UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

Problem Set 6

Fall 2019

1. **Interesting Bernoulli Convergence**

   Consider an independent sequence of random variables where $X_n \sim B(\frac{1}{n})$.

   (a) Prove that X_n converges to 0 in probability.

   (b) Prove that X_n **does not** converge almost surely to 0.

2. **Convergence in Probability**

   Let $(X_n)_{n=1}^{\infty}$, be a sequence of i.i.d. random variables distributed uniformly in $[-1, 1]$. Show that the following sequences $(Y_n)_{n=1}^{\infty}$ converge in probability to some limit.

   (a) $Y_n = \prod_{i=1}^{n} X_i$.

   (b) $Y_n = \max\{X_1, X_2, \ldots, X_n\}$.

   (c) $Y_n = (X_1^2 + \cdots + X_n^2)/n$.

3. **More Almost Sure Convergence**

   (a) Suppose that, with probability 1, the sequence $(X_n)_{n\in\mathbb{N}}$ oscillates between two values $a \neq b$ infinitely often. Is this enough to prove that $(X_n)_{n\in\mathbb{N}}$ does *not* converge almost surely? Justify your answer.

   (b) Suppose that Y is uniform on $[-1, 1]$, and X_n has distribution

   $$\mathbb{P}\big(X_n = (y + n^{-1})^{-1} \bigm| Y = y\big) = 1.$$

   Does $(X_n)_{n=1}^{\infty}$ converge a.s.?

   (c) Define random variables $(X_n)_{n\in\mathbb{N}}$ in the following way: first, set each X_n to 0. Then, for each $k \in \mathbb{N}$, pick j uniformly randomly in $\{2^k, \ldots, 2^{k+1} - 1\}$ and set $X_j = 2^k$. Does the sequence $(X_n)_{n\in\mathbb{N}}$ converge a.s.?

(d) Does the sequence $(X_n)_{n\in\mathbb{N}}$ from the previous part converge in probability to some X? If so, is it true that $\mathbb{E}[X_n] \to \mathbb{E}[X]$ as $n \to \infty$?

4. **Compression of a Random Source**

Suppose I'm trying to send a text message to my friend. In general, I know I need $\log_2(26)$ bits for every letter I want to send because there are 26 letters in the alphabet. However, it turns out if I have some information on the distribution of the letters, I can do better. For example, I might give the letter e a shorter bit representation because I know it's the most common. Actually, it turns out the number of bits I need on average is the entropy, and in this problem, we try to show why this is true in general.

Let $(X_i)_{i=1}^{\infty} \stackrel{\text{i.i.d.}}{\sim} p(\cdot)$, where p is a discrete PMF on a finite set $\mathcal{X}$. We know the entropy of a random variable X is

$$H(X) = -\sum_{x\in\mathcal{X}} p(x)\log_2 p(x)$$

Since entropy is really a function of the distribution, we could write the entropy as $H(p)$.

(a) Show that

$$-\frac{1}{n}\log_2 p(X_1,\dots,X_n) \xrightarrow{n\to\infty} H(X_1) \qquad \text{almost surely.}$$

(Here, we are extending the notation $p(\cdot)$ to denote the joint PMF of $(X_1,\dots,X_n)$: $p(x_1,\dots,x_n) := p(x_1)\cdots p(x_n)$.)

(b) Fix $\epsilon > 0$ and define $A_\epsilon^{(n)}$ as the set of all sequences $(x_1,\dots,x_n) \in \mathcal{X}^n$ such that:

$$2^{-n(H(X_1)+\epsilon)} \le p(x_1,\dots,x_n) \le 2^{-n(H(X_1)-\epsilon)}.$$

Show that $\mathbb{P}((X_1,\dots,X_n) \in A_\epsilon^{(n)}) > 1-\epsilon$ for all n sufficiently large. Consequently, $A_\epsilon^{(n)}$ is called the **typical set** because the observed sequences lie within $A_\epsilon^{(n)}$ with high probability.

(c) Show that $(1-\epsilon)2^{n(H(X_1)-\epsilon)} \le |A_\epsilon^{(n)}| \le 2^{n(H(X_1)+\epsilon)}$, for n sufficiently large. Use the union bound.

Parts (b) and (c) are called the **asymptotic equipartition property (AEP)** because they say that there are $\approx 2^{nH(X_1)}$ observed sequences which each have probability $\approx 2^{-nH(X_1)}$. Thus, by discarding the sequences outside of $A_\epsilon^{(n)}$, we need only keep track of $2^{nH(X_1)}$ sequences, which means that a length-n sequence can be compressed into $\approx nH(X_1)$ bits, requiring $H(X_1)$ bits per symbol.

(d) (**optional**) Now show that for any $\delta > 0$ and any positive integer n, if $B_n \subseteq \mathcal{X}^n$ is a set with $|B_n| \leq 2^{n(H(X_1)-\delta)}$, then $\mathbb{P}((X_1, \ldots, X_n) \in B_n) \to 0$ as $n \to \infty$.

This says that we cannot compress the observed sequences of length n into any set smaller than size $2^{nH(X_1)}$.

[*Hint*: Consider the intersection of B_n and $A_\epsilon^{(n)}$.]

(e) (**optional**) Next we turn towards using the AEP for compression. Recall that in order to encode a set of size n in binary, it requires $\lceil \log_2 n \rceil$ bits. Therefore, a naïve encoding requires $\lceil \log_2 |\mathcal{X}| \rceil$ bits per symbol.

From (b) and (d), if we use $\log_2 |A_\epsilon^{(n)}| \approx nH(X_1)$ bits to encode the sequences in $A_\epsilon^{(n)}$, ignoring all other sequences, then the probability of error with this encoding will tend to 0 as $n \to \infty$, and thus an asymptotically error-free encoding can be achieved using $H(X_1)$ bits per symbol.

Alternatively, we can create an error-free code by using $1 + \lceil \log_2 |A_\epsilon^{(n)}| \rceil$ bits to encode the sequences in $A_\epsilon^{(n)}$ and $1 + n\lceil \log_2 |\mathcal{X}| \rceil$ bits to encode other sequences, where the first bit is used to indicate whether the sequence belongs in $A_\epsilon^{(n)}$ or not. Let L_n be the length of the encoding of $X_1, \ldots, X_n$ using this code; show that $\lim_{n\to\infty} \mathbb{E}[L_n]/n \leq H(X_1) + \epsilon$. In other words, asymptotically, we can compress the sequence so that the number of bits per symbol is arbitrary close to the entropy.