Asymptotic Equipartition Property - Seminar 3, part 1

October 22, 2013

Problem 1 (Calculation of typical set) *To clarify the notion of a typical set* $A_\varepsilon^{(n)}$ *and the smallest set of high probability* $B_\delta^{(n)}$*, we will calculate the set for a simple example. Consider a sequence of i.i.d. binary random variables,* $X_1, X_2, ..., X_n$*, where the probability that* $X_i = 1$ *is 0.6 (and therefore the probability that* $X_i = 0$ *is 0.4).*

(a) Calculate $H(X)$*.*

(b) With $n = 25$ *and* $\varepsilon = 0.1$*, which sequences fall in the typical set* $A(n)$*? What is the probability of the typical set? How many elements are there in the typical set? (This involves computation of a table of probabilities for sequences with k 1's,* $0 \leq k \leq 25$*, and finding those sequences that are in the typical set.)*

(c) How many elements are there in the smallest set that has probability 0.9?

(d) How many elements are there in the intersection of the sets in parts (b) and (c)? What is the probability of this intersection?

Solution.

$$X : \begin{pmatrix} 0 & 1 \\ 0.4 & 0.6 \end{pmatrix}$$

(a) $H(x) = -0.4 \cdot \log_2 0.4 - 0.6 \cdot \log_2 0.6 = 0.970\,95$ bits.

(b) $\varepsilon = 0.1$, $[H(X) - \varepsilon, H(X) + \varepsilon] = [0.870\,95, 1.07095]$.

$$A_\varepsilon^{(n)} = \{x \in \{0,1\}^n : H(X) - \varepsilon \leq \frac{-1}{n} \log p(x^n) \leq H(X) + \varepsilon\},$$

We generate a table of probabilities using the following MATLAB code

```
%Bernoulli AEP
p=0.6;
epsilon=0.1;
```

```
q=1-p;
n=25;
k=(0:n)';
Hx=entropy([p,q]);
li=Hx-epsilon; ls=Hx+epsilon;
for i=0:n
    combi(i+1)=nchoosek(n,i);
end
M=[k,combi',binocdf(k,n,p),-1/n*log2(p.^k.*q.^(n-k))];
```

The output table is

0	1	1.1259e-010	1.3219
1	25	4.3347e-009	1.2985
2	300	8.0333e-008	1.2751
3	2300	9.5431e-007	1.2517
4	12650	8.1646e-006	1.2283
5	53130	5.359e-005	1.2049
6	1.771e+005	0.00028072	1.1815
7	4.807e+005	0.0012054	1.1581
8	1.0816e+006	0.0043264	1.1347
9	2.043e+006	0.013169	1.1113
10	3.2688e+006	0.034392	1.0879
11	4.4574e+006	0.077801	1.0645
12	5.2003e+006	0.15377	1.0411
13	5.2003e+006	0.26772	1.0177
14	4.4574e+006	0.41423	0.99435
15	3.2688e+006	0.57538	0.97095
16	2.043e+006	0.72647	0.94755
17	1.0816e+006	0.84645	0.92415
18	4.807e+005	0.92643	0.90076
19	1.771e+005	0.97064	0.87736
20	53130	0.99053	0.85396
21	12650	0.99763	0.83056
22	2300	0.99957	0.80716
23	300	0.99995	0.78376
24	25	1	0.76036
25	1	1	0.73697

The forth column contains -$\frac{1}{n}\log p(x^n)$. The values within the range [$0.870\,95, 1.07095$] are for $11 \leq k \leq 19$

```
i=find(M(:,4)>=li & M(:,4)<=ls);
M
M(i,:)
```

with output

11	4.4574e+006	0.077801	1.0645
12	5.2003e+006	0.15377	1.0411
13	5.2003e+006	0.26772	1.0177
14	4.4574e+006	0.41423	0.99435
15	3.2688e+006	0.57538	0.97095
16	2.043e+006	0.72647	0.94755
17	1.0816e+006	0.84645	0.92415
18	4.807e+005	0.92643	0.90076
19	1.771e+005	0.97064	0.87736

The probability that the number of 1's lies between 11 and 19 is equal to $F(19) - F(10) = 0.970638 - 0.034392 = 0.936246$. Note that this is greater than $1 - \varepsilon$, i.e., the n is large enough for the probability of the typical set to be greater than $1 - \varepsilon$.

The number of elements in the typical set can be found using the third column.

$$\left|A_\varepsilon^{(n)}\right| = \sum_{k=11}^{19} \binom{25}{k} = 26\,366\,510.$$

(c) To find the smallest set $B_\delta^{(n)}$ of probability 0.9, we can imagine that we are filling a bag with pieces such that we want to reach a certain weight with the minimum number of pieces. To minimize the number of pieces that we use, we should use the largest possible pieces. In this case, it corresponds to using the sequences with the highest probability. Thus we keep putting the high probability sequences into this set until we reach a total probability of 0.9. Looking at the fourth column of the table, it is clear that the probability of a sequence increases monotonically with k. Thus the set consists of sequences of $k = 25, 24; \ldots,$ until we have a total probability 0.9.

