Homework 1 Answers, CS225 | ECE205A, UCSB, 2016 Winter

Copyright:

Question 1
Prove that for each discrete random variable $X \sim p(x)$ it holds that

$$H(X) \leq 2\log\Big(\sum_{x\in\mathcal{X}} \sqrt{p(x)}\Big).$$

Answer: We have that

$$\begin{aligned} 2\log\Big(\sum_{x\in\mathcal{X}} \sqrt{p(x)}\Big) &= 2\log\Big(\sum_{x\in\mathcal{X}} p(x)\cdot 1/\sqrt{p(x)}\Big) \\ &= 2\log \mathrm{E}\left[1/\sqrt{p(x)}\right] \\ &\geq 2\,\mathrm{E}\left[\log\left(1/\sqrt{p(x)}\right)\right] \end{aligned}$$

(by Jensen's Inequality $f(E[Y]) \geq E[f(Y)]$ with the concave $f(Y) = 2\log(Y)$ and $Y = 1/\sqrt{p(X)}$)

$$\begin{aligned} &= \mathrm{E}\left[-\log\left(p(x)\right)\right] \\ &= H(x) \end{aligned}$$

Note that Jensen's inequality can be applied (in the direction sought) only when $f(Y)$ is concave, which is indeed the case for $f(Y) = 2\log(Y)$.

Question 2
Consider two random variables X and Y with alphabets $\mathcal{X} = \{0,1\}$ and $\mathcal{Y} = \{a,b,c\}$ and the following joint distribution:

$$\begin{aligned} &\Pr[X=0,Y=a] = 0.15, &&\Pr[X=0,Y=b] = 0.3, &&\Pr[X=0,Y=c] = 0.05, \\ &\Pr[X=1,Y=a] = 0.25, &&\Pr[X=1,Y=b] = 0.15, &&\Pr[X=1,Y=c] = 0.1. \end{aligned}$$

Determine the entropic quantities $H(X)$, $I(X;Y)$, et cetera, involved in the Venn diagram of the random variables X and Y.

Answer: The entropy of X is straightforward:

$$H(X) = -\Pr[X=0]\log\Pr[X=0] - \Pr[X=1]\log\Pr[X=1] = 1.$$

Next, we calculate explicitly the probabilities for $p(Y)$:

$$\begin{aligned} \Pr[Y=a] &= \Pr[Y=a \mid X=0]\Pr[X=0] + \Pr[Y=a \mid X=1]\Pr[X=1] = 0.4 \\ \Pr[Y=b] &= \Pr[Y=b \mid X=0]\Pr[X=0] + \Pr[Y=b \mid X=1]\Pr[X=1] = 0.45 \\ \Pr[Y=c] &= \Pr[Y=c \mid X=0]\Pr[X=0] + \Pr[Y=c \mid X=1]\Pr[X=1] = 0.15. \end{aligned}$$

With these probabilities $p(Y)$ we get the entropy of Y:

$$\begin{aligned} H(Y) &= -\Pr[Y=a]\log\Pr[Y=a] - \Pr[Y=b]\log\Pr[Y=b] - \Pr[Y=c]\log\Pr[Y=c] \\ &= 1.4577\ldots \end{aligned}$$

The conditional entropy $H(Y \mid X)$ is the last one that we have to calculate explicitly.

$$\begin{aligned}
H(Y \mid X) &= \Pr[X=0]H(Y \mid X=0) + \Pr[X=1]H(Y \mid X=1) \\
&= \frac{1}{2} \sum_{y \in \{a,b,c\}} \Pr[Y=y \mid X=0] \log \Pr[Y=y \mid X=0] \\
&+ \frac{1}{2} \sum_{y \in \{a,b,c\}} \Pr[Y=y \mid X=1] \log \Pr[Y=y \mid X=1] \\
&= 1.3905\ldots
\end{aligned}$$

The remaining entropies then follow in a straightforward manner.

$$\begin{aligned}
H(X \mid Y) &= H(X) + H(Y \mid X) - H(Y) = 0.9328\ldots \\
H(X,Y) &= H(X) + H(Y \mid X) = 2.3905\ldots \\
I(X;Y) &= H(X) - H(X \mid Y) = 0.0672\ldots
\end{aligned}$$

