Problem Set 4

MATH 778C, Spring 2009, Austin Mohr (with John Boozer)

April 30, 2009

1. Two n-sided dice with sides labeled 1 through n are rolled, resulting in the i.i.d. random variables X and Y. Let $Z = X + Y$. Compute $H(X)$, $H(X, Y)$, $H(Z)$, $H(X|Z)$, $I(X;Y)$ and $I(X;Y|Z)$.

We compute the entropy of X, we compute directly.

$$\begin{aligned} H(X) &= -\sum_{i=1}^{n} \mathbf{P}(X = i) \log(\mathbf{P}(X = i)) \\ &= -\sum_{i=1}^{n} \frac{1}{n} \log\left(\frac{1}{n}\right) \\ &= \log(n) \end{aligned}$$

We compute the the joint entropy of X and Y directly.

$$\begin{aligned} H(X, Y) &= H(X) + H(Y \mid X) \\ &= H(X) + H(Y) \qquad \text{(since } X \text{ and } Y \text{ are independent)} \\ &= 2\log(n) \end{aligned}$$

To determine the entropy of $Z = X+Y$, observe that there are a total of n^2 possible outcomes. There is no way to get $Z = 1$. For $2 \le k \le n+1$, there are $k-1$ ways to achieve $Z = k$ (for a given k, we have the results $(1, k-1)$, $(2, k-2)$, $\cdots$, $(k-1, 1)$). For $n+2 \le k \le 2n-1$, we see by a similar argument that there are $2n - (k-1)$ ways. Observe that for sums other than $n+1$, there is a kind of parity (both 2 and $2n$ have

1 possibility, 3 and $2n-1$ have 2 possibilities, etc.). Combining this information, we have

$$\begin{aligned}
H(Z) &= -\sum_{i=2}^{2n} \mathbf{P}(Z=i)\log(\mathbf{P}(Z=i)) \\
&= -2\sum_{i=2}^{n+1} \mathbf{P}(Z=i)\log(\mathbf{P}(Z=i)) + \mathbf{P}(Z=n+1)\log(\mathbf{P}(Z=n+1)) \\
&= -2\sum_{i=2}^{n+1} \frac{i-1}{n^2}\log\left(\frac{i-1}{n^2}\right) + \frac{n}{n^2}\log\left(\frac{n}{n^2}\right) \\
&= -\frac{2}{n^2}\sum_{i=1}^{n} i\log\left(\frac{i}{n^2}\right) + \frac{1}{n}\log\left(\frac{1}{n}\right) \\
&= -\frac{2}{n^2}\sum_{i=1}^{n} i\left(\log(i) - \log(n^2)\right) + \frac{1}{n}\log\left(\frac{1}{n}\right) \\
&= -\frac{2}{n^2}\left(\sum_{i=1}^{n} i\log(i) - \sum_{i=1}^{n} i\log(n^2)\right) + \frac{1}{n}\log\left(\frac{1}{n}\right) \\
&= -\frac{2}{n^2}\sum_{i=1}^{n} i\log(i) + \frac{2}{n^2}\cdot\frac{n(n+1)\log(n^2)}{2} + \frac{1}{n}\log\left(\frac{1}{n}\right) \\
&= -\frac{2}{n^2}\sum_{i=1}^{n} i\log(i) + \frac{(2n+1)\log(n)}{n}
\end{aligned}$$

To determine the entropy of X given the value of Z, we employ a similar parity argument to obtain

$$\begin{aligned}
H(X \mid Z) &= -\sum_{i=2}^{2n}\left(\sum_{j=1}^{n} \mathbf{P}(X=j \mid Z=i)\log(\mathbf{P}(X=j \mid Z=i))\right) \\
&= -2\sum_{i=2}^{n+1}\left(\frac{i-1}{n^2}\sum_{j=1}^{i-1}\frac{1}{i-1}\log\left(\frac{1}{i-1}\right)\right) + \frac{1}{n}\sum_{j=1}^{n}\frac{1}{n}\log\left(\frac{1}{n}\right) \\
&= -2\sum_{i=2}^{n+1}\left(\frac{i-1}{n^2}\log\left(\frac{1}{i-1}\right)\right) + \frac{1}{n}\log\left(\frac{1}{n}\right) \\
&= \frac{2}{n^2}\sum_{i=1}^{n} i\log(i) - \frac{1}{n}\log(n)
\end{aligned}$$

