Math 4280 HW #13 Solutions.

3.2) Let (X_i, Y_i) be identically distributed random variables $\sim p(x,y)$. 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

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

$$\frac{1}{n}\log\left(\frac{p(X^n)p(Y^n)}{p(X^n, Y^n)}\right) = \frac{1}{n}\left(\log\left(\frac{1}{p(X^n, Y^n)}\right) - \log\left(\frac{1}{p(X^n)}\right) - \log\left(\frac{1}{p(Y^n)}\right)\right)$$

Note that X^n is a random variable taking values in A^n, i.e. $X^n = X_1 X_2 \cdots X_n$. Likewise $Y^n = Y_1 Y_2 \cdots Y_n$ and (X^n, Y^n) takes values in $A^n \times B^n \cong (A \times B)^n$

$$(X^n, Y^n) = (X_1 X_2 \cdots X_n, Y_1 Y_2 \cdots Y_n)$$
$$\cong (X_1, Y_1)(X_2, Y_2) \cdots (X_n, Y_n).$$

From the asymptotic equipartition property (Th'm 3.1.1)

$$-\frac{1}{n}\log\left(\frac{1}{p(X_n)}\right) = \frac{1}{n}\log(p(X_1 X_2 \cdots X_n))$$
approaches $H(A)$

and

$$-\frac{1}{n}\log\left(\frac{1}{p(Y_n)}\right) = \frac{1}{n}\log(p(Y_1 Y_2 \cdots Y_n))$$
approaches $H(B)$

3.2)

and,

$$\frac{1}{n}\log\left(\frac{1}{p(X_n,Y_n)}\right) = \frac{1}{n}\log\frac{1}{p(X_1,Y_1)p(X_2,Y_2)\cdots p(X_n,Y_n)}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{1}{p(X_i,Y_i)}\right)$$

and from the law of large numbers this approaches the expected value of $\log\left(\frac{1}{p(X,Y)}\right)$ as $n \to \infty$ with probability approaching 1.

$$E\left(\log\left(\frac{1}{p(X,Y)}\right)\right) = \sum_{i,j} R_{ij}\log\left(\frac{1}{R_{ij}}\right)$$

since $R_{ij} = pr\{(X,Y) = (a_i, b_j)\}$ is the joint probability.

So $E\left(\log\left(\frac{1}{p(X,Y)}\right)\right) = H(A,B)$.

Putting these together

$$\frac{1}{n}\log\left(\frac{p(X^n)p(Y^n)}{p(X^n,Y^n)}\right) \text{ approaches } H(A,B) - H(A) - H(B)$$
$$= -I(A,B)$$

with probability approaching 1 as $n \to \infty$

$$\boxed{-I(A,B)}$$

3.4 Let X_i be iid $\sim p(x)$, $x \in \{1, 2, \ldots, m\}$
Let $\mu = E(X)$ and $H = -\sum p(x) \log p(x)$
Let $A^n = \{x^n \in \mathcal{X}^n \,|\, |-\frac{1}{n} \log p(x^n) - H| \leq \varepsilon\}$.
Let $B^n = \{x^n \in \mathcal{X}^n \,|\, |\frac{1}{n} \sum_{i=1}^{n} X_i - \mu| \leq \varepsilon\}$.

(a) Does $\Pr\{X^n \in A^n\} \longrightarrow 1$?
yes, this is exactly the A.E.P.

(b) Does $\Pr\{X^n \in A^n \cap B^n\} \longrightarrow 1$?
yes.
$\Pr\{X^n \in B^n\} \longrightarrow 1$ is the law of large numbers.

$$\Pr\{A^n \cup B^n\} = \Pr\{A^n\} + \Pr\{B^n\} - \Pr\{A^n \cap B^n\}$$

So

$$\Pr\{A^n \cap B^n\} = \Pr\{A^n\} + \Pr\{B^n\} - \Pr\{A^n \cup B^n\}$$

For any $\delta > 0$ there exists N_1 and N_2 such that for all $n \geq N_1$

$$\Pr\{A^n\} = \Pr\left\{\left|-\frac{1}{n} \log p(X^n) - H\right| \leq \varepsilon\right\} \geq 1 - \delta/2$$

and for all $n \geq N_2$

$$\Pr\{B^n\} = \Pr\left\{\left|\frac{1}{n} \sum X_i - \mu\right| \leq \varepsilon\right\} \geq 1 - \delta/2.$$

Let $N = \max\{N_1, N_2\}$, and for all $n \geq N$

$$\Pr\{A^n \cap B^n\} \geq 1 - \delta/2 + 1 - \delta/2 - 1 = 1 - \delta.$$

So $\Pr\{X^n \in A^n \cap B^n\} \longrightarrow 1$ as $n \to \infty$

3.4)

(c) Show that $|A^n \cap B^n| \leq 2^{n(H+\varepsilon)}$ for all n.

Since $|A^n \cap B^n| \leq |A^n|$

and $|A^n| \leq 2^{n(H+\varepsilon)}$ from Th'm 3.1.2.

$|A^n \cap B^n| \leq 2^{n(H+\varepsilon)}$.

(d) Show that $|A^n \cap B^n| \geq (\frac{1}{2}) 2^{n(H-\varepsilon)}$ for n sufficiently large.

