$H(P_e) + P_e\log|\mathcal{X}| \geq H(X|\hat{X}) \geq H(X|Y)$$

References

[1] Cover, Thomas; Thomas, Joy: ”Elements of Information Theory”. In: John Willey & Sons, second edition, 2006

[2] Csiszár, Imre; Körner, János: ”Information Theory - Coding Theorems for Discrete Memoryless Systems”. In: Cambridge University Press, second edition, 2011

[3] El Gamal, Abbas; Kim, Young-Han: "Network Information Theory". In: Cambridge University Press, 2011

[4] MacKay, David: ”Information Theory, Inference, and Learning Algorithms”. In: Cambridge University Press, 2003

[5] Wolf, Michael: ”http://www−m5.ma.tum.de/foswiki/pub/M5/Allgemeines/MA5103_2012W/lecture1.pdf”. 2012

[6] Wolf, Michael: ”http://www−m5.ma.tum.de/foswiki/pub/M5/Allgemeines/MA5103_2012W/lecture2.pdf”. 2012

[7] Wolf, Michael: ”http://www−m5.ma.tum.de/foswiki/pub/M5/Allgemeines/MA5103_2012W/lecture3.pdf”. 2012