We compute the mutual information of X and Y directly.

$$\begin{aligned} I(X;Y) &= H(X) - H(X \mid Y) \\ &= H(X) - H(X) \qquad \text{(since } X \text{ and } Y \text{ independent)} \\ &= 0 \end{aligned}$$

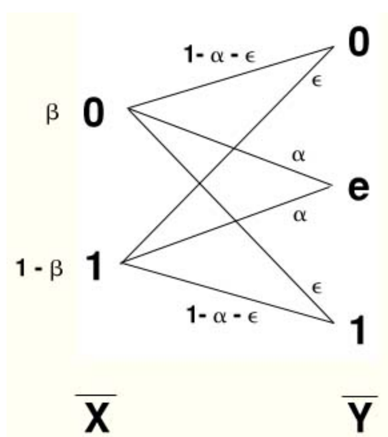
We compute the mutual information of X and Y given the value of Z directly.

$$\begin{aligned} I(X;Y \mid Z) &= H(X \mid Z) - H(X \mid (Y,Z)) \\ &= H(X \mid Z) \end{aligned}$$

since knowing both Y and Z completely determines X.

2. Consider a channel with binary inputs that has both erasures and errors. Let the probability of error be ϵ and the probability of erasure be α. (Hence, the probability of correct transmission is $1 - \alpha - \epsilon$.) What is the capacity of this channel?

We can visualize the channel in the following way



where e denotes erasure and we write the probability distribution of the input in terms of β. The capacity of the channel is given by

$$\max_{\mathbf{P}(X)} I(X;Y) = \max_{\mathbf{P}(X)} [H(Y) - H(Y \mid X)]$$

Now

$$
\begin{aligned}
H(Y) &= -\sum_{y \in Y} \mathbf{P}\left(Y=y\right) \log\left(\mathbf{P}\left(Y=y\right)\right) \\
&= -[\mathbf{P}(Y=0)\log(\mathbf{P}(Y=0)) \\
&\quad + \mathbf{P}(Y=e)\log(\mathbf{P}(Y=e)) \\
&\quad + \mathbf{P}(Y=1)\log(\mathbf{P}(Y=1))] \\
&= -[(\beta(1-\alpha-\epsilon)+(1-\beta)\epsilon)\log(\beta(1-\alpha-\epsilon)+(1-\beta)\epsilon) \\
&\quad +(\beta\alpha+(1-\beta)\alpha)\log(\beta\alpha+(1-\beta)\alpha) \\
&\quad +(\beta\epsilon+(1-\beta)(1-\alpha-\epsilon))\log(\beta\epsilon+(1-\beta)(1-\alpha-\epsilon))] \\
&= -[(\beta(1-\alpha-\epsilon)+(1-\beta)\epsilon)\log(\beta(1-\alpha-\epsilon)+(1-\beta)\epsilon) \\
&\quad +\alpha\log(\alpha) \\
&\quad +(\beta\epsilon+(1-\beta)(1-\alpha-\epsilon))\log(\beta\epsilon+(1-\beta)(1-\alpha-\epsilon))]
\end{aligned}
$$

and

$$
\begin{aligned}
H(Y \mid X) &= -\sum_{x \in X} \mathbf{P}\left(X=x\right)\left(\sum_{y \in Y} \mathbf{P}(Y=y \mid X=x)\log(\mathbf{P}(Y=y \mid X=x))\right) \\
&= -\mathbf{P}(X=0)[\mathbf{P}(Y=0 \mid X=0)\log(\mathbf{P}(Y=0 \mid X=0)) \\
&\quad + \mathbf{P}(Y=e \mid X=0)\log(\mathbf{P}(Y=e \mid X=0)) \\
&\quad + \mathbf{P}(Y=1 \mid X=0)\log(\mathbf{P}(Y=1 \mid X=0))] \\
&\quad - \mathbf{P}(X=1)[\mathbf{P}(Y=0 \mid X=1)\log(\mathbf{P}(Y=0 \mid X=1)) \\
&\quad + \mathbf{P}(Y=e \mid X=1)\log(\mathbf{P}(Y=e \mid X=1)) \\
&\quad + \mathbf{P}(Y=1 \mid X=1)\log(\mathbf{P}(Y=1 \mid X=1))] \\
&= -\beta[(1-\alpha-\epsilon)\log(1-\alpha-\epsilon) \\
&\quad +\alpha\log(\alpha) \\
&\quad +\epsilon\log(\epsilon)] \\
&\quad -(1-\beta)[\epsilon\log(\epsilon) \\
&\quad +\alpha\log(\alpha) \\
&\quad +(1-\alpha-\epsilon)\log(1-\alpha-\epsilon)] \\
&= -(\epsilon\log(\epsilon)+\alpha\log(\alpha)+(1-\alpha-\epsilon)\log(1-\alpha-\epsilon))
\end{aligned}
$$