Using the cumulative probability column, it follows that the set $B_\delta^{(n)}$ consist of sequences with $k \geq 13$ and some sequences with $k = 12$. The sequences with $k \geq 13$ provide a total probability of $1 - 0.153768 = 0.846232$ to the set $B_\delta^{(n)}$. The remaining probability of $0.9 - 0.846232 = 0.053768$ should come from sequences with$k = 12$. The number of such sequences needed to fill this probability is at least$0.053768 = p(x^n) = 0.053768/1.460813 \cdot 10^{-8} = 3680690.1$, which we round up to 3680691. Thus the smallest set with probability 0.9 has $33554432 - 16777216 + 3680691 = 20457907$ sequences. Note that the set $B_\delta^{(n)}$ is not uniquely defined - it could include any 3680691 sequences with $k = 12$. However, the size of the smallest set is a well defined number.

(d) The intersection of the sets $A_\varepsilon^{(n)}$ and $B_\delta^{(n)}$ in parts (b) and (c) consists of all sequences with k between 13 and 19, and 3680691 sequences with $k = 12$. The probability of this intersection$= 0.970638 - 0.153768 +$

$0.053768 = 0.870638$, and the size of this intersection $= 33486026 - 16777216 + 3680691 = 20389501$.

■

Problem 2 (Markov's inequality and Chebyshev's inequality) *(a) (Markov's inequality.) For any non-negative random variable X and any $t > 0$, show that*

$$P(X \geq t) \leq \frac{E(X)}{t}.$$

Exhibit a random variable that achieves this inequality with equality.

(b) (Chebyshev's inequality.) Let Y be a random variable with mean μ and variance σ^2. By letting $X = (Y-\mu)^2$, show that for any $\varepsilon > 0$,

$$P(|Y - \mu| > \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}.$$

Solution.

(a)

$$\begin{aligned} E(X) &= \int_0^\infty x dF(x) = \int_0^t x dF(x) + \int_t^\infty x dF(x) \\ &\geq \int_t^\infty x dF(x) \geq \int_t^\infty t dF(x) \\ &= tP(X \geq t). \end{aligned}$$

Rearranging sides and dividing by t we get,

$$P(X \geq t) \leq \frac{E(X)}{t}.$$

Example for "="

$$X = \begin{cases} t & \text{with probability } \frac{\mu}{t} \\ 0 & \text{with probability } 1 - \frac{\mu}{t} \end{cases}$$

where $\mu \leq t$.

(b) In Markov inequality, take $X = (Y-\mu)^2$.

$$P\left((Y-\mu)^2 > \varepsilon^2\right) \leq \frac{E(Y-\mu)^2}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2},$$

and noticing that

$$P\left((Y-\mu)^2 > \varepsilon^2\right) = P(|Y-\mu| > \varepsilon)$$

we get ChebIn.

∎

Problem 3 (AEP and mutual information) *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*

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

Solution.

$$\begin{aligned}\frac{1}{n}\log\frac{p\left(X^n\right)p\left(Y^n\right)}{p\left(X^n, Y^n\right)} &= \frac{1}{n}\log\prod_{i=1}^{n}\frac{p\left(X_i\right)p\left(Y_i\right)}{p\left(X_i, Y_i\right)} = \\ &= \frac{1}{n}\sum_{i=1}^{n}\log\frac{p\left(X_i\right)p\left(Y_i\right)}{p\left(X_i, Y_i\right)} \\ &\to E\left(\log\frac{p\left(X_i\right)p\left(Y_i\right)}{p\left(X_i, Y_i\right)}\right) = -I(X,Y).\end{aligned}$$

Thus, $\frac{p(X^n)p(Y^n)}{p(X^n,Y^n)} \to 2^{-nI(X,Y)}$, which will converge to 1 if X and Y are indeed independent. ∎

Problem 4 (Piece of cake) *A cake is sliced roughly in half, the largest piece being chosen each time, the other pieces discarded. We will assume that a random cut creates pieces of proportions:*

$$X = \begin{cases} \left(\frac{2}{3}, \frac{1}{3}\right) & w.p.\ \frac{3}{4} \\ \left(\frac{2}{5}, \frac{3}{5}\right) & w.p.\ \frac{1}{4} \end{cases}$$

Thus, for example, the first cut (and choice of largest piece) may result in a piece of size $3/5$. Cutting and choosing from this piece might reduce it to size $(3/5)(2/3)$ at time 2, and so on. How large, to first order in the exponent, is the piece of cake after n cuts?