Since $\Pr\{A^n \cap B^n\} \to 1$, for sufficiently large n, $\Pr\{A^n \cap B^n\} \geq \frac{1}{2}$

So

$$\sum_{\vec{x} \in A^n \cap B^n} p(\vec{x}) \geq \frac{1}{2}$$

because $\vec{x} \in A^n \cap B^n \subseteq A^n$ $p(\vec{x}) \leq 2^{-n(H-\varepsilon)}$,

So $$\frac{1}{2} \leq \sum_{\vec{x} \in A^n \cap B^n} p(\vec{x}) \leq \sum_{\vec{x} \in A^n \cap B^n} 2^{-n(H-\varepsilon)}$$

$$\Rightarrow \qquad = |A^n \cap B^n| \, 2^{-n(H-\varepsilon)}$$

$|A^n \cap B^n| \geq \frac{1}{2} 2^{n(H-\varepsilon)}$, for sufficiently large n.

3.5] Let $X_1, X_2 \cdots$ be an i.i.d. sequence of discrete random variables with entropy $H(X)$.

Let $C_n(t) = \{ x^n \in \mathcal{X}^n \mid p(x^n) \geq 2^{-nt} \}$.

denote the subset of n-sequences with probabilities $\geq 2^{-nt}$

(a) Show that $|C_n(t)| \leq 2^{nt}$.

$$1 \geq pr\{C_n(t)\} = \sum_{\vec{x} \in C_n(t)} p(\vec{x}) \geq \sum_{\vec{x} \in C_n(t)} 2^{-nt} = |C_n(t)| \, 2^{-nt}$$

So

$$|C_n(t)| \, 2^{-nt} \leq 1 \iff |C_n(t)| \leq 2^{nt}.$$

(b) For what values of t does $P\{X^n \in C_n(t)\} \to 1$.

A] For $t > H(\mathcal{X})$.

If $t > H(X)$ and $\varepsilon = t - H(X)$, $\varepsilon > 0$.

$$A_\varepsilon^{(n)} = \{ \vec{x} \in \mathcal{X}^n \mid 2^{-n(H(X)-\varepsilon)} \geq p(x) \geq 2^{-n(H(X)+\varepsilon)} \}$$

Since $t = H(X) + \varepsilon$, $A_\varepsilon^{(n)} \subseteq C_n(t)$. & $pr\{A_\varepsilon^{(n)}\} \leq pr\{C_n(t)\}$.

Thus $Pr\{A_\varepsilon^{(n)}\} \to 1 \Rightarrow Pr\{C_n(t)\} \to 1$.

If $t < H(X)$

then there exists an ε, $H - t > \varepsilon > 0$.
for this ε,

$$A_\varepsilon^{(n)} = \{\vec{x} \mid 2^{-n(H-\varepsilon)} \geq p(x) \geq 2^{-n(H+\varepsilon)}\}$$

since $t < H - \varepsilon$, $2^{-nt} > 2^{-n(H-\varepsilon)}$.

So if $\vec{x} \in C_n(t)$ $p(x) \geq 2^{-nt} > 2^{-n(H-\varepsilon)}$

$\Rightarrow$ $\vec{x} \notin A_\varepsilon^{(n)}$.

So $C_n(t) \subseteq (A_\varepsilon^{(n)})^c$, $\Pr\{C_n(t)\} \leq \Pr\{(A_\varepsilon^{(n)})^c\}$.

Since $\Pr\{A_\varepsilon^{(n)}\} \longrightarrow 1$, $\Pr\{(A_\varepsilon^{(n)})^c\} \longrightarrow 0$

and $\Pr\{C_n(t)\} \longrightarrow 0$ also.

<u>If $t = H(X)$</u> I'm not sure yet.

certainly if the probability is uniform and $|X| = r$

$$pr(\vec{x}) = \frac{1}{r^n} = r^{-n} = 2^{\log_2 r^{-n}} = 2^{-n \log_2 r} = 2^{-nH}$$

for all $\vec{x} \in X^n$

so $C_n(t) = X^n$ and $pr\{C_n(t)\} = 1$.

If the probability is not uniform, I suspect $\Pr\{X^n \in C_n(t)\} \not\longrightarrow 1$, but I'm not sure yet. I'll work on it.

3.7 AEP and source coding. A discrete memoryless source emits a sequence of statistically independent binary digits with probabilities $p(1) = .005$ and $p(0) = .995$. The digits are taken 100 at a time and a binary codeword is provided for every sequence of 100 digits containing 3 or fewer 1's.

(a) Assume that all codewords are the same length find the minimum length required to provide codewords for all sequences with 3 or fewer 1's.

Count the # of sequences with 3 or fewer ones.

0 ones $\binom{100}{0} = 1$
1 one $\binom{100}{1} = 100$
2 ones $\binom{100}{2} = 4950$
3 ones $\binom{100}{3} = 161700$

So total # of sequences with 3 or fewer ones is

166,751.

$\lceil \log_2 166{,}751 \rceil = 18$, so 18 bits are required

3.7)

(b) Calculate the probability of observing a source sequence for which no code word has been assigned.

Probability of a source sequence with 3 or fewer ones

$$\binom{100}{0}\cdot(.995)^{100} + \binom{100}{1}\cdot(.995)^{99}(.005) + \binom{100}{2}\cdot(.995)^{98}(.005)^{2}$$

$$\binom{100}{3}\cdot(.995)^{97}(.005)^{3}$$

$$\approx .9983$$

So the probability that a sequence of 100 bits will not have a code word assigned is

$$\approx 1-.9983 = .0017.$$