Since $H(Y \mid X)$ does not depend on β, we need only to maximize $H(Y)$, which occurs at $\beta = \frac{1}{2}$. So

$$\begin{aligned}\max_{\mathbf{P}(X)} I(X;Y) &= \max_{\mathbf{P}(X)}[H(Y) - H(Y \mid X)] \\ &= -(1-\alpha)\log(\frac{1}{2}(1-\alpha)) + \epsilon\log(\epsilon) + \alpha\log(\alpha) + (1-\alpha-\epsilon)\log(1-\alpha-\epsilon)\end{aligned}$$

3. Let $s_t(n)$ denote the number of graphs on n vertices not containing a K_t, $t \geq 3$. Show that

$$1 - \frac{1}{t-1} + o(1) \leq \frac{\log s_t(n)}{\log N} \leq 1 - \epsilon_t,$$

where $N = 2^{\binom{n}{2}}$ is the number of total graphs on n vertices and

$$\epsilon_t = \frac{1}{\binom{t}{2}2^{\binom{t}{2}}}.$$

Hint: Shearer's Inequality.

Proof. Let Y be the random variable representing choosing a random graph on n vertices without a K_t. Let X_i represent the ith edge of that graph.

$$log(S_t(n)) = H(Y) = H(X_1 \ldots X_{\binom{n}{2}})$$

Let G_i be the set of possible edges in a t-vertex subgraph. This is a $\binom{n-2}{t-2}$ covering of the edges. So by Shearer's inequality:

$$\begin{aligned}H(X_1 \ldots X_{\binom{n}{2}}) &\leq \frac{1}{\binom{n-2}{t-2}} \sum_{G_i} H(X_{G_i}) \\ &\leq \frac{1}{\binom{n-2}{t-2}} \sum_{G_i} log(2^{\binom{t}{2}} - 1) \\ &= \frac{\binom{n}{t}}{\binom{n-2}{t-2}} log(2^{\binom{t}{2}} - 1)\end{aligned}$$

$$= \frac{\binom{n}{2}}{\binom{t}{2}} log(2^{\binom{t}{2}} - 1)$$

Now a little calculus tells us $log(2^{\binom{t}{2}} - 1) \leq \binom{t}{2} - \frac{1}{ln(2)2^{\binom{t}{2}}}$, since the derivative of $log(x)$ is $\frac{1}{ln(2)x}$ and $log(x)$ is concave down. Also $\binom{t}{2} - \frac{1}{ln(2)2^{\binom{t}{2}}} \leq \binom{t}{2} - \frac{1}{2^{\binom{t}{2}}}$. Thus

$$log(S_t(n)) \leq \frac{\binom{n}{2}}{\binom{t}{2}} \binom{t}{2} - \frac{1}{2^{\binom{t}{2}}}$$
$$\frac{log(S_t(n))}{log(2^{\binom{n}{2}})} \leq 1 - \frac{1}{\binom{t}{2} 2^{\binom{t}{2}}}$$

□

4. Find the capacity C of the union of two channels $(\mathcal{X}_1, p_1(y_1|x_1), \mathcal{Y}_1)$ and $(\mathcal{X}_2, p_2(y_2|x_2), \mathcal{Y}_2)$, where at each time, one can send a symbol over channel 1 or channel 2 but not both. Assume that the output alphabets are disjoint. Express your answer as a function of the two channel capacities C_1 and C_2.

