Homework Set #2

1. **Entropy and pairwise independence.**
   Let X, Y, Z be three binary Bernoulli($\frac{1}{2}$) random variables that are pairwise independent; that is, $I(X;Y) = I(X;Z) = I(Y;Z) = 0$.

   (a) Under this constraint, what is the minimum value for $H(X, Y, Z)$?

   (b) Give an example achieving this minimum.

   (c) Now suppose that X, Y, Z are three random variables each uniformly distributed over the alphabet $\{1, 2, \ldots, m\}$. Again, they are pairwise independent. What is the minimum value for $H(X, Y, Z)$?

2. **The value of a question.**
   Let $X \sim p(x), \quad x = 1, 2, \ldots, m$.

   We are given a set $S \subseteq \{1, 2, \ldots, m\}$. We ask whether $X \in S$ and receive the answer

   $$Y = \begin{cases} 1, & \text{if } X \in S \\ 0, & \text{if } X \notin S. \end{cases}$$

   Suppose $\Pr\{X \in S\} = \alpha$.

   Find the decrease in uncertainty $H(X) - H(X|Y)$.

   Apparently any set S with a given probability α is as good as any other.

3. **Random questions.**
   One wishes to identify a random object $X \sim p(x)$. A question $Q \sim r(q)$ is asked at random according to $r(q)$. This results in a deterministic answer $A = A(x, q) \in \{a_1, a_2, \ldots\}$. Suppose the object X and the question Q are independent. Then $I(X; Q, A)$ is the uncertainty in X removed by the question-answer (Q, A).

   (a) Show $I(X; Q, A) = H(A|Q)$. Interpret.

   (b) Now suppose that two i.i.d. questions $Q_1, Q_2 \sim r(q)$ are asked, eliciting answers A_1 and A_2. Show that two questions are less valuable than twice the value of a single question in the sense that $I(X; Q_1, A_1, Q_2, A_2) \leq 2I(X; Q_1, A_1)$.

4. **Bottleneck.**
Suppose a (non-stationary) Markov chain starts in one of n states, necks down to $k < n$ states, and then fans back to $m > k$ states. Thus $X_1 \to X_2 \to X_3$, $X_1 \in \{1, 2, \ldots, n\}$, $X_2 \in \{1, 2, \ldots, k\}$, $X_3 \in \{1, 2, \ldots, m\}$, and $p(x_1, x_2, x_3) = p(x_1)p(x_2|x_1)p(x_3|x_2)$.

   (a) Show that the dependence of X_1 and X_3 is limited by the bottleneck by proving that $I(X_1; X_3) \leq \log k$.

   (b) Evaluate $I(X_1; X_3)$ for $k = 1$, and conclude that no dependence can survive such a bottleneck.

5. **Conditional mutual information.**
Consider a sequence of n binary random variables $X_1, X_2, \ldots, X_n$. Each n-sequence with an even number of 1's has probability $2^{-(n-1)}$ and each n-sequence with an odd number of 1's has probability 0. Find the mutual informations

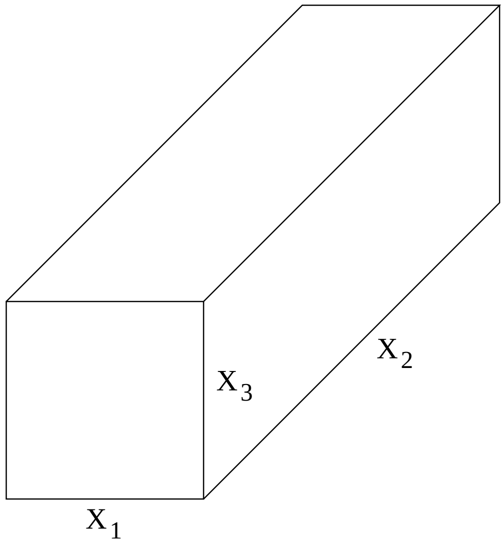
$$I(X_1; X_2), \quad I(X_2; X_3|X_1), \ldots, I(X_{n-1}; X_n|X_1, \ldots, X_{n-2}).$$

6. **Fano's inequality.**
Let $\Pr(X = i) = p_i, i = 1, 2, \ldots, m$ and let $p_1 \geq p_2 \geq p_3 \geq \cdots \geq p_m$. The minimal probability of error predictor of X is $\hat{X} = 1$, with resulting probability of error $P_e = 1 - p_1$. Maximize $H(\mathbf{p})$ subject to the constraint $1 - p_1 = P_e$ to find a bound on P_e in terms of H. This is Fano's inequality in the absence of conditioning.