In summary:

$$\begin{aligned}
H(X) &= 1 \\
H(X \mid Y) &\approx 0.9328 \\
H(Y) &\approx 1.4577 \\
H(Y \mid X) &\approx 1.3905 \\
H(X,Y) &\approx 2.3905 \\
I(X;Y) &\approx 0.0672.
\end{aligned}$$

Question 3
Let p_0 and p_1 be probabilities distributions over a finite alphabet $\mathcal{X} = \{1,\ldots,D\}$. Consider a 'mixture' of p_0 and p_1 described by $p_\lambda = (1-\lambda)\cdot p_0 + \lambda \cdot p_1$ with $0 \leq \lambda \leq 1$, such that indeed for $\lambda = 0$ one gets p_0, for $\lambda = 1$ we get p_1, for $\lambda = 1/2$ we have the 50/50 mixture of the two probability distributions and so on. What can you prove about the entropy $H(p_\lambda)$ in terms of $H(p_0)$, $H(p_1)$, λ, and D? Prove your statements.

Answer: Consider the entropy function H on the space of probability distributions $\{(q(1),\ldots,q(D)) \in \mathbb{R}^D_{\geq 0} \mid q(1)+\cdots q(D) = 1\}$. The function H is concave and hence we have by Jensen's Inequality:

$$H((1-\lambda)p_0 + \lambda p_1) \geq (1-\lambda)H(p_0) + \lambda H(p_1).$$

Note that we used here the concavity of the function $\sum_j -p_j \log p_j$, not *the concavity of* $\log p$.

An alternative proof of this lower bound goes as follows. The events of the probability distribution p_λ can be taught of as having a hidden bit b that indicates if p_0 was used or p_1. I.e. with probability $1-\lambda$ we have $b=0$ and we sample from p_0, and with probability λ we have $b=1$ and we sample from p_1. If we know this hidden bit, we are dealing with the conditional entropy $H(p_\lambda \mid b)$. We thus have

$$\begin{aligned}
H(p_\lambda) &\geq H(p_\lambda \mid b) \\
&= \Pr[b=0]H(p_\lambda \mid b=0) + \Pr[b=1]H(p_\lambda \mid b=1) \\
&= (1-\lambda)H(p_0) + \lambda H(p_1).
\end{aligned}$$

In addition this point of view also gives an upper bound through the joint entropy:

$$\begin{aligned}
H(p_\lambda) &\leq H(p_\lambda, b) \\
&= H(p_\lambda \mid b) + H(b) \\
&= (1-\lambda)H(p_0) + \lambda H(p_1) + H((1-\lambda,\lambda)).
\end{aligned}$$

Of course $H(p_\lambda) \leq \log D$ holds as well, but that is trivial.

Question 4
Prove or disprove the triangle inequality for the relative entropy function D. In other words, does it hold that for all probability distributions p, q, r over the alphabet $\mathcal{X}$ we have $D(p \parallel q) + D(q \parallel r) \geq D(p \parallel r)$?

Answer: We can disprove the Triangle inequality by the following counterexample. Consider $x \in \mathcal{X} = \{0, 1\}$ with

	$x = 0$	$x = 1$
$p(x)$	0.3	0.7
$q(x)$	0.4	0.6
$r(x)$	0.5	0.5

Then, we have

$$\begin{aligned} D(p \parallel q) &= p(0) \log \frac{p(0)}{q(0)} + p(1) \log \frac{p(1)}{q(1)} = 0.0302\ldots \\ D(q \parallel r) &= q(0) \log \frac{q(0)}{r(0)} + q(1) \log \frac{q(1)}{r(1)} = 0.0290\ldots \\ D(p \parallel r) &= p(0) \log \frac{p(0)}{r(0)} + p(1) \log \frac{p(1)}{r(1)} = 0.1187\ldots \end{aligned}$$

As a result, $D(p \parallel q) + D(q \parallel r) = 0.0592 \cdots < 0.1187 \cdots = D(p \parallel r)$, hence $D(p \parallel q) + D(q \parallel r) \geq D(p \parallel r)$ does not always hold.