Elements of Information Theory
A Graphical Summary

Pushpendre Rastogi

October 2020

1 Basic Information Theory concepts

Concept	Symbol	Definition
Entropy	$\mathrm{H}(X)$	$-\sum p_x(x) \log p_x(x)$
Joint Entropy	$\mathrm{H}(X, Y)$	$-\sum p_{xy}(x, y) \log p_{xy}(x, y)$
Cond Entropy	$\mathrm{H}(Y\|X)$	$\mathrm{E}_x[\mathrm{H}(Y\|x)] = -\sum_x p_x(x) \sum_y p_{y\|x}(y) \log p_{y\|x}(y)$
Cross Entropy	$C(Y; X)$	$-\sum p_x(x) \log p_y(x)$
KL-Divergence	$\mathrm{D}(X\|\|Y) = \mathrm{D}(p_x\|\|p_y)$	$-\sum p_x(x) \log \big(p_x(x)/p_y(x)\big)$
Mutual Info	$\mathrm{MI}(X; Y) = \mathrm{MI}(Y; X)$	$\mathrm{KL}(p_{xy}\|\|p_x p_y)$

These quantities follow three types of relations

1. Basic Inequalities. Based on concavity of $\log x$. They are summarized in the diagram below.
2. Chain rule.
3. Concavity and Convexity. Based on convexity of $x \log x$.
4. Data processing inequality.
5. Fano's inequality, and a simpler result that for two IID random variables with entropy H. $\Pr(X = X') \geq \exp(-\mathrm{H}(X))$, assuming H is measured in nats.

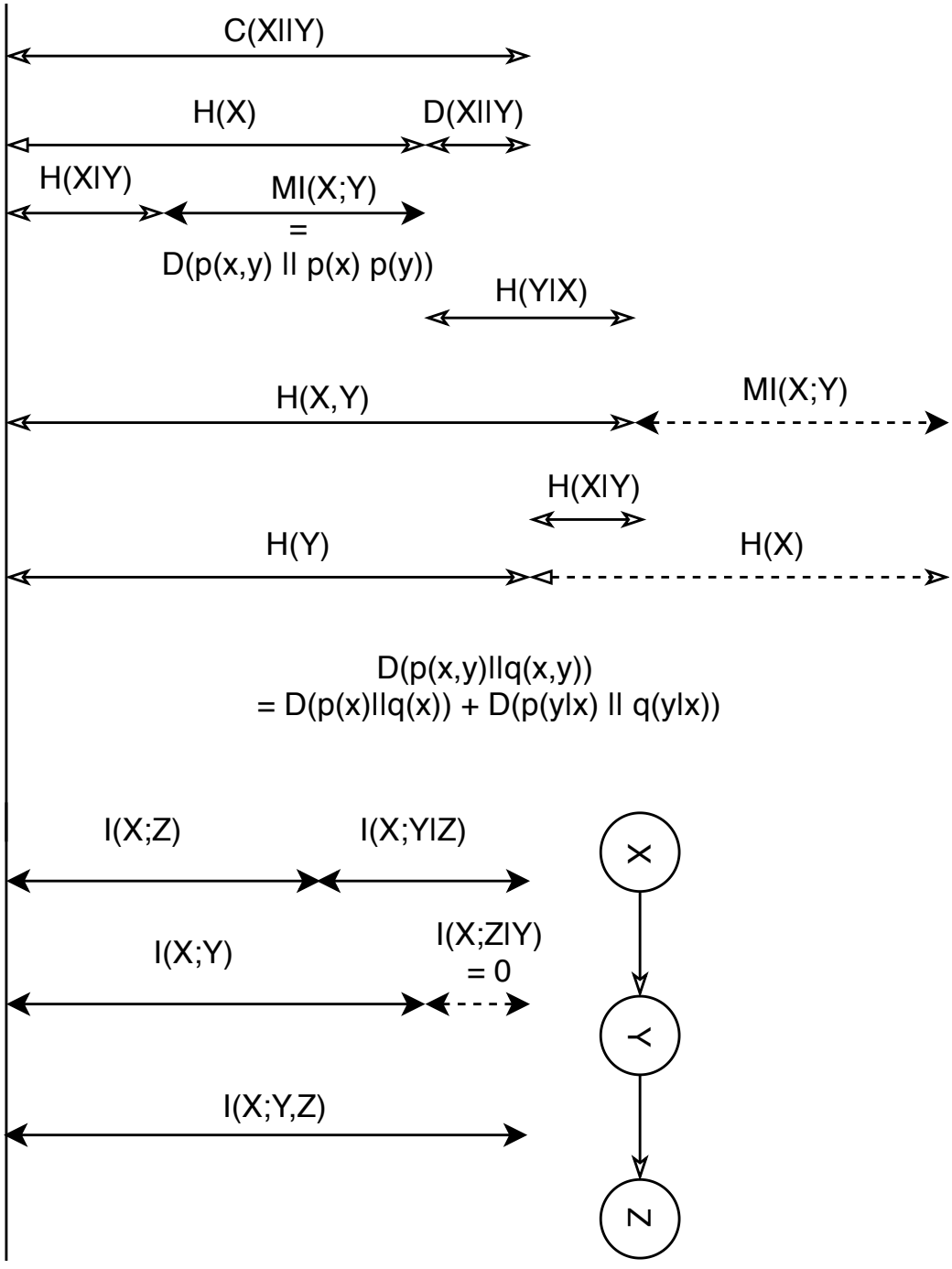

Figure 1: These concepts obey a lot of inequalities as shown in this diagram.