Since the output alphabets are disjoint, we can assume without loss of generality that the input alphabets are also disjoint (for any letter $x \in \mathcal{X}_1 \cap \mathcal{X}_2$, remove it an add a letter x_1 to $\mathcal{X}_1$ only and x_2 to $\mathcal{X}_2$ only). Now,

$$P(x) = \begin{cases} \lambda P_1(x) & : x \in \mathcal{X}_1 \\ (1-\lambda) P_2(x) & : x \in \mathcal{X}_2 \end{cases}$$

where $\lambda \in [0, 1]$ is the probability of choosing to send through channel 1 (and so $(1 - \lambda)$ is the probability of sending through channel 2). To determine C, we need

$$\begin{aligned} H(Y) &= -\sum_{y \in \mathcal{Y}} P(y) \log(P(y)) \\ &= -\sum_{y \in \mathcal{Y}_1} P(y) \log(P(y)) - \sum_{y \in \mathcal{Y}_2} P(y) \log(P(y)) \end{aligned}$$

$$
\begin{aligned}
&= -\sum_{y\in\mathcal{Y}_1} \frac{\lambda P(y)}{\lambda} \log(\frac{\lambda P(y)}{\lambda}) - \sum_{y\in\mathcal{Y}_2} \frac{(1-\lambda)P(y)}{1-\lambda} \log(\frac{(1-\lambda)P(y)}{1-\lambda}) \\
&= -\lambda \sum_{y\in\mathcal{Y}_1} P_1(y)\log(P_1(y)) - \lambda \sum_{y\in\mathcal{Y}_1} P_1(y)\log(\lambda) \\
&\quad -(1-\lambda)\sum_{y\in\mathcal{Y}_2} P_2(y)\log(P_2(y)) - (1-\lambda)\sum_{y\in\mathcal{Y}_2} P_2(y)\log(1-\lambda) \\
&= \lambda H(Y_1) + (1-\lambda)H(Y_2) + \lambda\log\left(\frac{1}{\lambda}\right) + (1-\lambda)\log\left(\frac{1}{1-\lambda}\right)
\end{aligned}
$$

and

$$
\begin{aligned}
H(Y \mid X) &= -\sum_{x\in\mathcal{X}} P(x)\left(\sum_{y\in\mathcal{Y}} P(y \mid x)log(P(y \mid x))\right) \\
&= -\left(\sum_{x\in\mathcal{X}_1} P(x) + \sum_{x\in\mathcal{X}_2} P(x)\right)\left(\sum_{y\in\mathcal{Y}} P(y \mid x)log(P(y \mid x))\right) \\
&= -\left(\lambda\sum_{x\in\mathcal{X}_1} \frac{P(x)}{\lambda} + (1-\lambda)\sum_{x\in\mathcal{X}_2} \frac{P(x)}{1-\lambda}\right)\left(\sum_{y\in\mathcal{Y}} P(y \mid x)log(P(y \mid x))\right) \\
&= -\left(\lambda\sum_{x\in\mathcal{X}_1} \frac{P(x)}{\lambda}\right)\left(\sum_{y\in\mathcal{Y}} P(y \mid x)log(P(y \mid x))\right) \\
&\quad -\left((1-\lambda)\sum_{x\in\mathcal{X}_2} \frac{P(x)}{1-\lambda}\right)\left(\sum_{y\in\mathcal{Y}} P(y \mid x)log(P(y \mid x))\right) \\
&= -\left(\lambda\sum_{x\in\mathcal{X}_1} \frac{P(x)}{\lambda}\right)\left(\sum_{y\in\mathcal{Y}_1} P(y \mid x)log(P(y \mid x))\right) \\
&\quad -\left((1-\lambda)\sum_{x\in\mathcal{X}_2} \frac{P(x)}{1-\lambda}\right)\left(\sum_{y\in\mathcal{Y}_2} P(y \mid x)log(P(y \mid x))\right) \\
&= \lambda H(Y_1 \mid X_1) + (1-\lambda)H(Y_2 \mid X_2)
\end{aligned}
$$

Now,

$$
\begin{aligned}
C &= \max_{P(X)} I(X;Y) \\
&= \max_{P(X)} H(X) - H(Y \mid X)
\end{aligned}
$$

$$
\begin{aligned}
&= \max_{\lambda} \lambda H(Y_1) + (1-\lambda)H(Y_2) + \lambda \log\left(\frac{1}{\lambda}\right) + (1-\lambda)\log\left(\frac{1}{1-\lambda}\right) \\
&\quad - (\lambda H(Y_1 \mid X_1) - (1-\lambda)H(Y_2 \mid X_2)) \\
&= \max_{\lambda} \lambda C_1 + (1-\lambda)C_2 + \lambda \log\left(\frac{1}{\lambda}\right) + (1-\lambda)\log\left(\frac{1}{1-\lambda}\right)
\end{aligned}
$$

Using calculus, we determine that this expression is maximized when $\lambda = \frac{1}{2^{b-a}+1}$.