7. **Random box size.**

   An n-dimensional rectangular box with sides $X_1, X_2, X_3, \ldots, X_n$ is to be constructed. The volume is $V_n = \prod_{i=1}^{n} X_i$. The edge-length l of an n-cube with the same volume as the random box is $l = V_n^{1/n}$. Let $X_1, X_2, \ldots$ be i.i.d. uniform random variables over the interval [0,a].

   Find $\lim_{n\to\infty} V_n^{1/n}$, and compare to $(EV_n)^{1/n}$. Clearly the expected edge length does not capture the idea of the volume of the box.



8. **An AEP-like limit and the AEP.**

   (a) Let $X_1, X_2, \ldots$ be i.i.d. drawn according to probability mass function $p(x)$. Find

   $$\lim_{n\to\infty} [p(X_1, X_2, \ldots, X_n)]^{\frac{1}{n}}.$$

   (b) Let $X_1, X_2, \ldots$ be drawn *i.i.d.* according to the following distribution:

   $$X_i = \begin{cases} 2, & \frac{1}{2} \\ 3, & \frac{1}{3} \\ 4, & \frac{1}{6} \end{cases}$$

   Find the limiting behavior of the product

   $$(X_1 X_2 \cdots X_n)^{1/n}.$$

   (c) Evaluate the limit of $p(X_1, X_2, \ldots, X_n)^{\frac{1}{n}}$ for the distribution in part b.

9. **AEP.**
   Let (X_i, Y_i) be *i.i.d.* $\sim p(x, y)$. We form the log likelihood ratio of the hypothesis that X and Y are independent vs. the hypothesis that X and Y are dependent. What is the limit of

   $$\frac{1}{n} \log \frac{p(X^n)p(Y^n)}{p(X^n, Y^n)}?$$

10. **Entropy of a disjoint mixture.**
Let X_1 and X_2 be discrete random variables drawn according to probability mass functions $p_1(\cdot)$ and $p_2(\cdot)$ over the respective alphabets $\mathcal{X}_1 = \{1, 2, \ldots, m\}$ and $\mathcal{X}_2 = \{m + 1, \ldots, n\}$. Notice that these sets do not intersect. Let

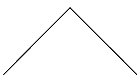
$$X = \begin{cases} X_1, & \text{with probability } \alpha, \\ X_2, & \text{with probability } 1 - \alpha. \end{cases}$$

   (a) Find $H(X)$ in terms of $H(X_1)$ and $H(X_2)$ and α.

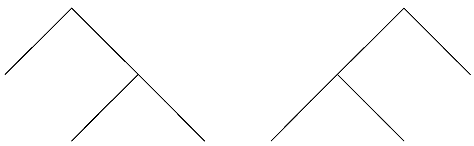
   (b) Maximize over α to show that $2^{H(X)} \leq 2^{H(X_1)} + 2^{H(X_2)}$ and interpret using the notion that $2^{H(X)}$ is the effective alphabet size.

   (c) Let X_1 and X_2 be uniformly distributed over their alphabets. What is the maximizing α and the associated $H(X)$?