Solution. Let C_i be the fraction of the piece of cake that is cut at the ith cut, and let T_n be the fraction of cake left after n cuts. Then we have $T_n = C_1 C_2 \ldots C_n = \prod_{i=1}^{n} C_i$. Hence

$$\begin{aligned}\lim\frac{1}{n}\log T_n &= \frac{1}{n}\sum_{i=1}^{n}\log C_i = E(\log C_1) \\ &= \frac{3}{4}\log\frac{2}{3} + \frac{1}{4}\log\frac{3}{5}.\end{aligned}$$

∎

Problem 5 (The AEP and source coding) *A discrete memoryless source emits a sequence of statistically independent binary digits with probabilities $p(1) = (0.005)$ and $p(0) = 0.995$. The digits are taken 100 at a time and a binary codeword is provided for every sequence of 100 digits containing three or fewer ones.*

(a) *Assuming that all codewords are the same length, find the minimum length required to provide codewords for all sequences with three or fewer ones.*

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

(c) *Use Chebyshev's inequality to bound the probability of observing a source sequence for which no codeword has been assigned. Compare this bound with the actual probability computed in part (b).*

Solution.

(a) The number of 100-bit binary sequences with three or fewer ones is

$$\binom{100}{0} + \binom{100}{1} + \binom{100}{2} + \binom{100}{3} = 166\,751.$$

The required codeword length is $\lceil \log_2 166\,751 \rceil = 18$. (Note that $H(0.005) = 0.0454$, so 18 is quite a bit larger than the 4.5 bits of entropy.)

(b) The probability that a 100-bit sequence has three or fewer ones is

$$\sum_{i=0}^{3} \binom{100}{i} (0.005)^i \, (0.995)^{100-i} = 0.998\,33.$$

Thus the probability that the sequence that is generated cannot be encoded is $1 - 0.99833 = 0.00167$.

(c) In the case of a random variable S_n that is the sum of n i.i.d. random variables $X_1, X_2, \ldots, X_n$, Chebyshev's inequality states that

$$P\left(|S_n - n\mu| \geq \varepsilon\right) \leq \frac{n\sigma^2}{\varepsilon^2},$$

where μ and σ^2 are the mean and the variance of X_i, respectively. $E(S_n) = n\mu$, $V(S_n) = n\sigma^2$. In this problem, $n = 100$, $\mu = 0.005$, and $\sigma^2 = (0.005)(0.995)$. Note that $S_{100} \geq 4$ if and only if $|S_{100} - 100(0.005)| \geq 3.5$, so we should choose $\varepsilon = 3.5$. Then

$$P\left(S_{100} \geq 4\right) \leq \frac{100(0.005)(0.995)}{(3.5)^2} \approx 4.\,061\,2 \times 10^{-2}.$$

This bound is much larger than the actual probability 0.00167.

■

Problem 6 (Sets defined by probabilities) *Let $X_1, X_2, \ldots$ be an i.i.d. sequence of discrete random variables with entropy $H(X)$. Let*

$$C_n(t) = \left\{x^n \in X^n : p(x^n) \geq 2^{-nt}\right\}$$

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

(a) Show $|C_n(t)| \le 2^{nt}$*.*

(b) For what values of t *does* $P(X_n \in C_n(t)) \to 1$*?*

Proof.

(a) Since the total probability of all sequences is less than 1,

$$|C_n(t)| \min_{x^n \in C_n(t)} p(x^n) \le 1 \Longrightarrow |C_n(t)|\, 2^{-nt} \le 1.$$

(b) Since $-\frac{1}{n}\log p(x^n) \to H$, if $t < H$, the probability that $p(x^n) > 2^{-nt}$ goes to 0, and if $t > H$, the probability goes to 1.

■

Problem 7 (An AEP-like limit) *Let* $X_1, X_2, \ldots$ *be i.i.d. drawn according to probability mass function* $p(x)$*. Find*

$$\lim_{n\to\infty} \left[P\left(X_1, X_2, \ldots, X_n\right)\right]^{1/n}.$$

Solution. $\log(X_i)$ are also i.i.d. and

$$\begin{aligned}
\lim_{n\to\infty} \left[P\left(X_1, X_2, \ldots, X_n\right)\right]^{1/n} &= \lim 2^{\log[P(X_1, X_2, \ldots, X_n)]^{1/n}} \\
&= 2^{\lim \frac{1}{n} \sum \log P(X_i)} \;\; a.e. \\
&= 2^{E(\log p(x))} \;\; a.e. \\
&= 2^{-H(X)} \;\; a.e.
\end{aligned}$$

by the strong law of large numbers (assuming of course that $H(X)$ exists). ■