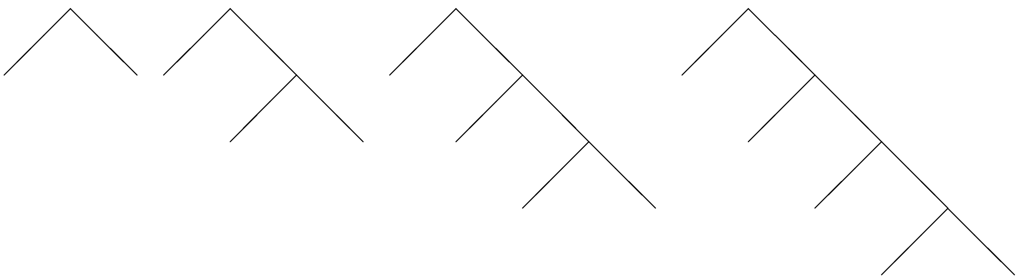
11. **Entropy of a random tree.**
Consider the following method of generating a random tree with n nodes. First expand the root node:



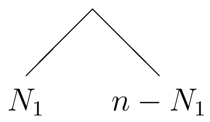
Then expand one of the two terminal nodes at random:



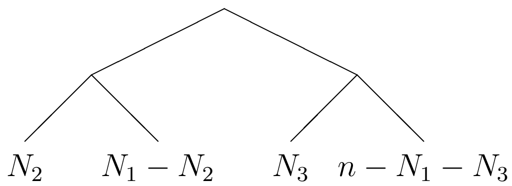
At time k, choose one of the $k - 1$ terminal nodes according to a uniform distribution and expand it. Continue until n terminal nodes have been generated. Thus a sequence leading to a five node tree might look like this:



Surprisingly, the following method of generating random trees yields the same probability distribution on trees with n terminal nodes. First choose an integer N_1 uniformly distributed on $\{1, 2, \ldots, n - 1\}$. We then have the picture.



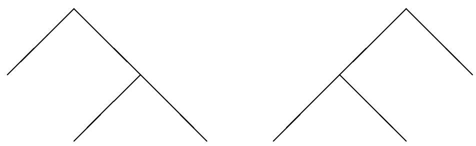
Then choose an integer N_2 uniformly distributed over $\{1, 2, \ldots, N_1 - 1\}$, and independently choose another integer N_3 uniformly over $\{1, 2, \ldots, (n - N_1) - 1\}$. The picture is now:



Continue the process until no further subdivision can be made. (The equivalence of these two tree generation schemes follows, for example, from Polya's urn model.)

Now let T_n denote a random n-node tree generated as described. The probability distribution on such trees seems difficult to describe, but we can find the entropy of this distribution in recursive form.

First some examples. For $n = 2$, we have only one tree. Thus $H(T_2) = 0$. For $n = 3$, we have two equally probable trees:



Thus $H(T_3) = \log 2$. For $n = 4$, we have five possible trees, with probabilities 1/3, 1/6, 1/6, 1/6, 1/6.

Now for the recurrence relation. Let $N_1(T_n)$ denote the number of terminal nodes of T_n in the right half of the tree. Justify each of the steps in the following:

$$\begin{aligned}
H(T_n) &\overset{(a)}{=} H(N_1, T_n) && (1)\\
&\overset{(b)}{=} H(N_1) + H(T_n|N_1) && (2)\\
&\overset{(c)}{=} \log(n-1) + H(T_n|N_1) && (3)\\
&\overset{(d)}{=} \log(n-1) + \frac{1}{n-1}\sum_{k=1}^{n-1}\left[H(T_k) + H(T_{n-k})\right] && (4)\\
&\overset{(e)}{=} \log(n-1) + \frac{2}{n-1}\sum_{k=1}^{n-1} H(T_k). && (5)\\
& && (6)
\end{aligned}$$

(f) Use this to show that

$$(n-1)H_n = nH_{n-1} + (n-1)\log(n-1) - (n-2)\log(n-2), \tag{7}$$

or

$$\frac{H_n}{n} = \frac{H_{n-1}}{n-1} + c_n, \tag{8}$$

for appropriately defined c_n. Since $\sum c_n = c < \infty$, you have proved that $\frac{1}{n}H(T_n)$ converges to a constant. Thus the expected number of bits necessary to describe the random tree T_n grows linearly with n.